Social Influence in Prosocial Behavior: Evidence from a Large-Scale Experiment*

Lorenz Goette Egon Tripodi

This version: January 28, 2019

Abstract

We propose an experiment that prevents social learning and allows to disentangle mechanisms of social influence. Subjects observe another individual's incentives, but not their behavior. We find conformity: when individuals believe that incentives make others contribute more, they also increase their contributions. Conformity is driven by individuals who feel socially close to their partner. However, when incentives don't raise others' contributions, individuals reduce contributions. This pattern cannot be explained by incentive inequality (Breza et al., 2017). We conclude that norm adherence is weakened when incentives are ineffective. Our results show that information about others' economic environment generates social influence.

*Goette: University of Bonn, Institute for Applied Microeconomics, Adenauerallee 24-42, 53113 Bonn; lorenz.goette@uni-bonn.de. Tripodi: European University Institute, Department of Economics, Via delle Fontanelle 18, 50014 San Domenico di Fiesole; egon.tripodi@eui.eu. This project has benefited from several conversations with Michele Belot, Fabian Dvorak, David K. Levine, and Simone Quercia. We also thank numerous participants of the San Diego Spring School and of the Bologna SEYE 2018 workshop, as well as seminar audience in Melbourne for useful feedback. We are grateful to Caterina Giannetti and Katharina Janezic for helpful discussions at SEYE 2018 and EDP Jamboree 2018. Research funding from the University of Bonn is gratefully acknowledged. Carina Lenze and Luis Wardenbach provided excellent research assistance. All errors remain our own.

1. Introduction

The increasing social connectivity of modern times fosters opportunities for social interactions and comparisons with others. A growing literature illustrates how information and cues about others' behavior can induce social influence: the effect of others' actions on individual behavior. Social influence plays an important role across a broad range of domains that includes charitable giving (Frey and Meier, 2004), financial decision making (Bursztyn et al., 2014), marketing (Bapna and Umyarov, 2015), political participation (Cantoni et al., 2017), tax evasion (Drago et al., 2015), and well-being (Aral and Nicolaides, 2017). While in most social influence studies individuals *observe* others' behavior, a common theme to many studies in this literature is that social influence can arise both from people's ability to extract instrumental information from others' behavior and from a preference for mimicking others. Disentangling the two mechanisms often remains an important open question. Various models (e.g. Bernheim (1994); Akerlof (1997)) explain the spread of social influence even for unobservable behavior via *conformity*, but isolating this behavioral mechanism requires ruling out the learning opportunities that derive from observing others.

In this paper, we study social influence in prosocial behavior through conformity, i.e. when actions are *not* directly observable. As we will illustrate, social influence makes actions of connected agents strategic complements, but such complementarities are often ignored in standard models of prosocial behavior.[1] This study investigates whether information about the economic environment can be sufficient to propagate social influence. Legislative changes often change the economic environment. One of the main goals of this project is to understand whether social influence can be relevant in empirical applications where the behavior of others cannot be observed but the environmental conditions under which people take actions can.

We analyze social influence through a conceptual framework and an experimental design that let us focus on a normative notion of *conformity* – the manifestation of an intrinsic desire to mimic the behavior of a relevant social reference – as a particular

[1] Much of the theoretical literature models prosocial behavior and public good contributions as games of strategic substitutes. The most prominent examples of such theories are represented by models of pure altruism (Becker, 1974) and impure altruism (Andreoni, 1989, 1990).

mechanism of social influence. *Conformists* adhere to the (expected) actions of a relevant social reference, which they perceive as having normative influence on their own decisions even when these actions carry no informational content about the benefits of taking one action over another. Because such a model assumes that people have in their utility function a desire to adhere to social norms of behavior, rational agents will try to mimic others' behavior even when this is unobservable.

In a large online experiment, 2,914 individuals engage in pairwise interactions before they independently take part in a real effort donation task. The two main outcomes of interest are (i) the amount of charitable donations individually generated through the donation task and (ii) expectations of the amount generated by the other player in the pair. In our task, individuals can generate donations to a charity through a tedious physical task. We experimentally manipulate the private incentives of the individuals to generate donations: for each of the two players in a pair, we simultaneously vary one of three levels of piece-rate (*zero*, *moderate*, and *high*) private incentives to generate charitable donations for *Médecins Sans Frontières*. Variation in the incentives of the other player in the pair allows to uncover social influence among partners: if an individual cares to adhere to the behavior of the partner, an increase in partner's incentives will have both a direct effect on partner's donations and an indirect effect on the individual's donations as she tries to minimize distance with the partner's actions. Endowed with a measure of expectations about partner's donations, we can identify the social influence effects of partner's incentives and evaluate different behavioral motives by estimating the contemporaneous effect of partner's incentives on both expectations (about partner's donations) and donations of the player whose incentives are held constant. Before the treatment manipulation, pairs of subjects participate in a joint problem solving task, which we adopt to induce social proximity between paired players (Chen and Li, 2009; Chen and Chen, 2011) and increase relevance of the partner as a social reference. After that, we elicit a survey measure of social proximity (Cialdini et al., 1997). We then use this measure to investigate how social proximity determines a differential propagation of social influence among more and less closely connected partners.

In our conceptual framework of conformist social influence, individual charitable

contributions are strategic complements with the contributions of others. These strategic complementarities are modulated by social proximity to the perceived social reference. In our experiment, where subjects interact in pairs, we observe complementarities in donations: when partner's incentives increase from *zero* to *moderate*, individuals expect their partner to increase donations and they donate more themselves. These effects are entirely driven by subjects who exhibit a close social connection to their partner, for whom the effect of increasing partner's incentives from *zero* to *moderate* on donations is as large as half the effect of increasing *their own* private incentives from *zero* to *moderate*. However, when partner's incentives further increase from *moderate* to *high*, we find a different result: individuals correctly expect their partners' donations to not be affected by higher incentives, and they themselves donate less. Thus, individual donations respond non-monotonically to partner's incentives. These effects are again driven by the subsample of individuals who feel socially close to their partner. To explain the non-monocity, we pit concave altruistic utility against the hypothesis that the magnitude of incentives faced by the social reference matters for norm adherence (Fuster and Meier, 2009); we provide evidence against concavity and discuss the plausibility of the change in norm adherence. As for the socially more distant half of the sample, we find that these subjects are unaffected by changes in partner's incentives, while they do seem responsive to altruistic and personal motives to complete the experimental task.[2]

Being asked to work for different incentives makes subjects in each pair susceptible to the morale effects of incentive inequality (Breza et al., 2017). We illustrate theoretically that such mechanism can often produce social influence effects akin to the ones predicted by conformity, and we highlight moments that allow to distinguish conformity from incentive inequality. An example of such violation is when individuals with *no incentives* increase donations in response to an increase in their partner's incentives. Morale effects cannot explain this pattern, because the incentive increase for the partner exacerbates incentive inequality and damages morale, but conformity can because

[2]A control treatment with no incentives shows that charitable contributions are unrelated to social proximity towards the partner. This indicates that systematic differences in response to partner's incentives across the social proximity dimension are unlikely to be due to differences in altruism.

individuals wish to minimize distance with donations of their social reference. In a series of tests that take into account the full set of restrictions implied by the morale effects of incentive inequality, we find that the behavior of subjects with a strong social connection to their partners violates the predictions of incentive inequality.

Our work broadly contributes to the large literature in economics and psychology that has studied empirically whether social information can produce social influence on prosocial behavior, both in the lab (Cason and Mui, 1998; Bohnet and Zeckhauser, 2004; Eckel and Wilson, 2007; Krupka and Weber, 2009; Servátka, 2009; Duffy and Kornienko, 2010) and in the field (Frey and Meier, 2004; Shang and Croson, 2009; Chen et al., 2010; Fellner et al., 2013; Cantoni et al., 2017). Our main contribution to this literature is to show that observing others' behavior is not necessary for people subject to social influence to try inferring how others behave and mimic their behavior.

We also contribute to a growing literature that tries to disentangle mechanisms of social influence. While we are not the first that try to separately identify social learning from conformity (Bursztyn et al., 2014; Lahno and Serra-Garcia, 2015; Gilchrist and Sands, 2016), our experiment is, to the best of our knowledge, the first with a focus on conformity in an environment that completely shuts down any opportunity for social learning. Moreover, compared to these papers, we are the first to study conformity in the prosocial domain: Bursztyn et al. (2014) investigate social learning and the shared experience of holding an asset as distinct mechanisms of peer effects in financial decisions; Lahno and Serra-Garcia (2015) isolate conformity in lottery choice through a decision environment stripped down of complexity to minimize the scope for social learning; Gilchrist and Sands (2016) use weather instruments to estimate the effect of cumulative movie viewership on the probability of going to watch a movie and run various robustness checks to rule out social learning about quality of the movie.

Our evidence can enrich the understanding of how norm adherence is affected by the economic environment, indicating that the effect of incentives on adherence to social norms of behavior needs not just depend on the incidence of incentives (Gneezy and Rustichini, 2000a,b; Fuster and Meier, 2009), and that also the *magnitude* of incentives matters.[3] We also add, to an empirical literature documenting the role of social

[3]The theory of *ostracism* from Dutta et al. (2018) predicts that sufficiently large incentives are needed for

proximity in social influence mediated by social information (see e.g. Topa (2001); Leider et al. (2009); Bond et al. (2012); Dimant (2018)), evidence that social proximity also modulates social influence in the absence of social information. This evidence is important because it shows that social proximity matters even when benefits of future interactions (heterogeneous for stronger and weaker ties) are absent.

Most closely related to ours is the work of Kessler (2017), who provides field and laboratory evidence that endorsement to a charitable cause can produce large complementarities in giving even when the actual amount of money donated is not observable. He proposes social learning and conformity as primary behavioral channels to explain such findings. We see our work as complementary to Kessler (2017) along two important dimensions: first, for the design that makes the first attempt to separately identify conformity from social learning in the prosocial domain; second, we shed light on some of the modulating factors of social influence that will help inform theoretical developments, the design of institutions, and future studies.

In the remainder of the paper, section 2 presents experimental design and predictions, section 3 illustrates the results and discusses mechanisms of social influence, section 4 concludes.

2. The Experimental Setup

2.1. Experimental Design

We conduct an online experiment with registered workers from Amazon Mechanical Turk. The study develops over five steps and features a full 3×3 between-subject design plus an additional control treatment. All subjects take part in the experiment in randomly formed pairs. Prior to being provided details about the main experimental task, subjects make contact with the other player in the pair. Pairs are formed after the first (registration) step, and the first three steps are common to all pairs. At the fourth step, each pair is randomly assigned to one of ten treatments. The experiment is concluded with a short survey and review of the payoffs. We present below each of

norm adherence to collapse. Although ostracism is very unlikely to play a role in our environment, our evidence is consistent with this theory.

the five steps in detail.[4]

1. Registration. Invited subjects accept the general conditions for participating in the experiment before accessing the software interface. The study begins with some general instructions that summarize the key steps of the experiment: subjects are told that they will be randomly paired to another player with whom they will jointly play a first task, and that a second task will follow for each subject to play independently. After reading the initial instructions, each subject chooses a number from 1 to 6, which they are told will matter for part of their variable pay at the end of the experiment. We introduce *tokens* as the experimental currency. This phase is concluded by a short survey to collect name, gender, age, and experience on Amazon Mechanical Turk, which subjects are told will be shared only with their partner.[5]

2. Joint problem solving task. As subjects get to this step of the experiment, pairs are formed at random and subjects are introduced to their partner: they read stated name, gender, age, country of residence, and experience on Amazon Mechanical Turk of their partner.[6] All our subjects are resident in the United States.

Similar to Chen and Li (2009), we use a joint problem solving task to favor the formation of a social connection between paired players. In this task, pairs of players see the same four famous paintings. For each painting, subjects are incentivized to identify – in coordination with their partner – the corresponding artist from a list of five: each subject in the pair earns 20 tokens each time *both* partners give the correct artist for the same painting.[7] Paired players can solve the task through a private online chat (see interface in Figure B.2). We differ from Chen and Li (2009) for making rewards dependent on both own and partner's answers to increase incentives to make social contact. Payoffs are revealed at the end of the experiment.

3. Oneness elicitation. We measure social proximity through the *oneness* scale. This is a natural choice for this study for two main reasons: Oneness has been found to ex-

[4]Full experimental instructions can be found in the supplemental material, Appendix C.

[5]We cannot verify that this information is truthfully provided. We ask people to provide a name to facilitate interactions, but we did not expect players to recognize the partner as acquaintance/friend. Chat scripts provide no evidence of pre-existing relationships among paired participants.

[6]The order of arrival to this page constitutes our random matching protocol.

[7]We make the task hard by listing possible artists from the same epoch and with a relatively similar style.

plain social proximity for dyadic relationships relatively well in comparison to more involved questionnaire-based scales from social psychology (Gächter et al., 2015), and it is fast and simple to administer.[8] The oneness scale was first proposed by Cialdini et al. (1997) as a simple mean of two underlying scores: (i) the Inclusion of Other in the Self (IOS) scale and the (ii) WE scale. The IOS scale (Aron et al., 1992) is an easy-to-administer pictorial measure of social proximity between the research subject and a related person, constructed by simply asking subjects to indicate which of seven diagrams, composed of two increasingly overlapping circles, best represents their connection to the related person of interest. Cialdini et al. (1997) later proposed to integrate the IOS scale with the WE scale, which asks subjects to express the extent to which they would refer to themselves and another person of interest as *we*, to capture complementary aspects of group membership embedded to social relationships. Both scales are elicited without incentives.

4. Donation task. For this task, subjects have to decide how many donations to generate for charity and make a point prediction about the number of donations the other player is going to generate. We treat such point prediction as proxy of beliefs of partner's giving.[9] To limit the scope of anchoring effects, we elicit expectations and desired number of donations simultaneously. After recording the two variables, subjects carry out the real effort task that generates these donations. Each donation requires entering 100 sequences of keystroke combinations "w"-"e" on a computer keyboard.[10]

Prior to eliciting beliefs and donations, subjects go through a small training to familiarize themselves with the real effort task, and the software randomly assigns pairs of subjects into one of ten different treatments.

Our experimental treatment manipulations simultaneously vary incentives to behave prosocially for both subjects in a pair. To make it very clear that variation in

[8] See Figure B.3 for an illustration of the elicitation method of these scales.

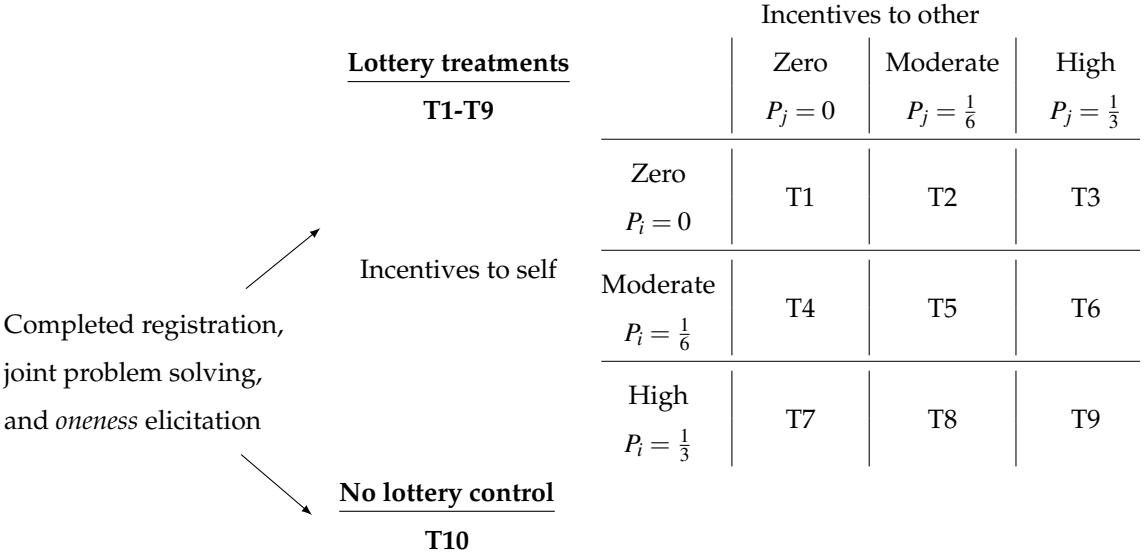
[9] For practical reasons we do not elicit the entire belief distribution, but instead use a measure that most likely captures the perceived mode of giving of the partner. To limit the scope for motivated reasoning, we incentivize correct predictions with a 20 tokens prize.

[10] We choose a sterile task to limit the scope for confounding factors. A similar task has been used in other experiments studying incentives for charitable giving (Ariely et al., 2009; Meyer and Tripodi, 2017), and effort provision (DellaVigna and Pope, 2016, 2017).

monetary incentives is random and independent between partners, all players in the nine incentivized treatment conditions are provided with ex-ante identical lottery incentives: they earn 50 tokens for each donation generated if the number picked in *stage 1* matches the roll of a fair die. Across incentivized treatments we vary, for each player, the *expected* stakes of monetary incentives by means of a simple information device that randomly determines whether to disclose if the matching die has a face number between the largest three or the smallest three figures of a die. When this signal is provided, depending on the initial number chosen, this either reduces to zero the chances of getting the piece-rate incentive to generate donations (incentives are *zero*), or it increase chances to 1 in 3 (incentives are *high*). When this signal is not provided, the probability of getting the piece-rate incentive for generating donations is not updated and remains 1 in 6 (incentives are *moderate*). To make incentives common knowledge within each pair, we reveal to subjects the partner's signal and initial chosen number. We also make sure that subjects understand both their own and the other player's incentives correctly by directly providing them with the updated probabilities of receiving the piece-rate to generate donations (see Figure B.4 for an example). This information revelation scheme produces variation in the magnitude of expected incentives for acting prosocially, for both player i and partner j of each pair, in a full 3×3 between-subject design. We enrich this design with a control *no lottery* condition. Figure 1 schematizes the experimental design.

5. Exit. After carrying out the real effort task that generates the chosen amount to be donated to the charity, subjects answer some unincentivized questions to check comprehension. The summary of individual payoffs concludes the experiment.

Figure 1: Overview of Experimental Design and Treatment Assignment



2.2. Conceptual Framework and Predictions

To organize ideas about our strategy for identifying social influence, consider the following simple model of prosocial behavior. Two agents $a = \{i, j\}$ are presented with the opportunity to make a donation d_a at a private cost $c(d_a)$. Personal benefit from overall donations $u(\sum_a d_a)$ is increasing and concave in the social value that donations generate.[11] Agents have an intrinsic preference for their own actions d_i to mimic those of their social reference d_j. Such preference is captured by a conformity function $v(\cdot)$ that is convex, monotonically increasing in the absolute distance between d_i and d_j, and satisfies $v(0) = 0$. We write the utility of agent i from contributing d_i as:

$$U(d_i) = u(d_i + d_j) - c(d_i) - \lambda_i \tau_{i,j} v(|d_i - d_j|) \tag{2.1}$$

Parameter λ_i captures the weight that an individual attributes to mimicking one's own social reference's behavior (Akerlof, 1997; Bernheim, 1994).[12] In a real-world situation, we see this as an individual preference for conforming to the behavior of an esteemed social reference, be it e.g. a parental figure, a role model, a friend. A conformist

[11] Function $u(\cdot)$ is monotonically increasing and concave; $c(\cdot)$ is monotonically increasing and convex.

[12] Conformity in this literature is distinct from theories of conformity where observing others' behavior is instrumental to uncertainty reduction see e.g. Banerjee (1992) and Bikhchandani et al. (1992).

adheres to a norm of behavior dictated by the action of a relevant social reference.[13] By introducing $\tau_{i,j}$, we allow for the desire to conform to vary in strength depending on the relationship with the social reference: we expect $\tau_{i,j}$ to be affected by the nature of the relationship between i and the social reference j. Social proximity is one measurable element to characterize the nature of social relationships and modulates social influence (Bond et al., 2012; Gioia, 2017).[14] Like in most models of prosocial behavior, donations embed strategic substitution due to concavity of $u(\cdot)$. Strategic complementarities are introduced by the conformity component of the utility function.[15]

2.2.1. Uncovering Social Influence Through Incentives

We focus on a simple structure of the utility function in equation (2.1) to straightforwardly illustrate how piece-rate incentives to behave prosocially m can be leveraged to identify the strategic complementarities induced by conformity. We assume the altruistic utility component $u(\cdot)$ to be linear in the social benefit of a donation A, the cost of effort $c(\cdot)$ and the conformity utility $v(\cdot)$ to be quadratic.[16]

[13] Sliwka (2007) describes conformists in a related way. In his model some agents are resolute while others are conformists – who think about resolute agents to mimic their expected modes of behavior.

[14] Tie strength as modulating factor of social influence is also common in more sophisticated models of conformism (Patacchini and Zenou, 2012) and status seeking (Immorlica et al., 2017).

[15] Super-modularity of the utility function, namely $\frac{\partial U(\cdot)}{\partial d_i \partial d_j} = u''(\cdot) + \lambda_i \tau_{i,j} v''(\cdot) > 0$, defines strategic complementarity. Whether donations are strategic complements around $D = d_i + d_j$ in the general formulation of utility in equation (2.1) depends on the exact shape of $U(\cdot)$; for standard utility formulation with $u''(\cdot) < 0$ and $u'''(\cdot) > 0$, donations are more likely to be substitutes around a small D and complements around a large D. Necessary condition for super-modularity is $\lambda_i \tau_{i,j} > 0$.

[16] Additionally taking into account impurely altruistic preferences for donating to charity (Andreoni, 1989, 1990) shifts the level of donations, and makes i's and j's donations imperfect substitutes, but cannot make donations strategic complements – and therefore is irrelevant to the illustration of social influence. The structure of conformity utility $v(\cdot)$ is typically assumed to be either an absolute value (Bernheim, 1994), or quadratic (Akerlof, 1997). Quadratic cost of effort has the main drawback of implying higher elasticity of effort than it is typically estimated (DellaVigna, 2018), but it makes the algebra with the conformity term a lot more manageable. Although typical structural assumption in the conformity literature impose symmetry around the behavior of the social reference, symmetry is not necessary for the strategic complementarities.

$$U(d_i) = (d_i + d_j)A + m_i d_i - c d_i^2 - \lambda_i \tau_{i,j}(d_i - d_j)^2 \tag{2.2}$$

We use symmetry to write the optimal donation level in this simple closed form

$$d_i^* = \frac{1}{2c}\left[A + m_i \frac{c + \lambda\tau}{c + 2\lambda\tau} + m_j \frac{\lambda\tau}{c + 2\lambda\tau}\right] \tag{2.3}$$

that leads to the following prediction.

Prediction 1 (Conformity)**.** *Increasing j's incentives to act prosocially has a positive direct effect on j's donations, and a smaller positive indirect effect on i's donations.*

The key implication of this statement is that varying incentives of the social reference to act prosocially can be a sufficient manipulation for detecting social influence in the form of conformity. Moreover, equation (2.3) implies that (i) in the absence of incentives, donations should be similar across agents irrespective of λ or τ, and (ii) the larger λ or τ, the less donations respond to changes in personal incentives m_i.

This theoretical framework offers two approaches to identify conformity through incentives. The first, less data demanding, resorts on estimating the indirect effect of changes in j's incentives to donate on i's donation behavior: conformity predicts that an increase in j's incentives should increase i's donations. The second, precisely identifies the strategic complementarities of conformity by considering the effect of changes in j's incentives on both i's expectations about j's donations and i's donations: if donations are affected by conformity, changes in j's incentives shift both i's beliefs about j's donations and i's donations in the same direction.

Not all agents are expected to conform to their social reference to the same degree. This framework assumes that social proximity modulates conformity, and that individuals have a potentially heterogeneous intrinsic preference to adhere to social norms of behavior. This framework does not say whether such intrinsic preference should be stable or potentially malleable to incentives. In the data, we expect to observe that a sufficiently weak intrinsic preference to conform (small λ), or sufficiently weak social ties to the social reference (small τ), leave donations insensitive by the incentives of others. Conversely, we also expect a stronger intrinsic preference to conform (large λ)

and stronger social ties to the social reference (large τ) to propagate significant conformity. Much like in theories of prosocial behavior with incentives, e.g. for social signaling (Benabou and Tirole, 2006) and peer punishment (Dutta et al., 2018), the extent to which agents wish to adhere to the behavior of a social reference λ can be endogenous to incentives. Providing a theory of endogenous norm adherence is beyond the scope of this paper, and we rather study this relationship empirically to inform future theoretical endeavors.

2.2.2. Incentive Inequality

One possible objection to leveraging heterogeneous monetary incentives to act prosocially for investigating the conformity channel of social influence is that incentive inequality in itself could be a source of strategic complementarities in donations. The recent research of Breza et al. (2017) shows that unjustifiably heterogeneous incentives in work environment can introduce a form of inequity aversion (Fehr and Schmidt, 1999) that damages morale to exert effort. The morale effect of incentive inequality is a salient form of inequity aversion even when opportunities for comparing realized payoffs are limited, and remains meaningful when payoff disparities depend on effort (rather than allocation decisions). In this section, we illustrate when this form of inequity aversion can induce strategic complementarities in donations.

Consider the simple model of prosocial behavior as in (2.2), but replace the conformity term with the morale utility term from Breza et al. (2017).

$$U(d_i) = (d_i + d_j)A + m_i d_i - c d_i^2 + M(m_i, m_j) d_i \tag{2.4}$$

Morale $M(\cdot)$, as illustrated below, is a function of the gap in incentives between i and j, and allows for additional direct psychological incentive effects. Parameters α and β capture the extent to which people differentially dislike disadvantageous and advantageous inequality, respectively. The function $g(m_i)$ captures any sort of direct psychological effects of incentives, and $f(\cdot)$ is monotonically increasing in the gap between incentives and satisfies $f(0) = 0$.

$$M(m_i, m_j) = g(m_i) - \alpha f(m_i - m_j | m_i < m_j) - \beta f(m_j - m_i | m_i > m_j)$$

From this simple model we can derive the closed form of the optimal donation level, which is interpreted in the prediction that follows.

$$d_i^* = \frac{1}{2c}\left[A + m_i - \alpha f(m_i - m_j | m_i < m_j) - \beta f(m_j - m_i | m_i > m_j) + g(m_i)\right]$$

Prediction 2 (Incentive Inequality)**.** *If donors' morale is damaged by incentive inequality, (i) at any m_i, i's donations are monotonically decreasing in the size of incentive inequality, and (ii) an increase (decrease) in either i's or j's incentives that reduces (increases) incentive inequality increases (decreases) donations of both i and j.*

The obvious implication of (i) is what we label a *main diagonal condition*: holding i's incentives constant, i's donations should be highest when incentives are homogeneous, and monotonically decreasing in the size of the $m_i - m_j$ gap.

Part (ii) further illustrates when incentive inequality does introduce strategic complementarities in donations. However, notice how an increase (decrease) in m_j that accentuates (reduces) the gap between m_i and m_j decreases (increases) d_i and has a mixed effect on d_j – strengthening the strategic substitution of donations when the direct incentive effect on d_j dominates the negative (positive) effect of increased (decreased) inequality on j's morale.

2.3. Procedures

To uncover the role and determinants of the conformity channel of social influence, we conduct six sessions of the experiment in 2017, between July 30 and August 4, recruiting 3,467 subjects on Amazon Mechanical Turk.[17] This is an online platform that is becoming increasingly popular for conducting economic experiments (DellaVigna and Pope, 2016) where thousands of registered workers are commonly employed in tasks that require human intelligence. Compared to lab subjects, workers on this platform are more heterogeneous in terms of socio-economic characteristics and have been found to exert more attention to experimental instructions (Hauser and Schwarz,

[17]The experimental software is programmed in oTree (Chen et al., 2016). We collect data over multiple sessions to minimize risks of overloading our server.

2016).[18] In our experiment, subjects that complete the study earn 1.20 USD participation fee plus bonus pay depending on their behavior during the experiment. Tokens constitute the experimental currency at the exchange rate of 1 token=0.005 USD. Completing the experiment took participants 17 minutes and 4 seconds on average. Including participation fee, on average, subjects earned 1.63 USD for themselves, and generated 1.13 USD donations for the charity of our choice – *Médecins Sans Frontières*. For subjects that do not spend time to help the charity, the experiment took only 10 minutes and 33 seconds; including participation fee these subjects earned 1.34 USD on average. Such average earnings are comparable to the 7.25 USD hourly earnings accumulated by the most productive 4% of workers on this platform and are significantly higher than the median hourly earnings of 2 USD (Hara et al., 2018). Participation in the experiment is allowed only once, and no retakes are granter to subjects that accidentally drop out of the study.

2.4. Randomization Checks

From the total of 3,467 subjects that began the experiment, we work with a sample of 2,914 subjects who completed both the joint problem solving (JPS) task and the donation task. In the JPS, subjects score an average of 40 out of the 80 available points. After the JPS subjects report a 2.8 oneness towards their partner on average (on a scale between 1 and 7). Across the ten treatment conditions, subjects on average generate 4.6 donations for *Médecins Sans Frontières*, and predict their partner to generate an average of 3.9 donations. Table 1 shows balance in pre-treatment measures, and lack of differential attrition across treatments.

2.5. Social Proximity

As argued in the conceptual framework, conformity requires some degree of social connection to the social reference.[19] This section discusses interpretation and deter-

[18] Like other studies conducted on this platform, we restrict participation in our experiment to workers with an approval rate above 90%. We also restrict participation to workers residing in the US.

[19] Studying behavioral mechanisms that operate via social interactions is methodologically complex. Some papers leverage existing social relationships and identities, while others induce the formation

Table 1: Summary Statistics of Observable Characteristics and Attrition

(Means and Standard Errors in Parentheses)

	Full sample	No lottery	Lottery									p-value
Incentives to self			Zero			Moderate			High			
Incentives to other			Zero	Moderate	High	Zero	Moderate	High	Zero	Moderate	High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					a) Measured before treatment							
Male	0.452	0.449	0.437	0.441	0.465	0.408	0.473	0.451	0.453	0.458	0.484	0.861
	(0.009)	(0.029)	(0.029)	(0.029)	(0.030)	(0.029)	(0.030)	(0.029)	(0.030)	(0.029)	(0.028)	
Age group	2.524	2.491	2.473	2.500	2.620	2.582	2.513	2.487	2.529	2.548	2.500	0.882
	(0.021)	(0.068)	(0.065)	(0.065)	(0.066)	(0.066)	(0.069)	(0.064)	(0.065)	(0.065)	(0.068)	
Experience	2.605	2.774	2.567	2.666	2.662	2.624	2.564	2.632	2.604	2.568	2.403	0.280
	(0.028)	(0.092)	(0.089)	(0.091)	(0.089)	(0.089)	(0.093)	(0.086)	(0.088)	(0.083)	(0.093)	
Points JPS task	40.199	37.979	40.333	40.966	42.324	39.443	41.392	39.934	41.079	39.535	39.226	0.924
	(0.619)	(1.957)	(1.906)	(1.985)	(1.982)	(2.004)	(2.034)	(1.965)	(1.955)	(1.795)	(2.029)	
Oneness	2.801	2.704	2.847	2.784	2.894	2.793	2.885	2.773	2.831	2.691	2.819	0.829
	(0.030)	(0.093)	(0.096)	(0.097)	(0.097)	(0.098)	(0.103)	(0.094)	(0.095)	(0.096)	(0.089)	
					b) Measured after treatment							
Dropout	0.159	0.138	0.167	0.167	0.147	0.171	0.152	0.163	0.165	0.169	0.151	0.974
	(0.006)	(0.019)	(0.020)	(0.020)	(0.019)	(0.020)	(0.020)	(0.019)	(0.020)	(0.019)	(0.020)	
Observations	2914 [3467]	287 [333]	300 [363]	290 [348]	284 [333]	287 [346]	273 [322]	304 [363]	278 [333]	301 [362]	310 [365]	

Notes: p-value in column (12) is for a one-way ANOVA on ranks (Kruskal-Wallis) test comparing the ten treatment groups in columns (2) to (11). Except for dropout rates ("Dropout"), all statistics refer to the final sample of subjects who completed the experiment. Dropout rates of subjects after treatment assignment computed on the samples reported in square brackets in the "Observations" row.

minants of our measure of social proximity, which we elicit among pairs of strangers after they interact in the JPS task.

Remember that in this task pairs of subjects have four paintings and they need to agree on the correct artist to associate from a list of five artists for each painting. Social contact within each pair occurs in the chat box that allows for instrumental coordination on answers and strategies to solve the task.[20] An average score of 40 out of 80 available points indicates significant coordinated effort to solve the common puzzles; random click-through from both subjects would predict an expected score of 3.2. The chat box also introduces each subject to the partner by reporting partner's stated first name, age, gender, level of experience on the Amazon Mechanical Turk platform, and common US residence. The *oneness* measure of social proximity is meant to capture the extent to which basic demographic information and contact with the other player in the JPS task facilitate the formation of perceived social proximity.[21]

To gain interpretation of the kind of social proximity captured by the oneness scale, it is worth comparing the levels we measure to existing estimates. In other studies, on the same scale from 1 to 7, oneness towards an acquaintance, non-close friend, and close relationship is measured to be on average 2.5, 4.0, and 5.4, respectively (Gächter et al., 2015). In our sample, we measure greatly different levels of oneness, with an inter-quartile range capturing half of the entire range of possible realizations: the first quartile of the distribution is 1, the median is 2.5, the third quartile is 4. Expectedly, many subjects exhibit no social proximity to their partner in the experiment. But it is interesting to notice that at least half of the sample exhibits social proximity towards

of social relationships and identities within the experiment (Goette et al. (2012) and Chen et al. (2014) for reviews of this literature). For our investigation, to avoid contaminating the conformity with other forms of social influence deriving from the prospects of future interactions, we choose the approach of building social relationships among randomly and anonymously matched strangers.

[20] To solve puzzles, many of the subjects realize that they can use Google image search, and they tend to split up paintings to search with their partner.

[21] Figure B.5 provides the distribution of the two psychological scales underlying oneness. These two scales are strongly correlated ($\rho = 0.731$), with the WE scale exhibiting a relatively multi-peaked distribution compared to the clear single peak of the IOS scale (at the lowest level of social proximity). All analyses presented in the results section are robust to replacing either of these two scales as measures of social proximity, and can be made available upon request.

their partner – a stranger with whom they have recently made contact to solve puzzles – similar to social proximity that other studies observe towards acquaintances. This is not a *causal* effect of JPS interactions on social proximity, but gives an indication that the JPS does harness social proximity. More direct causal evidence can be found in Gioia (2017).

In Table A.6, least squares regressions illustrate the correlates of social proximity in this experiment, and highlights the role of both *homophily* (Marmaros and Sacerdote, 2006) and chat box contact (Chen and Li, 2009) in the formation of social proximity. Although age difference between the partners does not seem to be highly predictive of social proximity, the partner being of the same gender and having similar experience on the platform predict significantly higher oneness. The fit of this simple linear regression model improves remarkably when we include a binary indicator – *contact* – for whether players made reciprocal contact through the chat box provided.[22] Players that make reciprocal contact with their partner report 67.5% higher social proximity, and although the decision to engage in chat interactions is endogenous, the relatively strong correlation of 0.294 (column (1)) is indicative of the role of social contact for the development of a social connection.

3. Experimental Results

3.1. Conformity as Social Influence

Table 2 summarizes average beliefs and donations across the ten treatments. We begin by pointing out some basic patterns. Donations monotonically increase in personal incentives, though the incentive effect is sizable when incentives increase from *zero* to

[22] 80.4% of subjects used the chatbox to contact the partner, and 64.6% of pairs managed to have a conversation (*contact* = 1). In these conversations, subject share their knowledge of the paintings, share relevant personal information and considerations (e.g., one says "If my husband was here he would know, he is an art teacher lol", some other says that "Modern art sucks".), and agree upon strategies to solve the task (e.g. "You betcha. I'm googling the heck out of it right now. I've got Miro for the first one, Botticelli for the second, Grant Wood for the 3rd, working on the 4th."). Scripts of these conversations can be made available upon request.

moderate and negligible when incentives further increase from *moderate* to *high*.[23] Beliefs indicate that individuals anticipate these patterns of direct incentive effects correctly, but systematically underestimate the levels of others' generosity.[24]

Moving to the main focus of this paper, notice how the economic environment faced by the partner affects individual donations. Donations are systematically affected in a non-monotonic way by partners' incentives: at every level of personal incentive, donations increase when partners' incentives go from *zero* to *moderate*, and decrease when these go from *moderate* to *high*. This highlights complementarities in donations that vary in strength with the magnitude of partners' incentives.

Beliefs will turn out to help shed light on the mechanisms that can explain this pattern. Notice that a change in partners' incentives from *zero* to *moderate* affects beliefs and donations in the same direction while changes in partners' incentives from *moderate* to *high* do not.

In the remainder of this section, we test the statistical significance of these patterns, we try to explain the evidence through the lenses of the conformity framework of social influence, and we illustrate the inconsistency of these patterns with standard models of prosocial behavior that ignore social influence.

We use the linear regression framework of equation (3.1) to estimate the indirect effect of j's incentives on i's donations. The regression model also includes treatment indicators for the direct effects of i's incentives on i's donations, the *no lottery* control treatment to isolate the *disappointment effect* discussed above, and controls for all observable characteristics of both players in the pair.

It is worth emphasizing that models of (im)pure altruistic giving predict that β_4 and β_5 – the effects of j's incentives on i's donations – should be non-positive. As illustrated in prediction 1 instead, positive values of β_4 and β_5 are consistent with the conformity mechanism of social influence. Our simple theoretical framework with linear altruistic utility and proportional conformity also predicts that a desire to conform

[23] Having mentioned the possibility of receiving incentives to donate leads to a *disappointment effect* on donations among those who are randomized out of incentives.

[24] This evidence is consistent with earlier studies finding that average forecasts of subject participants accurately predict experimental results (DellaVigna and Pope, 2016), and that people underestimate others' prosocial attitudes (e.g. Goette et al. (2006)).

Table 2: Beliefs and Donations Across Treatments (Means and Standard Errors)

		Beliefs of partner's donations			Own donations		
Incentives offered							
No (control)		3.585			3.934		
		(0.205)			(0.222)		
Yes (3x3 treatments)							
		Incentives to other			Incentives to other		
		Zero	Moderate	High	Zero	Moderate	High
Incentives to self	Zero	2.540	4.331	4.637	3.233	3.417	3.190
		(0.182)	(0.215)	(0.208)	(0.217)	(0.230)	(0.210)
	Moderate	2.585	4.832	5.086	5.042	5.546	5.155
		(0.193)	(0.213)	(0.207)	(0.233)	(0.235)	(0.224)
	High	2.374	4.100	4.374	5.299	5.575	5.187
		(0.174)	(0.201)	(0.195)	(0.233)	(0.229)	(0.212)

should lead us to estimate a β_5 larger than β_4, but we do not stress this quantitative prediction that so heavily relies on the convenient assumption of linear altruistic utility.

$$\begin{aligned} Donation_i = \alpha + \beta_1 Lottery_i + \beta_2 Moderate_i + \beta_3 High_i + \\ + \beta_4 Moderate_j + \beta_5 High_j + X_{i,j}\gamma + \varepsilon_i \end{aligned} \tag{3.1}$$

To more directly identify the strategic complementarities of social influence, we additionally estimate the mirror least squares regression (3.2) for individual predictions of donations of the social reference. If donations exhibit strategic complementarity, we should observe that when a change in the incentives of player j leads to an expected increase in j's donations i will increase her own donations as well.

$$\begin{aligned} Belief_i = \gamma + \delta_1 Lottery_j + \delta_2 Moderate_j + \delta_3 High_j + \\ + \delta_4 Moderate_i + \delta_5 High_i + X_{i,j}\omega + \varepsilon_i \end{aligned} \tag{3.2}$$

Estimation of regression models (3.1) and (3.2) is presented in panels (a) and (b) of Table 3, respectively. The regression framework replicates the direct incentive effects

summarized in Result 1. Considering the effects of partner's incentives on individual outcomes, we find that when partners' incentives increase from *zero* to *moderate* individual beliefs about partner's donations increase remarkably by 1.962 ($p < 0.001$) units. At the same time, individual donations also increase by 0.356 ($p = 0.055$) units. These effects are hard to reconcile with a model that ignores the strategic complementarities of prosocial behavior. A further increase in partner's incentives further increases beliefs about partner's donations by 0.278 ($p = 0.095$) units, but causes a drop in donations of the player whose incentives are held constant by 0.357 ($p = 0.046$) units.

The non-monotonicity of donations in partners' incentives is driven by subjects with a strong connection to their partner. In fact, when we estimate (3.1) and (3.2) separately for subjects above and below the median level of social proximity we find that socially distant subjects monotonically increase donations with monetary incentives, they expect their social reference to do the same, and their giving behavior is not significantly affected by the incentives provided to the partner.[25] If at all, consistent with concave altruistic utility, monetary incentives to the partner monotonically decrease one's own giving: donations decrease by 0.214 units and 0.251 when the partner gets moderate and high incentives, respectively, but these point estimates are not significantly different from zero.

Socially close subjects instead behave consistently with the framework of prosocial giving augmented by social influence: when partner's incentives increase from zero to moderate, subjects expect the partner to increase donations by 2.155 ($p < 0.001$) units and they donate 0.837 ($p < 0.001$) units more themselves. However, again, when partner's incentives further increase from moderate to high, donations *decrease* by 0.667 ($p = 0.007$) units and individuals believe that the incentive increase does not affect partner's donations ($p = 0.750$) – as it is indeed the case (the direct effect of increasing incentives from moderate to high for individuals exhibiting social closeness to their partner are not significantly ($p = 0.395$) different from zero).[26]

[25] We decided to partition the sample at the median score of oneness. For robustness, we have tried sample splits at the median score of the JPS task and at the median score of just one of the two psychological scales that are used to construct oneness; the results are qualitatively the same and can be made available upon request.

[26] Importantly, differences in behavior across socially close and socially distant individuals does not

Table 3: OLS for the Effect of Partner's Incentives on Donations and Beliefs

(a) Outcome: Donations	Full sample		Split by oneness		p-
			High	Low	value
	(1)	(2)	(3)	(4)	(5)
Provided Lottery	-0.712***	-0.831***	-0.665*	-1.066***	0.464
	(0.262)	(0.283)	(0.389)	(0.403)	
Incentives to self *(baseline: Zero)*					0.052
Moderate	1.964***	1.970***	1.921***	2.037***	
	(0.183)	(0.182)	(0.254)	(0.260)	
High	2.047***	2.044***	1.712***	2.502***	
	(0.179)	(0.179)	(0.242)	(0.259)	
Incentives to other *(baseline: Zero)*					0.016
Moderate		0.356*	0.837***	-0.214	
		(0.186)	(0.259)	(0.268)	
High		-0.001	0.170	-0.251	
		(0.180)	(0.236)	(0.269)	
Constant	4.663***	4.650***	4.896***	4.248***	0.369
	(0.368)	(0.368)	(0.500)	(0.539)	
H0: Incentives to self *Moderate* = *High*, p-value	0.649	0.686	0.395	0.087	
H0: Incentives to other *Moderate* = *High*, p-value		0.046	0.007	0.888	
H0: Incentives to other *Zero* = *Moderate* = *High* = *0*, p-value		0.080	0.003	0.607	

(b) Outcome: Beliefs	Full sample		Split by oneness		p-
			High	Low	value
	(1)	(2)	(3)	(4)	(5)
Provided Lottery	-1.155***	-1.188***	-1.207***	-1.315***	0.822
	(0.237)	(0.256)	(0.358)	(0.348)	
Incentives to other *(baseline: Zero)*					0.391
Moderate	1.948***	1.962***	2.155***	1.773***	
	(0.161)	(0.160)	(0.222)	(0.221)	
High	2.237***	2.240***	2.227***	2.218***	
	(0.158)	(0.158)	(0.211)	(0.229)	
Incentives to self *(baseline: Zero)*					0.435
Moderate		0.336**	0.420*	0.257	
		(0.167)	(0.222)	(0.240)	
High		-0.253	-0.337	-0.105	
		(0.160)	(0.221)	(0.227)	
Constant	4.273***	4.274***	4.800***	3.625***	0.075
	(0.341)	(0.341)	(0.458)	(0.495)	
H0: Incentives to other *Moderate* = *High*, p-value	0.085	0.095	0.750	0.065	
H0: Incentives to self *Moderate* = *High*, p-value		0.000	0.000	0.109	
H0: Incentives to self *Zero* = *Moderate* = *High* = *0*, p-value		0.001	0.003	0.267	
Observations	2914	2914	1571	1343	

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: All specifications include gender, age group, and experience, of both the player and the partner each player faces, as well as session dummies. Column (5) presents joint F-tests for the null hypotheses that point estimates – for each group of variables – are equal in the high and low oneness subsamples. Standard errors are clustered at the pair level. Results are qualitatively very similar in a seemingly unrelated regression framework that allows for correlation in the error term of individual beliefs and donations.

How can these findings be reconciled through conformity? If we allow for concave altruistic utility in our conformity framework,[27] there are two natural ways of interpreting the puzzling evidence. The first is that because the strategic substitution of (im)purely altruistic preferences and the strategic complementarities of conformity pull donations in opposite directions, strategic complementarities happen to dominate at moderate incentives to the partner and succumb to strategic substitution when partner's incentives are higher. A second plausible interpretation that we put forward is that norm adherence (λ in the theoretical framework) is weakened when the behavior of a social reference is more strongly influenced by monetary self interest. The latter should be the case if, for example, the strength of one's desire to conform is a function of the likely intentions that determine the behavior of the social reference.

Endowed with belief data we can test for the substitution effect. In particular, we can test whether socially close subjects expect their partner to give more when partner's incentives increase from *moderate* to *high*. We find that this is not the case: when partner's incentives increase from *moderate* to *high* beliefs are shifted by just 2 percent of a standard deviation ($p = 0.750$). Therefore, we conclude that substitution effects are an unlikely candidate to explain the non-monotonic effects of partners' incentives.[28]

With this in mind, we conclude in favor of our alternative interpretation: the economic environment that determines the choice of the social reference can affect the strength of conformity.

This interpretation is reminiscent of influential papers by Gneezy and Rustichini (2000a,b) and the more recent study of Fuster and Meier (2009). From their experiments, these authors conclude that incentives weaken adherence to the norms of behavior dictated by the actions of a social reference – let this be a small group or society. An important novel element of distinction of our findings is that incentives do not

appear to be driven by differences in pro social orientation. In fact, we can use the control treatment to show that in the absence of incentives subjects with high social proximity to their partner do not systematically donate differently from subjects with low social proximity to their partner ($p = 0.861$).

[27] We called $u(\cdot)$ in Equation (2.1) the altruistic utility term.

[28] The same test on the whole sample indicates some potential role for substitution ($p = 0.095$). However, differences in beliefs when partner's incentives increase from *moderate* to *high* are largely driven by low social proximity subjects, who, as predicted by conformity, are more strongly affected by personal incentives.

seem to simply shut down adherence to social norms: in fact, the magnitude of incentives matters. Relatively small incentives to act prosocially can preserve a certain level of norm adherence and produce social influence.[29] When this is the case, our evidence suggests that larger incentives are more likely to backfire on the positive spillovers of social influence, and perhaps the power of small (but not large) incentives could be leveraged to crowd-in donations by naturally inducing *epidemics* of prosociality.

3.2. Incentive Inequality and Donor's Morale

The presentation of the results so far has been ignoring the possibility that incentive inequality in itself can affect the morale of an agent to work on a task to generate donations for a charity. Although we think that making very clear to subjects that incentives are allocated randomly and independently from their partner greatly reduces the scope for incentive inequality to be perceived as unfair, we deem appropriate to consider in this section (i) what would be the morale effects of incentive inequality with the partner on individual donations, (ii) how they confound conformity, and (iii) whether incentive inequality effects alone can explain the observed strategic complementarities in donations.

Section 2.2.2 illustrated the potential morale effects of incentive inequality, as proposed by Breza et al. (2017), highlighting the predicted *main diagonal condition* of this theoretical framework: conditional on one's own incentives, donation levels should be highest when incentives for both players in a pair are equal, and monotonically decrease in the gap between one's own and partner's incentives.

In this section, we present a test of the joint hypothesis of the *main diagonal condition* to understand whether the morale effects of incentive inequality alone can explain the variation in donation behavior, or the framework of social influence is necessary to reconcile the findings. Similar to Burks et al. (2009), we devise a likelihood ratio test of the joint null hypothesis that the *main diagonal condition* is an adequate set of

[29] Ostracism as in Dutta et al. (2018) allows to endogenize social norms to demonstrate that it is not the mere incidence of payments that damages norm following, but sufficiently large incentives are instead needed. Albeit aligned with our evidence, for the absence of social interactions *after* the donation, we cannot meaningfully use this theory to explain our findings.

restrictions on the first moments. Table 4 summarizes these restrictions.

Average donations in the nine incentivized treatments of our experiment can be treated as a nine-dimensional normal distribution with means μ_{p_i,p_j} (which we treat as unknown) and diagonal covariance matrix $\Sigma = \sigma^2{}_{p_i,p_j}\mathbb{I}$ (which we treat as known). For the joint test, we use maximum likelihood to determine the vector $\hat{\mu}_{p_i,p_j}$ that best fits the nine dimensional vector of sample means $\overline{Donation}_{p_i,p_j}$ - with and without the inequality constraints imposed by the *main diagonal condition*. A Likelihood Ratio test from the constrained and unconstrained likelihood functions is used to jointly assess these constraints. The test statistic is $\chi^2_{(d)}$ distributed with degrees of freedom d equal to the number of binding inequality constraints.

Table 4: Inequalities in Average Donations between Incentivized Treatments Predicted by the Main Diagonal Condition

		Incentives to other				
		Zero		Moderate		High
Incentives to self	Zero	$\mu_{n,n}$	$>$	$\mu_{n,m}$	$>$	$\mu_{n,h}$
	Moderate	$\mu_{m,n}$	$<$	$\mu_{m,m}$	$>$	$\mu_{m,h}$
	High	$\mu_{h,n}$	$<$	$\mu_{h,m}$	$<$	$\mu_{h,h}$

Qualitatively, it can be noticed from Table 5 (columns (1) to (3)) that the main diagonal condition appears violated when incentives to self are both zero and high. As already noted, this is because incentives to the partner non-monotonically affect individual donations. Also notice, that when we focus on average donations from subjects with strong social proximity to their partner, separately from subjects with weak social proximity, violations of the main diagonal condition of incentive inequality become much starker for the former group and virtually disappear for the latter group. This distinction is important because it already suggests that the morale effects of incentive inequality are less likely to fit the behavior of subjects with a strong connection to their partner.

Moving to a test of the significance of these patterns, Table 5 reports the results of Maximum Likelihood estimation constrained by the main diagonal condition and

Likelihood Ratio tests. In spite of some local violations of the *main diagonal condition*, which cause some of the inequality constraints to be binding, the test does not reject the constrained model in favor of the unconstrained model (panel (a)). Local violations of the theory are not strong enough to reject the null hypothesis that donation behavior in the sample can be fit by social influence effects of incentive inequality on morale.

As previously noted, the effects of partner's incentives on individual donations, which were attributed in the previous section to conformity, are very different for subjects who exhibit a close (oneness above median) connection to the social reference and those who do not. We have highlighted that subjects with below median oneness to the partner display virtually no connection to the partner, and partner's incentives have no influence on the donations of these subjects. Testing the *main diagonal condition* separately for low oneness and high oneness subjects helps shed light on the role for conformity to explain the effects attributed to conformity in the previous section.

In panel (b), we confirm that the restrictions imposed by inequity aversion cannot be rejected among low oneness subjects ($p = 0.737$). In panel (c), instead, we strongly reject the *main diagonal condition* among high oneness subjects ($p = 0.002$). To understand how inequity aversion is rejected for more socially close subjects, it is worth interpreting the two main local violations that determine the results of the joint test. The first local violation is due to the change in average donations between groups of players who get randomized out of incentives: increases in their partner's incentives – that *ceteris paribus* increase incentive inequality – increase their own donations. This result is clearly inconsistent with the morale effects of incentive inequality, and is also inconsistent with a concave altruistic utility of giving.[30] The second local violation is due to the change in average donations between groups of players who get randomized into relatively high incentives (*good news*): decreases in their partners' incentives – that *ceteris paribus* increase incentive inequality – increase their own donations. This result is significant for the decrease in partners' incentives from high to moderate, and

[30]The standard framework of inequity aversion (Fehr and Schmidt, 1999), is less tractable in our setting because realized payoff inequality depends both on incentives provided and effort choices. Such a framework does however makes the clear prediction that partner's incentives should not affect individual donations when an agent gets no incentives, and the t-test for one of the two local violations ($\hat{\mu}_{n,n} = \hat{\mu}_{n,m}$) reported in Table 5 panel (c) provides some evidence against this prediction.

Table 5: Average Donations in Lottery Treatments, Maximum Likelihood Estimates (Coefficient Estimates and Standard Errors in Parentheses)

(a) Full sample		Data			$\hat{\theta}^{ML}_{constrained}$				
		Incentives to other			Incentives to other			Main Diagonal	
		Zero	Moderate	High	Zero	Moderate	High		p-value
		(1)	(2)	(3)	(4)	(5)	(6)		(7)
Incentives to self	Zero	3.233	3.417	3.190	3.320	3.320	3.190	LR: $\chi^2_{(2)} = 1.877$	0.391
		(0.217)	(0.230)	(0.209)	(0.217)	(0.230)	(0.209)		
	Moderate	5.042	5.546	5.155	5.042	5.546	5.155	Local Violations: t-tests	
		(0.233)	(0.235)	(0.224)	(0.233)	(0.235)	(0.224)	H0: $\hat{\mu}_{n,n} = \hat{\mu}_{n,m}$	0.551
	High	5.299	5.575	5.187	5.299	5.366	5.366	H0: $\hat{\mu}_{h,m} = \hat{\mu}_{h,h}$	0.215
		(0.233)	(0.229)	(0.212)	(0.233)	(0.229)	(0.212)		

(b) Low oneness		Data			$\hat{\theta}^{ML}_{constrained}$				
		Incentives to other			Incentives to other			Main Diagonal	
		Zero	Moderate	High	Zero	Moderate	High		p-value
		(1)	(2)	(3)	(4)	(5)	(6)		(7)
Incentives to self	Zero	3.190	2.667	2.593	3.190	2.667	2.593	$\chi^2_{(1)} = 0.113$	0.737
		(0.320)	(0.304)	(0.299)	(0.320)	(0.304)	(0.299)		
	Moderate	4.778	5.105	4.622	4.778	5.105	4.622	Local Violations: t-tests	
		(0.337)	(0.339)	(0.332)	(0.337)	(0.339)	(0.332)	H0: $\hat{\mu}_{h,n} = \hat{\mu}_{h,h}$	0.727
	High	5.549	4.889	5.382	5.456	4.889	5.456		
		(0.370)	(0.323)	(0.331)	(0.370)	(0.323)	(0.331)		

(c) High oneness		Data			$\hat{\theta}^{ML}_{constrained}$				
		Incentives to other			Incentives to other			Main Diagonal	
		Zero	Moderate	High	Zero	Moderate	High		p-value
		(1)	(2)	(3)	(4)	(5)	(6)		(7)
Incentives to self	Zero	3.270	4.099	3.614	3.635	3.635	3.614	$\chi^2_{(2)} = 12.443$	0.002
		(0.295)	(0.333)	(0.285)	(0.295)	(0.333)	(0.285)		
	Moderate	5.263	5.913	5.627	5.263	5.913	5.627	Local Violations: t-tests	
		(0.322)	(0.323)	(0.298)	(0.322)	(0.323)	(0.298)	H0: $\hat{\mu}_{n,n} = \hat{\mu}_{n,m}$	0.057
	High	5.103	6.293	5.034	5.103	5.581	5.581	H0: $\hat{\mu}_{h,m} = \hat{\mu}_{h,h}$	0.003
		(0.297)	(0.316)	(0.277)	(0.297)	(0.316)	(0.277)		

Notes: Degrees of freedom of the Likelihood Ratio test statistic equal the number of binding inequality constraints imposed by the composite null hypothesis. Empirical standard errors of the means are directly fed into the maximum likelihood routine.

may be explained by substitution due to concave (altruistic) utility of giving. However, the fact that expectations about partners' levels of giving are virtually identical between these two groups makes this explanation unlikely.

The *main diagonal condition* has a mirror set of conditions on beliefs across treatments. Table A.7 in appendix shows rejection of the conditions imposed by the morale effects on incentive in equality on beliefs.

Taken together, the results of the analyses in this section indicate that the morale effects of inequality capture many but not all patterns of strategic complementarities in donations. Some of the complementarities observed in the data are in stark contrast with the predictions of inequity aversion. This contrast is particularly strong among subjects with a closer connection to their partner. Hence, the conformity framework of social influence remains the most plausible to explain the data.

3.3. Other Mechanisms of Social Influence

Mechanisms such as social learning, social consumption, reciprocity, and conformism have been proposed to explain the large evidence in support of the hypothesis that most individuals are conditional co-operators (Frey and Meier, 2004). Signaling motives to engage in charitable behavior can endogenously determine strategic complementarities to arise. Having provided consistent evidence of the potential for and determinants of strategic complementarities in the decision to behave pro-socially in the previous sections, we conclude with a brief discussion of some of the several mechanisms that, in addition to conformity, are often at the root of strategic complementarities in prosocial behavior. This discussion also helps rack the arguments that curb the potential for other channels of social influence to explain our findings.

Social learning. When people are asymmetrically informed about relevant parameters, observing others' behavior can facilitate information aggregation. In any charitable giving context, the social value of a prosocial activity is definitely uncertain, and the attitudes of others towards the charitable activity may indeed be informative about the quality of the charity or the social norm of giving to the specific cause. Our experiment excludes any scope for social learning. We make clear to subjects that the value generated from a donation is 0.25 USD and that this is common knowledge. Yet, the

effectiveness of *Médecins Sans Frontières* in generating social value may be uncertain and some subjects may know the charity better than others. Our experiment rules out this channel by making others' donations not observables.

Joint consumption. Especially when it comes to volunteer work, this mechanism plays a major role in producing social influence. A multitude of prosocial actions may involve some sort of social gathering. As a result, individuals jointly engage in the prosocial activity enjoying consumption utility that is determined by common experiences and interactions during the activity. The lack of social interactions among participants during the donation makes it easy to rule out this mechanism as potential concurrent for explaining our findings.

Reciprocity. This mechanism of social influence is often appealed in the context of *local* social dilemmas - where agents directly benefit from the prosocial behavior of others. In most cases, charitable giving can be instead regarded as a *global* social dilemma - in the sense that agents benefit from the prosocial behavior of others only to a very marginal extent. We cannot rule out that reciprocity plays some role for highly altruistic types who enjoy private utility from anybody's contributions to the charity, but we deem the relevance of this mechanism as absolutely minor in our charitable giving setup (and in global social dilemmas more generally).

Signaling motives. The unifying theory of Benabou and Tirole (2006) proposes the signaling of altruism and greed as channels that can endogenuously lead to strategic complementarity or substitutability of donations. For binary donation decisions, they show that complementarities arise when, as more people decide to donate, the image of the pool of donors deteriorates faster than the image of non-donors. While our context is highly anonymous, and our results are unlikely to be driven by *social* signaling, we recognize that the Benabou and Tirole (2006) model admits a self-image interpretation that abilitates the theory of signaling as potential channel underlying the complementarities in giving that we present. The self-signaling theory, on the other hand, struggles to capture how individual behavior should be affected by the behavior of agents that act under different economic incentives, and it is silent on whether the behavior of more or less socially close individuals should induce the complementarities of behavior. For these reasons, we think that conformity is the most reasonable

channel of social influence that operates in our data, but we cannot entirely rule out self-signaling motives as a competing and possibly inter-related explanation.[31]

4. Conclusion

This study proposes a novel experiment to study social influence independently of social learning. In our experiment pairs of players collaborate on a task that provides the opportunity to develop social proximity with the partner. Each individual then independently generates donations to a charity through a tedious task knowing both her incentives and the incentives of her partner. We study how the economic environment faced by the partner spreads social influence by randomly varying incentives for both players in each pair.

We find that information about the economic environment faced by the partner is sufficient to spread social influence. When partner's incentives increase from *zero* to *moderate*, individuals generate more donations for the charity and also expect their partner to donate more. Corroborating the social influence interpretation, we find that the effect of partner's incentives on donations is entirely driven by the behavior of the half of the sample most closely connected to the partner. However, when partner's incentives further increase from *moderate* to *high*, individuals correctly anticipate that this incentive change will not affect partner's charitable donations and their own donations drop significantly. Again, this effect is entirely driven by subjects that feel socially close to their partner. One challenge to the interpretation of this evidence is that individuals may respond with frustration to incentive inequality. We derive the predictions of the model proposed by Breza et al. (2017) to study the effects of incentive inequality in a related setup and reject them for subjects who exhibit social influence – those who feel close to their partner.

Having ruled out incentive inequality as a confound, as well as other alternative explanations, we interpret the non-monotonic effects of partner's incentives on individ-

[31] Jones and Linardi (2014) find that making signaling motives more salient increases conformism; calling this "wallflower" behavior and reinforcing our perception that signaling motives and conformity may have related behavioral roots.

ual donations in the following way: high incentives that are ineffective at increasing partner's donations reduce norm adherence. Higher incentives in our setup, do not affect partners' behavior, but obviously change the weight attached to the different motives for donating to charity, and individuals are less prone to adhere to behavior more prominently driven by self serving motives. This interpretation is related to Fuster and Meier (2009) who find that incentives and norm enforcement are substitutes. In line with (Gneezy and Rustichini, 2000b), our evidence indicates that incentives and norm *adherence* are substitutes, with the important qualification that the incidence of incentives is not a binary switch of norm adherence. The magnitude of incentives matters.

Our results also have methodological implications. Increasingly, social scientists are becoming interested in studying the relationship between beliefs about others' behavior and individual behavior. Such empirical efforts often have to overcome several challenges, which include the notorious reverse causality issue entrenched to the *false consensus* in belief formation.[32] An approach that is growingly being used in the experimental literature, to overcome similar challenges and study the effect of beliefs about others' on individual behavior, is to introduce sources of belief variation that serve as instruments for elicited beliefs (see e.g. Smith (2013); Costa-Gomes et al. (2014)). Our results are important for illustrating a violation of the exclusion restriction that is necessary to use incentives to a social reference as an instrument for beliefs.[33]

This evidence is informative of the mechanisms underlying conformity. As noted by Dutta et al. (2018), whether conformity is a behavioral preference or a social norm in itself is a difficult question to answer in most empirical settings if one considers that individuals can find it optimal to *internalize* social norms instead of investing in introspection to figure out their own favorite strategies. While we do not claim to characterize conformity exclusively as a behavioral preference, we think that our design makes it hard for individuals to internalize social norms for these not being readily available.

[32]The concern that beliefs on others behavior reflects more the response function of the *observer* than the *observed*.

[33]Had we run a partition of our experimental design, with a binary manipulation of incentives, we could be presenting very different (spurious) effects of beliefs on donations, depending on the magnitude of incentives, without being able to test the validity of the exclusion restriction.

In fact, one of the merits of our study design is that, because others' behavior is not observable, in order to enjoy any of the benefits of internalization, subjects would first have to accurately assess what is the social norm in this relatively unfamiliar environment. Although, subjects are asked to predict the social norm (the partner's behavior), their incentives for accuracy are relatively low and they have little information to ground an assessment. In this specific environment, introspection seems relatively more accessible than internalization and conformity seems fairly characterized as a behavioral preference.

An implication of our results is that incentives can be calibrated to be *moderate* in order not to spoil the desire to conform to others' prosocial actions and lead to *epidemics* of socially desirable behavior. We do not expect this calibration exercise to be always easy to carry out. Meyer and Tripodi (2017) illustrate what is possibly a successful execution of such calibration exercise: the German blood collection system. An institutional landscape where some of the main blood collection agencies provide small monetary incentives to donate, producing the *largest* per capita supply of blood donations world-wide. Market designers, however, should also be cautious when they ponder increasing incentives for activities that are partly regulated by a social contract because larger incentives are more likely to backfire on social influence. Consistent with this interpretation is the surprising evidence that *better* paid police officers in West Africa become *more* corrupt (Foltz and Opoku-Agyemang, 2015).

Future research should extend our work in at least two directions. First, relatively little research has been studying the factors that modulate social influence. In line with some of the evidence in Fuster and Meier (2009), Bond et al. (2012) and Kessler (2017), our findings highlight that both the economic environment and social proximity to the social reference modulate the strategic complementarities of social influence, but more research will be needed to establish the robustness of this evidence and better understand the sources of heterogeneous social influence effects. Second, field evidence from less controlled settings will be of paramount importance to determine if, and under what conditions, the strategic complementarities of social influence can dominate the well documented substitutability embedded in the free-riding problem that afflicts the provision of private contributions towards common pools of resources. We learn

from this paper that the intrinsic desire to mimic the behavior of a close social reference can have a significant empirical relevance, especially when policy changes occur that affect the economic incentives to engage in certain behaviors, but more evidence is needed to quantify the economic significance of conformity and better appreciate its relevance for the design of markets and institutions.

References

Akerlof, George A., "Social Distance and Social Decisions," *Econometrica*, 1997, *65* (5), 1005–1027.

Andreoni, James, "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, December 1989, *97* (6), 1447–1458.

_ , "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal*, June 1990, *100* (401), 464.

Aral, Sinan and Christos Nicolaides, "Exercise contagion in a global social network," *Nature Communications*, April 2017, *8*, ncomms14753.

Ariely, Dan, Anat Bracha, and Stephan Meier, "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially," *American Economic Review*, February 2009, *99* (1), 544–555.

Aron, Arthur, Elaine N Aron, and Danny Smollan, "Inclusion of other in the self scale and the structure of interpersonal closeness.," *Journal of personality and social psychology*, 1992, *63* (4), 596.

Banerjee, Abhijit V., "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, August 1992, *107* (3), 797–817.

Bapna, Ravi and Akhmed Umyarov, "Do Your Online Friends Make You Pay? A Randomized Field Experiment on Peer Influence in Online Social Networks," *Management Science*, April 2015, *61* (8), 1902–1920.

Becker, Gary S., "A Theory of Social Interactions," *Journal of Political Economy*, November 1974, *82* (6), 1063–1093.

Benabou, Roland and Jean Tirole, "Incentives and Prosocial Behavior," *American Economic Review*, December 2006, *96* (5), 1652–1678.

Bernheim, B. Douglas, "A Theory of Conformity," *Journal of Political Economy*, 1994, *102* (5), 841–877.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, October 1992, *100* (5), 992–1026.

Bohnet, Iris and Richard Zeckhauser, "Social Comparisons in Ultimatum Bargaining," *Scandinavian Journal of Economics*, October 2004, *106* (3), 495–510.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler, "A 61-million-person experiment in social influence and political mobilization," *Nature*, September 2012, *489* (7415), 295–298.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani, "The Morale Effects of Pay Inequality," *The Quarterly Journal of Economics*, 2017.

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini, "Cognitive skills affect economic preferences, strategic behavior, and job attachment," *Proceedings of the National Academy of Sciences*, May 2009, *106* (19), 7745–7750.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman, "Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions," *Econometrica*, 2014, *82* (4), 1273–1301.

Cantoni, Davide, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang, "Are Protests Games of Strategic Complements or Substitutes? Experimental Evidence from Hong Kong's Democracy Movement," Working Paper 23110, National Bureau of Economic Research January 2017. DOI: 10.3386/w23110.

Cason, Timothy N. and Vai-Lam Mui, "Social Influence in the Sequential Dictator Game," *Journal of Mathematical Psychology*, June 1998, *42* (2), 248–265.

Chen, Daniel L., Martin Schonger, and Chris Wickens, "oTree-An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, March 2016, *9*, 88–97.

Chen, Roy and Yan Chen, "The Potential of Social Identity for Equilibrium Selection," *American Economic Review*, October 2011, *101* (6), 2562–2589.

Chen, Yan and Sherry Xin Li, "Group identity and social preferences," *American Eco-*

nomic Review, 2009, *99* (1), 431–57.

_ , **F. Maxwell Harper, Joseph Konstan, and Sherry Xin Li**, "Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens," *American Economic Review*, September 2010, *100* (4), 1358–1398.

_ , **Sherry Xin Li, Tracy Xiao Liu, and Margaret Shih**, "Which hat to wear? Impact of natural identities on coordination and cooperation," *Games and Economic Behavior*, March 2014, *84* (Supplement C), 58–86.

Cialdini, Robert B., Stephanie L. Brown, Brian P. Lewis, Carol Luce, and Steven L. Neuberg, "Reinterpreting the empathyâĂŞaltruism relationship: When one into one equals oneness.," *Journal of Personality and Social Psychology*, 1997, *73* (3), 481–494.

Costa-Gomes, Miguel A., Steffen Huck, and Georg Weizsäcker, "Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect," *Games and Economic Behavior*, November 2014, *88* (Supplement C), 298–309.

DellaVigna, Stefano, "Structural behavioral economics," Technical Report, National Bureau of Economic Research 2018.

_ **and Devin Pope**, "Predicting Experimental Results: Who Knows What?," Working Paper 22566, National Bureau of Economic Research August 2016. DOI: 10.3386/w22566.

_ **and** _ , "What motivates effort? Evidence and expert forecasts," *The Review of Economic Studies*, 2017, *85* (2), 1029–1069.

Dimant, Eugen, "Contagion of Pro-and Anti-Social Behavior Among Peers and the Role of Social Proximity," Technical Report 2018.

Drago, Francesco, Friederike Mengel, and Christian Traxler, "Compliance behavior in networks: Evidence from a field experiment," 2015.

Duffy, John and Tatiana Kornienko, "Does competition affect giving?," *Journal of Economic Behavior & Organization*, May 2010, *74* (1), 82–103.

Dutta, Rohan, David K Levine, and Salvatore Modica, "Peer Monitoring, Ostracism and the Internalization of Social Norms," Technical Report, David K. Levine 2018.

Eckel, Catherine C. and Rick K. Wilson, "Social learning in coordination games: does status matter?," *Experimental Economics*, September 2007, *10* (3), 317–329.

Fehr, Ernst and Klaus M. Schmidt, "A Theory of Fairness, Competition, and Cooper-

ation," *The Quarterly Journal of Economics*, August 1999, *114* (3), 817–868.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler, "Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information," *Journal of the European Economic Association*, June 2013, *11* (3), 634–660.

Foltz, Jeremy D and Kweku A Opoku-Agyemang, "Do higher salaries lower petty corruption?," *A Policy Experiment on West Africa?s Highways*, 2015.

Frey, Bruno S. and Stephan Meier, "Social Comparisons and Pro-Social Behavior: Testing "Conditional Cooperation" in a Field Experiment," *The American Economic Review*, 2004, *94* (5), 1717–1722.

Fuster, Andreas and Stephan Meier, "Another Hidden Cost of Incentives: The Detrimental Effect on Norm Enforcement," *Management Science*, October 2009, *56* (1), 57–70.

Gächter, Simon, Chris Starmer, and Fabio Tufano, "Measuring the Closeness of Relationships: A Comprehensive Evaluation of the 'Inclusion of the Other in the Self' Scale," *PLOS ONE*, June 2015, *10* (6), e0129478.

Gilchrist, Duncan Sheppard and Emily Glassberg Sands, "Something to talk about: Social spillovers in movie consumption," *Journal of Political Economy*, 2016, *124* (5), 1339–1382.

Gioia, Francesca, "Peer effects on risk behaviour: the importance of group identity," *Experimental Economics*, Mar 2017, *20* (1), 100–129.

Gneezy, Uri and Aldo Rustichini, "A Fine Is a Price," *Journal of Legal Studies*, 2000, *29*, 1–18.

_ and _, "Pay Enough or Don't Pay at All," *The Quarterly Journal of Economics*, 2000, *115* (3), 791–810.

Goette, Lorenz, David Huffman, and Stephan Meier, "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups," *The American Economic Review*, 2006, *96* (2), 212–216.

_ , _ , and _, "The Impact of Social Ties on Group Interactions: Evidence from Minimal Groups and Randomly Assigned Real Groups," *American Economic Journal: Microeconomics*, February 2012, *4* (1), 101–115.

Hara, Kotaro, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch,

and Jeffrey P Bigham, "A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk," in "Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems" ACM 2018, p. 449.

Hauser, David J. and Norbert Schwarz, "Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants," *Behavior Research Methods*, March 2016, *48* (1), 400–407.

Immorlica, Nicole, Rachel Kranton, Mihai Manea, and Greg Stoddard, "Social status in networks," *American Economic Journal: Microeconomics*, 2017, *9* (1), 1–30.

Jones, Daniel and Sera Linardi, "Wallflowers: Experimental Evidence of an Aversion to Standing Out," *Management Science*, March 2014, *60* (7), 1757–1771.

Kessler, Judd B., "Announcements of Support and Public Good Provision," *American Economic Review*, December 2017, *107* (12).

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

Lahno, Amrei M and Marta Serra-Garcia, "Peer effects in risk taking: Envy or conformity?," *Journal of Risk and Uncertainty*, 2015, *50* (1), 73–95.

Leider, Stephen, Markus M. Mobius, Tanya Rosenblat, and Quoc-Anh Do, "Directed Altruism and Enforced Reciprocity in Social Networks," *The Quarterly Journal of Economics*, November 2009, *124* (4), 1815–1851.

Marmaros, David and Bruce Sacerdote, "How do friendships form?," *The Quarterly Journal of Economics*, 2006, *121* (1), 79–119.

Meyer, Christian Johannes and Egon Tripodi, "Sorting into Incentives for Prosocial Behavior," SSRN Scholarly Paper ID 3058195, Social Science Research Network, Rochester, NY October 2017.

Patacchini, Eleonora and Yves Zenou, "Juvenile Delinquency and Conformism," *The Journal of Law, Economics, and Organization*, April 2012, *28* (1), 1–31.

Servátka, Maroš, "Separating reputation, social influence, and identification effects in a dictator game," *European Economic Review*, 2009, *53* (2), 197–209.

Shang, Jen and Rachel Croson, "A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods," *The Eco-*

nomic Journal, October 2009, *119* (540), 1422–1439.

Sliwka, Dirk, "Trust as a signal of a social norm and the hidden costs of incentive schemes," *American Economic Review*, 2007, *97* (3), 999–1012.

Smith, Alexander, "Estimating the causal effect of beliefs on contributions in repeated public good games," *Experimental Economics*, September 2013, *16* (3), 414–425.

Topa, Giorgio, "Social Interactions, Local Spillovers and Unemployment," *The Review of Economic Studies*, April 2001, *68* (2), 261–295.

A. Additional Tables

Table A.6: OLS for Determinants of Social Proximity
(Coefficient Estimates and Standard Errors in Parentheses)

Outcome: Oneness scale	(1)	(2)
Contact		1.434***
		(0.057)
Male	0.120*	0.139**
	(0.062)	(0.056)
Same gender	0.236***	0.180***
	(0.061)	(0.056)
Age, absolute difference	-0.003	-0.001
	(0.003)	(0.003)
Experience, absolute difference	-0.080***	-0.072***
	(0.024)	(0.022)
Constant	3.051***	2.118***
	(0.122)	(0.116)
Observations	2914	2914
R^2	0.014	0.189
Correlation in regression residuals (oneness scale) between partners	0.294 (0.340)	0.167 (0.340)

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: All specifications include age group, experience, and session dummies. Standard errors are clustered at the pair level.

Table A.7: Average Beliefs in Lottery Treatments, Maximum Likelihood Estimates (Coefficient Estimates and Standard Errors in Parentheses)

(a) Full sample		Data			$\hat{\theta}^{ML}_{constrained}$				
		Incentives to self			Incentives to self			Likelihood Ratio	
		Zero	Moderate	High	Zero	Moderate	High	Test statistic	p-value
Incentives to other	Zero	2.540	2.585	2.374	2.561	2.561	2.374		
		(0.182)	(0.193)	(0.174)	(0.182)	(0.193)	(0.174)		
	Moderate	4.331	4.832	4.100	4.331	4.832	4.100	$\chi^2_{(2)} = 6.277$	0.043
		(0.215)	(0.214)	(0.201)	(0.215)	(0.214)	(0.201)		
	High	4.637	5.086	4.374	4.637	4.708	4.708		
		(0.208)	(0.207)	(0.195)	(0.208)	(0.207)	(0.195)		

(b) Low oneness		Data			$\hat{\theta}^{ML}_{constrained}$				
		Incentives to self			Incentives to self			Likelihood Ratio	
		Zero	Moderate	High	Zero	Moderate	High	Test statistic	p-value
Incentives to other	Zero	2.124	2.053	1.754	2.124	2.053	1.754		
		(0.241)	(0.264)	(0.225)	(0.241)	(0.264)	(0.225)		
	Moderate	3.703	3.992	3.357	3.703	3.992	3.357	$\chi^2_{(1)} = 0.041$	0.840
		(0.303)	(0.296)	(0.262)	(0.303)	(0.296)	(0.262)		
	High	3.907	4.301	4.213	3.907	4.256	4.256		
		(0.316)	(0.310)	(0.303)	(0.316)	(0.310)	(0.303)		

(c) High oneness		Data			$\hat{\theta}^{ML}_{constrained}$				
		Incentives to self			Incentives to self			Likelihood Ratio	
		Zero	Moderate	High	Zero	Moderate	High	Test statistic	p-value
Incentives to other	Zero	2.890	3.032	2.859	2.959	2.959	2.859		
		(0.265)	(0.272)	(0.249)	(0.265)	(0.272)	(0.249)		
	Moderate	4.901	5.530	4.878	4.901	5.530	4.878	$\chi^2_{(3)} = 12.248$	0.007
		(0.297)	(0.292)	(0.293)	(0.297)	(0.292)	(0.293)		
	High	5.157	5.783	4.500	5.124	5.124	5.124		
		(0.270)	(0.267)	(0.254)	(0.270)	(0.267)	(0.254)		

Notes: Degrees of freedom of the Likelihood Ratio test statistic equal the number of binding inequality constraints imposed by the composite null hypothesis. Empirical standard errors of the means are directly fed into the maximum likelihood routine.

B. Additional Figures

Figure B.2: Joint Problem Solving Task Software Interface

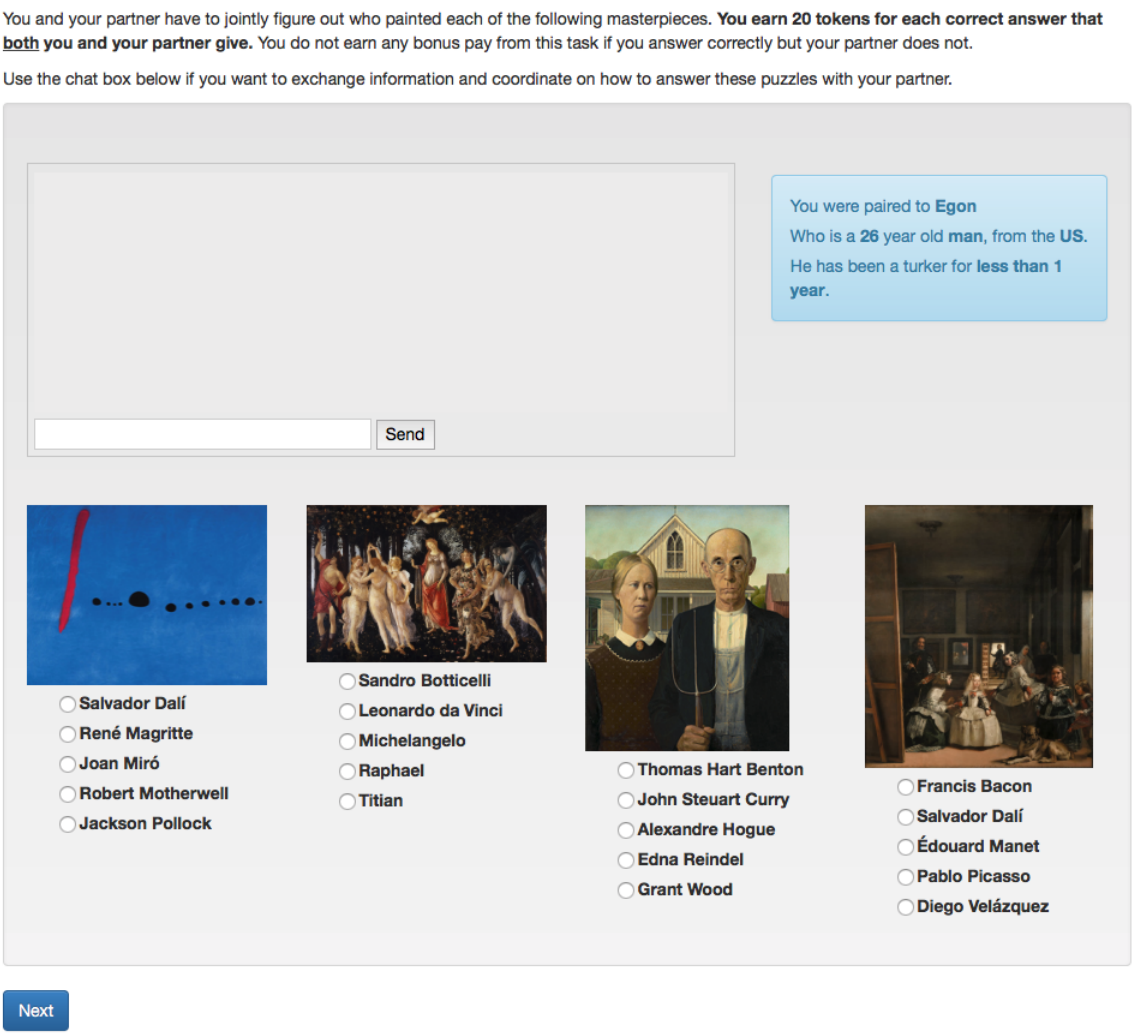


Figure B.3: Elicitation of the IOS (top) and WE (bottom) Scales

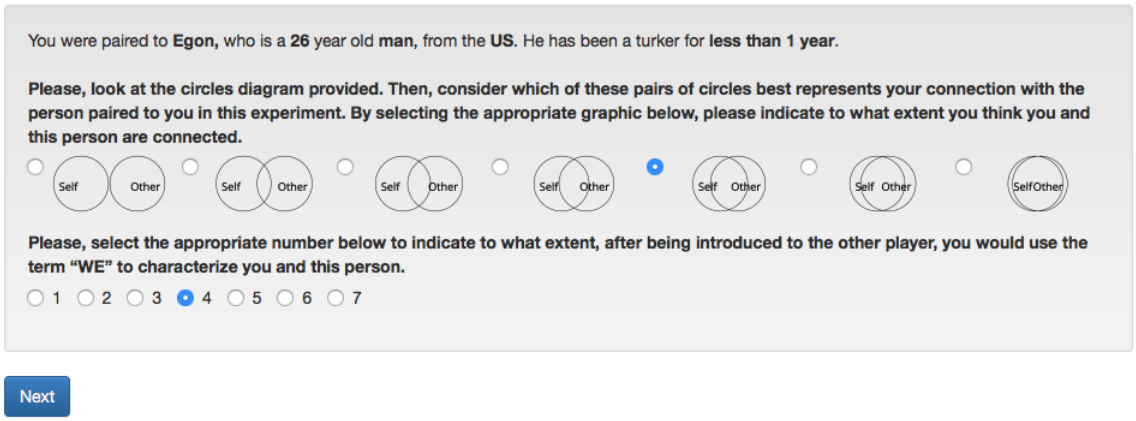


Figure B.4: Elicitation of Beliefs and Donations, and Treatment Assignment

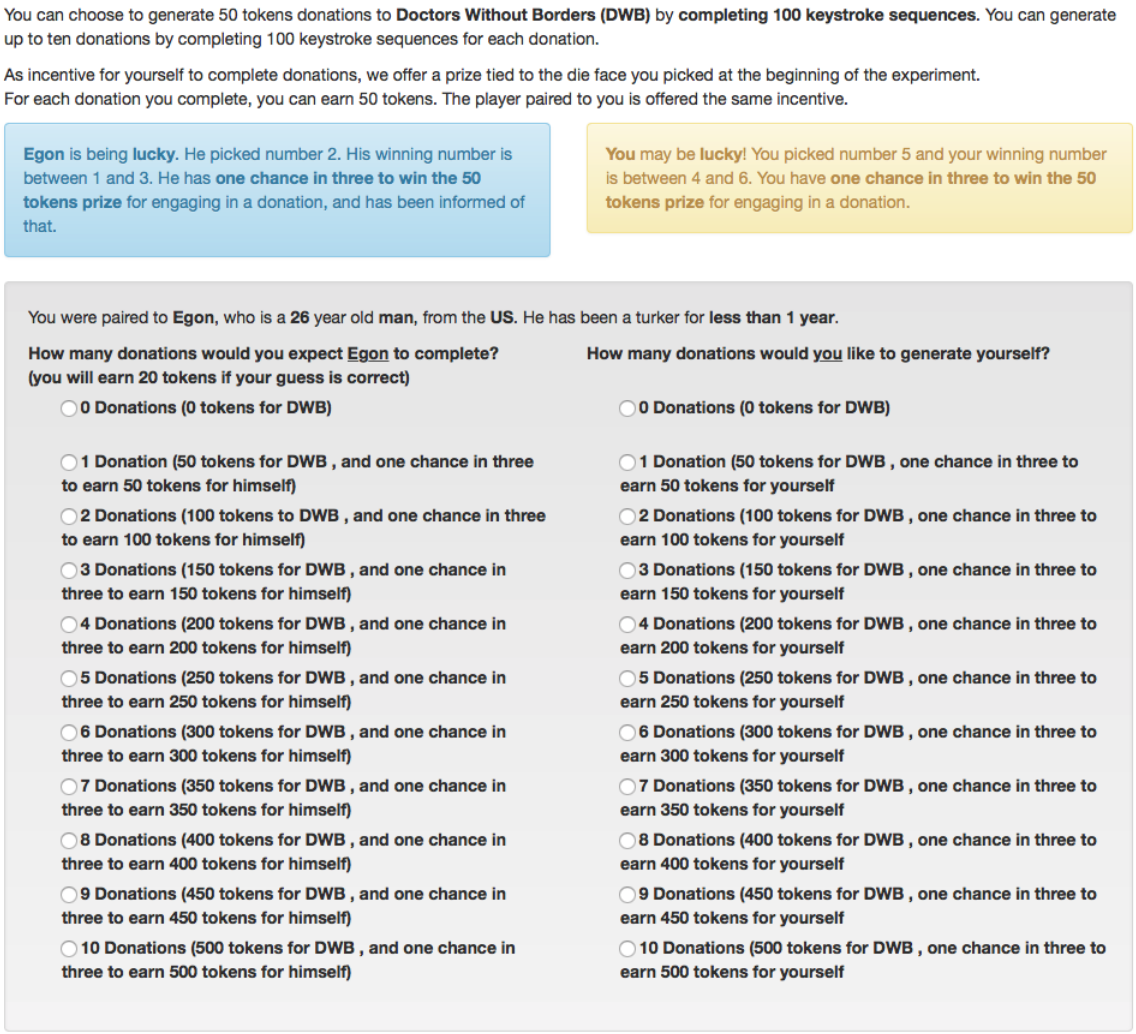
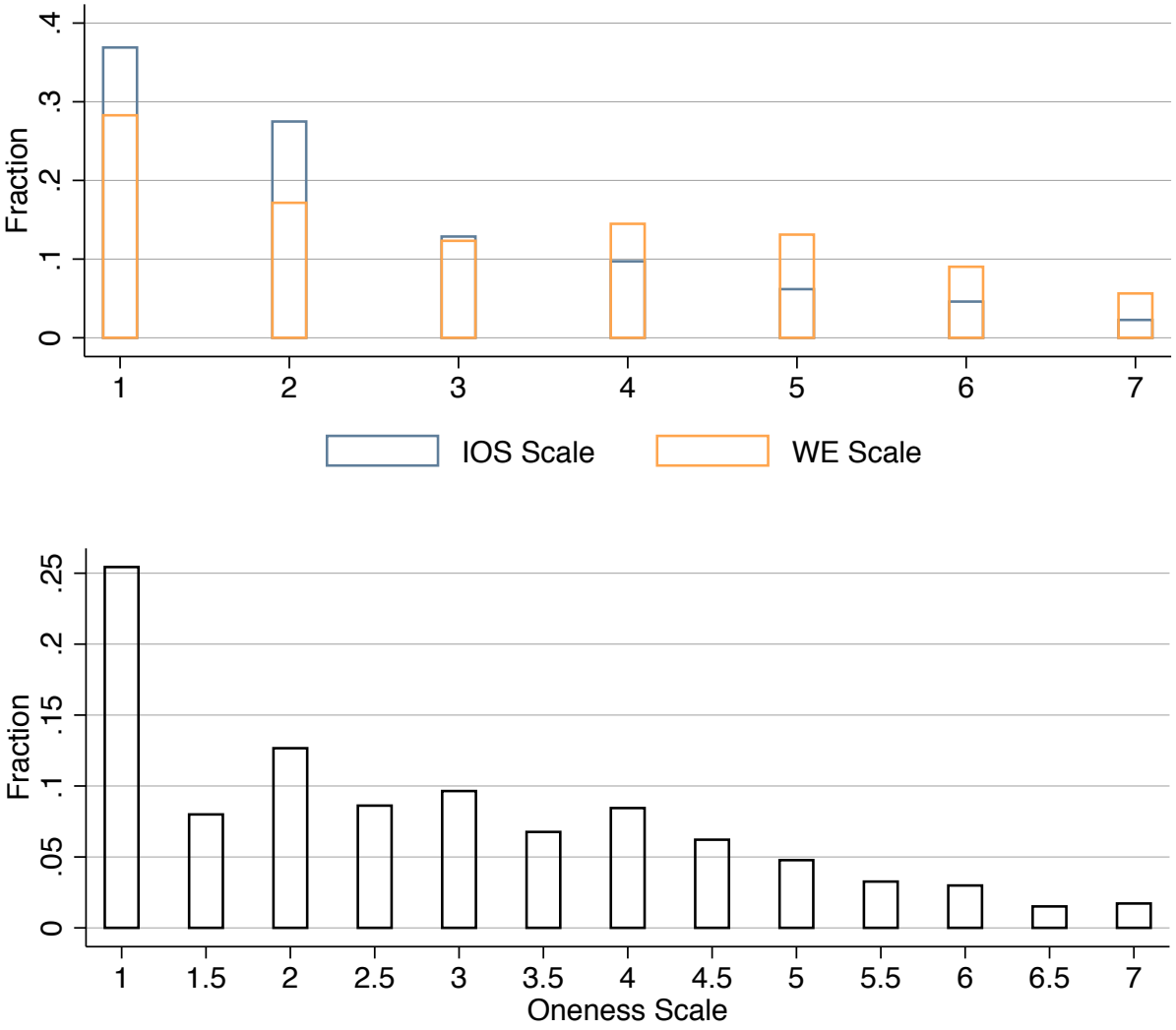


Figure B.5: Distribution of Social Proximity Scales



C. Complete Instructions

C.1. Page 0: Consent

Please read this before clicking "Accept"

This HIT is an academic experiment on economic decision making. Based on how you play the experiment, we will donate money to a charitable organization.

By participating in this experiment, you are participating in a study performed by researchers at the University of Bonn. All data collected in this study are for research purposes only.

The experiment requires you to press keys on your keyboard. You thus need full dexterity in at least one hand. The experimental software complies with modern web standards, but may require a physical keyboard to detect your keystrokes. For part of the experiment you will be interacting with another player. To ensure that interactions occur in a timely manner we give each participant 5 minutes maximum to complete each of the following two pages. For the rest of the experiment a session timeout applies. Your session expires 40 minutes after you accept this HIT. If you do not want to complete the HIT within 40 minutes, we advise to return the HIT. We will not be able to approve work for timed out HITs.

Compensation: After completing this HIT, you will receive your reward plus a bonus payment that is based on how you play the experiment.

Legal information: Your participation is voluntary. You may stop participating at any time by closing the browser window or the program to withdraw from the study. Any reports and presentations about the findings from this study will not include any information that could identify you. We may share the data we collect in this study with other researchers doing future studies; if we share your data, we will not include any information that could identify you. By accepting this HIT, you indicate that you are older than 18 years and agree to participate in this experiment.

C.2. Page 1: Introduction

In this study each participant will be given the opportunity to engage independently in a real effort game. By participating you create value for a charity.

Part of your variable bonus may be uncertain. For this we will ask you to pick a number between 1 and 6, which the experimental software will match to a digital roll of die.

What face of a die would you pick? [drop-down list]

For this experiment you will be paired to another player that is currently participating in the same experiment. Given that part of the experiment will involve common problem solving, we would like pairs of players to get to know each other. For this, on the next page we are collecting some basic socio-demographic information, which will be shared with the paired participant. The socio-demographic information collected is minimal and does not make you personally identifiable.

Throughout the experiment, you will engage in tasks that will determine your variable bonus. Completing tasks you accumulate tokens. Tokens are converted to USD at the end of the HIT. One token is worth 0.005 USD.

This experiment is a research effort to understand economic behavior. In what follows there will be no deception: we will do nothing different from what is explained to you. For any question do not esitate to contact us.

C.3. Page 2: Survery on Demographic Information

We would like paired players to know a bit about each other. For this, we are collecting some basic socio-demographic information, which will be shared with the other participant.

What is your first name? [text field]

What is your age? [drop-down list]

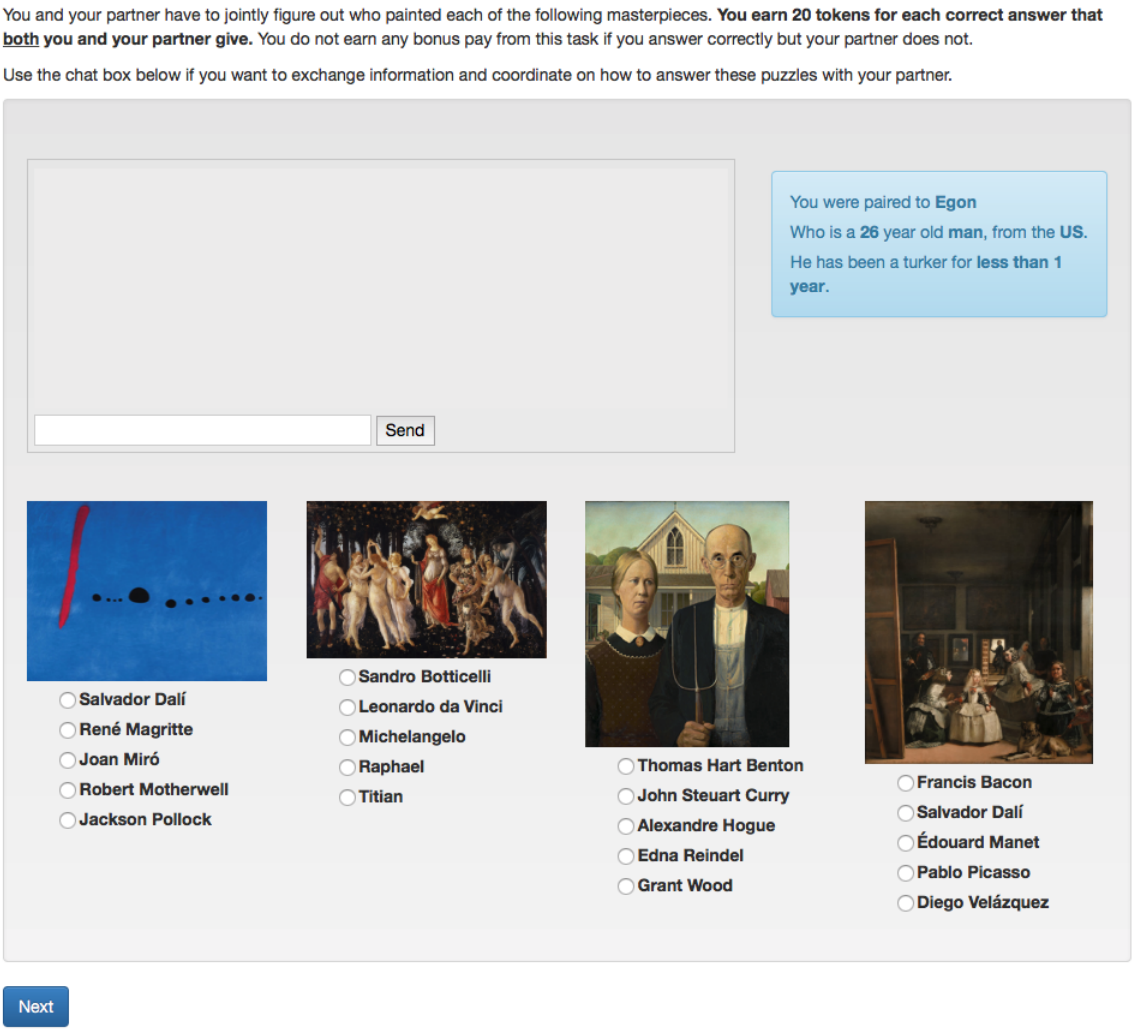
What is your gender? [drop-down list]

For how long have you been a turker? [drop-down list]

C.4. Page 3: Wait Page

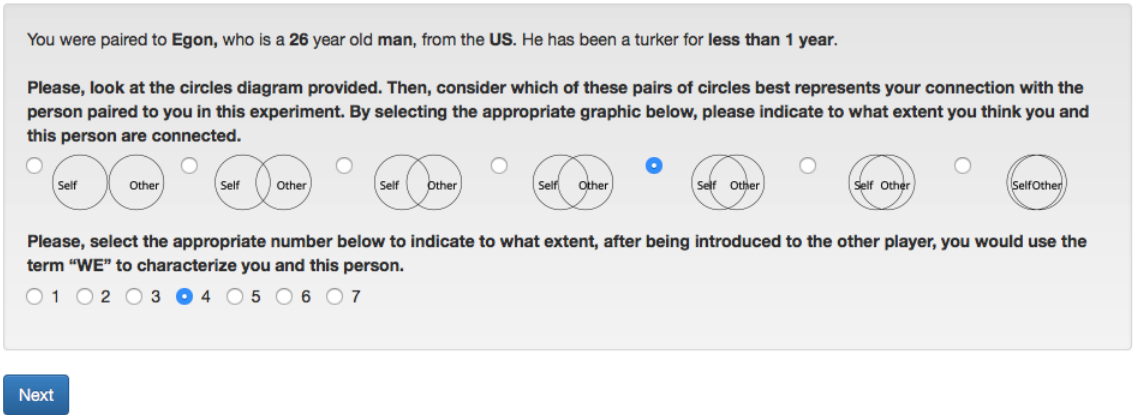
Please wait. Pairs are being formed.[34]

C.5. Page 4: Joint Problem Solving Task



[34] At this point of the experiment, each subject gets paired, randomly and anonymously, to another study participant.

C.6. Page 5: Oneness Elicitation



C.7. Page 6: Instructions for Donations

You will be able to engage in charitable giving by working on a simple assignment. Please carefully read the instructions below. Shortly, you will have the chance to familiarize yourself with this assignment in a training session. This will not affect your donation or payoffs. After the training, we will explain the payoffs for this task.

The assignment involves consecutively pressing the keys w e on your keyboard. You need to press the keys in this order. The keys are highlighted on the keyboard below. The software will display the number of successfully completed sequences.

You generate a donation to Doctors without Borders by completing a given number of sequences. A bar will indicate your progress towards this number.

In this example, you are asked to complete 100 keystroke sequences to generate a donation. Remember that this is just an example so that you can familiarize yourself with this assignment.

Please complete the training by pressing **w e** on your keyboard.

C.8. Page 7: Elicitation of Beliefs and Donations, and Treatment Assignment

You can choose to generate 50 tokens donations to Doctors Without Borders (DWB) by completing 100 keystroke sequences for each donation.

As incentive for yourself to complete donations, we offer a prize tied to the die face you picked at the beginning of the experiment. For each donation you complete, you can earn 50 tokens. The player paired to you is offered the same incentive.[35]

[Name_other_player] is being [lucky/unlucky]. [He/She] picked number [n]. [His/Her] winning number is between [1 and 3/4 and 6]. [He/She] has [no chance/one chance in three] to win the 50 tokens prize for engaging in a donation, and has been informed of that.[36]

[Name_other_player] picked number [n]. [He/She] has one chance in six to win the 50 tokens prize for engaging in a donation, and is aware of that.[37]

You may be [lucky/unlucky]. You picked number [m] and your winning number is between [1 and 3/4 and 6]. You have [no chance/one chance in three] to win the 50 tokens prize for engaging in a donation.[38]

You picked number [m]. You have one chance in six to win the 50 tokens prize for engaging in a donation.[39]

You were paired to [Name_other_player], who is a [age_other_player] year old[man/woman] from the US. [He/She] has been a turker for [less than 1 year/1 year/2 years/more than 2 years].

[35] Text displayed only if incentives were available.
[36] Text displayed only if other player's incentives were either *Zero* or *High*.
[37] Text displayed only if other player's incentives were *Moderate*.
[38] Text displayed only if personal incentives were either *Zero* or *High*.
[39] Text displayed only if personal incentives were *Moderate*.

How many donations would you expect [Name_other_player] to complete? (you will earn 20 tokens if your guess is correct)

○ 0 Donations (0 tokens for DWB)
○ 1 Donation (50 tokens for DWB, and one chance in [six/three] to earn 50 tokens for [him/her]self)
○ 2 Donations (100 tokens for DWB, and one chance in [six/three] to earn 100 tokens for [him/her]self)
○ 3 Donations (150 tokens for DWB, and one chance in [six/three] to earn 150 tokens for [him/her]self)
○ 4 Donations (200 tokens for DWB, and one chance in [six/three] to earn 200 tokens for [him/her]self)
○ 5 Donations (250 tokens for DWB, and one chance in [six/three] to earn 250 tokens for [him/her]self)
○ 6 Donations (300 tokens for DWB, and one chance in [six/three] to earn 300 tokens for [him/her]self)
○ 7 Donations (350 tokens for DWB, and one chance in [six/three] to earn 350 tokens for [him/her]self)
○ 8 Donations (400 tokens for DWB, and one chance in [six/three] to earn 400 tokens for [him/her]self)
○ 9 Donations (450 tokens for DWB, and one chance in [six/three] to earn 450 tokens for [him/her]self)
○ 10 Donations (500 tokens for DWB, and one chance in [six/three] to earn 500 tokens for [him/her]self)

How many donations would you like to generate yourself?

○ 0 Donations (0 tokens for DWB)
○ 1 Donation (50 tokens for DWB, and one chance in [six/three] to earn 50 tokens for yourself)
○ 2 Donations (100 tokens for DWB, and one chance in [six/three] to earn 100 tokens for yourself)
○ 3 Donations (150 tokens for DWB, and one chance in [six/three] to earn 150 tokens for yourself)
○ 4 Donations (200 tokens for DWB, and one chance in [six/three] to earn 200 tokens for yourself)
○ 5 Donations (250 tokens for DWB, and one chance in [six/three] to earn 250 tokens for yourself)
○ 6 Donations (300 tokens for DWB, and one chance in [six/three] to earn 300 tokens for yourself)
○ 7 Donations (350 tokens for DWB, and one chance in [six/three] to earn 350 tokens for yourself)
○ 8 Donations (400 tokens for DWB, and one chance in [six/three] to earn 400 tokens for yourself)
○ 9 Donations (450 tokens for DWB, and one chance in [six/three] to earn 450 tokens for yourself)
○ 10 Donations (500 tokens for DWB, and one chance in [six/three] to earn 500 tokens for yourself[a])

[a] Text displayed only if private incentives are available with positive ex-interim probability.

C.9. Page 8: Donation Task

You have chosen to make [D] donations. For this you will have to complete [D×100] keystroke sequences to generate these donations

Please complete the donation to Doctors without Borders by pressing w e on your keyboard.

C.10. Page 9: Short Questionnaire

Thank you for completing the donation task. Please fill out the short questionnaire below and then go to the next page to review payoffs and complete the HIT.

You and your partner could earn 20 tokens for guessing correctly how many donations the other did. Aside from the guessing question, was it clear to you that the number of donations that YOU made was not directly affecting the payoff of the other player? [Yes/No]

You and your partner could earn 20 tokens for guessing correctly how many donations the other did. Aside from the guessing question, was it clear to you that the number of donations that the OTHER made was not directly affecting your payoff? [Yes/No]

Did you realize that the amount donated to charity was increasing in the donations that both you and the other player made? [Yes/No]

In choosing how many donations to make, were you influenced by the thought of the number of donations the other person was going to make? [Yes/No]

Expecting that the other person could make more donations, makes you want to donate [More/Less/Indifferent]

In other contexts, when you are about to make a charitable donation, do you ever consider whether and how much other people are contributing to the same cause? [Always/Very often/Sometimes/Rarely/Never]

In other contexts, when you are about to make a charitable donation, expecting that other people could make more donations, makes you want to donate [More/-Less/Indifferent]

Please recall the screen where you chose how many donations to make. What

were the chances YOU had to win the lottery for participating in the donation task? [No chances/One chance in six/One chance in three/Cannot recall]

Please recall the screen where you chose how many donations to make. What were the chances the OTHER player had to win the lottery for participating in the donation task? [No chances/One chance in six/One chance in three/Cannot recall][40]

[40] Questions displayed only if incentives were available.