DISCUSSION PAPER SERIES

DP13078

SOCIAL INFLUENCE IN PROSOCIAL BEHAVIOR:EVIDENCE FROM A LARGE-SCALE EXPERIMENT

Lorenz Goette and Egon Tripodi

PUBLIC ECONOMICS

ISSN 0265-8003

SOCIAL INFLUENCE IN PROSOCIAL BEHAVIOR:EVIDENCE FROM A LARGE-SCALE EXPERIMENT

Lorenz Goette and Egon Tripodi

Discussion Paper DP13078
Published 25 July 2018
Submitted 25 July 2018

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

SOCIAL INFLUENCE IN PROSOCIAL BEHAVIOR:EVIDENCE FROM A LARGE-SCALE EXPERIMENT

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

JEL Classification: N/A

Keywords: prosocial behavior, social influence, Online Experiment

Lorenz Goette - lorenz.goette@uni-bonn.de
University of Bonn and CEPR

Egon Tripodi - egon.tripodi@eui.eu
European University Institute

Acknowledgements
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 for a helpful discussion at SEYE 2018. Research funding from the University of Bonn is gratefully acknowledged. Carina Lenze and Luis Wardenbach provided excellent research assistance.

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

Lorenz Goette Egon Tripodi

This version: July 3, 2018

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.

Keywords: Prosocial behavior, social influence, online experiment.

JEL Classification:

*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 for a helpful discussion at SEYE 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 peers in many day-to-day decisions. A growing literature illustrates how social information propagates social influence – the effect of the observed (or expected) actions of a relevant social reference group on individual behavior – both in online and in offline interactions.[1] People vote because they observe their friends vote (Bond et al., 2012), they exercise because they observe their friends exercise (Aral and Nicolaides, 2017), and they donate to charity because their university colleagues (Frey and Meier, 2004) or co-workers (Kessler, 2017) are observed to donate. A key feature of these studies is that individuals observe others' behavior. A common theme to many studies in this literature is that social influence can arise from both people's ability to learn about the environment from others' behavior and from a genuine preference for mimicking what others do. Disentangling the two mechanisms typically remains an important open question. Various models (e.g. Bernheim (1994); Akerlof (1997)) can explain that social influence spreads even for unobservable behavior, but to study this second mechanism of social influence it is necessary to rule out the learning opportunities that derive from observing others.

In this project, we study how social influence can affect prosocial behavior 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.[2] This study investigates whether information about the economic environment can be sufficient to propagate social influence. One of the main goals of this venture is to understand if and how 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 the spread of 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 –

[1] From the recent empirical literature, we notice that social influence plays an important role across a broad set of domains that include charitable giving (Lacetera et al., 2016; Kessler, 2017) marketing (Aral and Walker, 2011, 2014; Bapna and Umyarov, 2015), political participation (Bond et al., 2012; Cantoni et al., 2017), and well-being (Aral and Nicolaides, 2017).

[2] 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; Warr, 1982; Roberts, 1984) and impure altruism (Andreoni, 1989, 1990).

as a particular mechanism of social influence. *Conformists* are thought to 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 actions are unobservable. The empirical investigation of conformity poses at least three important methodological challenges. First, preventing that the behavior of another person gives the opportunity to update either the perceived social norm or the perceived valence of certain actions. Second, identifying what is the relevant social reference whose behavior is taken as a norm to which agents try to adhere. Third, introducing exogenous variation in the (expected) behavior of a social reference. We propose an experimental design that overcomes all three such challenges. To clarify our approach, we illustrate theoretically how heterogeneous monetary incentives to act prosocially can be used to experimentally identify the conformity channel of social influence. Our rich experimental design also provides the tools to shed light on some of the key mitigating factors of social influence.

In a large online experiment, 3,467 individuals engage in pair-wise 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 also 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 (*none*, *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 Bob cares to adhere to the behavior of Abigail, an increase in the Abigail's incentives should affect not only Abigails's donations but also Bob's. 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 were unaffected. Before the treatment manipulation, pairs of subject participate in a joint problem solving task, which we adopt to induce social proximity between paired players (Chen and Li, 2009)

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 how social influence differentially propagates among more and less closely connected partners.

In our conceptual framework of conformist social influence, individual charitable contributions are predicted to be 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 strategic complementarities in donations: when partner's incentives increase from *none* 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: the effect on donation of increasing partner's incentives from *none* to *moderate* is as large as about half the size of the effect of increasing *their own* private incentives from *none* 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. The socially more distant half of the sample is instead neutral to changes in partner's incentives. Importantly, we show that differences in response to partner's incentives are not due to differences in altruism towards the charity: in a control treatment with no incentives charitable contributions are unrelated to individual social proximity towards the partner.

An important challenge in the interpretation of the evidence is that heterogeneous incentives may damage donors' morale and confound conformity with inequity aversion. We illustrate the predictions of a conceptual framework that incorporates morale effects of incentive inequality (Breza et al., 2017). These predictions are consistent with the behavior of subjects that exhibit a weak connection (below median oneness) to their partner, but fail to capture patterns observed among subjects with a strong connection (above median oneness) to their partner. Socially close individuals (i) increase donations when higher incentives to the partner increase incentive inequality and (ii) decreases donations when higher incentives to the partner decrease incentive inequality. These subjects expect virtually identical donation levels from partners with *moderate* and *high* incentives, but compared to baseline they increase donations by 0.213 ($p = 0.001$) standard deviations when partners get *moderate* incentives and

increase donations by just 0.043 ($p = 0.479$) standard deviations when partners get *high* incentives. Having ruled out that these patterns can be explained by incentive inequality, we interpret our results as evidence of reduced norm adherence due to the ineffectiveness of larger incentives, which are not expected to affect partner's behavior, but necessarily make partner's motive to donate more self-serving. As a result, when the partner receives *high* incentives, the individual may reduce the importance attached to conformism, similar to Fuster and Meier (2009).

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; 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). Perhaps our key contribution to this literature is to show that, even when behavior is not observable, people subject to social influence try to infer how others behave and mimic their behavior.

The strategic complementarities of social influence are often attributed to behavioral mechanisms that include social learning (Vesterlund, 2003), joint consumption (Bruhin et al., 2015), reciprocity (Rabin, 1993), and conformity (Bernheim, 1994). To this literature, concerned with understanding the behavioral mechanisms of social influence, we make an empirical contribution by providing evidence of the empirical relevance of conformity and showing how social proximity and the economic environment shape social influence.

Compared to existing literature (Gneezy and Rustichini, 2000a,b; Fuster and Meier, 2009), our evidence enriches 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 be binary. In fact, we find that the *size* of incentives matters.[3] We also add, to an empirical literature documenting the role of social 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 completely absent.

[3]The theory of *ostracism* from Dutta et al. (2018) predicts that sufficiently large incentives are needed for norm adherence to collapse. Although ostracism is very unlikely to play a role in our environment, our evidence is consistent with this theory.

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, we develop an experimental design that makes, to the best of our knowledge, the first attempt to separately identify conformity from social learning; second, we shed light on some of the modulating factors of social influence that are relevant for theory and for the design of institutions and future studies.

The remainder of the paper is structured as follows: 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

The experiment is conducted online with people registered as workers on Amazon Mechanical Turk. The experimental design 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 the five steps in detail.

1. Registration. Invited subjects access our software interface. They read some general instructions in which we only provide summary information about 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 these 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 they are told will be shared only with their partner. [4]

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.[5] 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. For this task, pairs of players are faced with four famous paintings. Next to each painting are the names of four artists, and pairs of subjects are paid 20 tokens for each correct name on which they manage to agree.[6] Paired players communicate through a private online chat (see interface in Figure B.3) . Differently from Chen and Li (2009), to further encourage contact, we require that *both* players in a pair guess each artist correctly for correct answers to be valid for bonus pay.[7] Payoffs for this task are not revealed until the end of the experiment.

3. Oneness elicitation. Oneness represents our measure of social proximity for this study. This measure was first proposed by Cialdini et al. (1997) and then validated by Gächter et al. (2015) as a reliable measure of the subjective degree of closeness to another person. The oneness score is a simple mean of two underlying scores: (i) the IOS scale and the (ii) WE scale.[8] Each of these scales ranges from 1 to 7 (with 1 indicating low social proximity, and 7 high social proximity). 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 a good proxy of beliefs

[4]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.

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

[6]Admittedly, we make the task quite hard by listing after each painting artists who are mostly contemporaries and share a relatively similar style.

[7]Subjects cannot skip the task, but they are free to solve the puzzles without making contact with their partner.

[8]See Figure B.4 for an illustration of the elicitation method of these scales.

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 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 a fair die roll. 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 feedback is provided, depending on the initial number chosen, this may reduce to zero the chances of getting the piece-rate incentive for generating donations (incentives are *none*), or increase chances to 1 in 3 (incentives are *high*). When this feedback 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*). Incentives of the two players in each pair are common knowledge for both players. To make sure that subjects understand both their own and the other player's incentives correctly we directly provide them with the updated probabilities of getting paid for generating donations (see Figure B.5 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.

[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, n.d.).

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 serving as comprehension checks. A review of the individual payoffs concludes the experiment.

Figure 1: Overview of Experimental Design and Treatment Assignment

Completed registration, joint problem solving, and *oneness* elicitation

Lottery treatments
T1-T9

Incentives to self \ Incentives to other	None $P_j = 0$	Moderate $P_j = \frac{1}{6}$	High $P_j = \frac{1}{3}$
None $P_i = 0$	T1	T2	T3
Moderate $P_i = \frac{1}{6}$	T4	T5	T6
High $P_i = \frac{1}{3}$	T7	T8	T9

No lottery control
T10

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 prospected 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 $v(.)$ that is convex and monotonically increasing in the gap 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 sit-

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

uation, 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 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).[14] Like in most models of prosocial behavior, donations are strategic substitutes due to the concavity of $u(.)$. 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 $u(.)$ to be linear in the social benefit of a donation A, the cost of effort $c(.)$ and the conformity utility $v(.)$ to be quadratic.[16]

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

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

[13] **?** 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] The notion that stronger ties are more relevant for social influence is also common in more sophisticated models of conformism (Patacchini and Zenou, 2012) and status seeking (Immorlica et al., 2017) in social networks.

[15] Super-modularity of the utility function, namely $\frac{\partial U(.)}{\partial d_i \partial d_j} = u''() + \lambda_i \tau_{i,j} v''(.) > 0$, defines strategic complementarity. Given concavity of $u(.)$, a necessary condition for supermodularity 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 $v(.)$ is typically assumed to be either an absolute value (Bernheim, 1994), or quadratic (Akerlof, 1997). 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.

$$d_i^* = [A + m_i(c + 2\lambda\tau) + m_j 2\lambda\tau](c + 4\lambda\tau)^{-2}$$

that leads to the following prediction.

Prediction 2.1 (Conformity)**.** *Conformity makes donations strategic complements. 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. 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 feature the strategic complementarities of conformity, changes in j's incentives produce positively correlated effects on these two outcome variables.

Not all agents are expected to conform to their social reference to the same degree. This framework assumes that social proximity, between the agent whose incentives are affected and the social reference, modulates conformity, and that individuals have a potentially heterogeneous intrinsic preference to adhere to social norms of behavior. This framework does not say whether, for an individual, 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. The extent to which one wishes to adhere to the behavior of a social reference λ may 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.

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) in fact shows that, in a work environment, unjustifiably heterogeneous incentives can introduce a form of inequity aversion (Fehr and Schmidt, 1999) that damages morale to exert effort.

In the setup of a prosocial activity, randomly determined heterogeneities in incentives could be detrimental for one's morale to generate donations for a charity. We explore the predictions of a simple model of prosocial behavior like (2.2), in which the morale utility term from Breza et al. (2017) replaces the conformity term.

$$U(d_i) = (d_i + d_j)A + m_i d_i - c d_i^2 + M(m_i, m_j) d_i \tag{2.3}$$

Morale $M(.)$, 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.

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

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

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

Prediction 2.2 (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 should be monotonically decreasing in the size of the $m_i - m_j$ gap.

Part (ii) of prediction 2.2 indicates that incentive inequality can make i's and j's donations strategic complements. However, notice that incentive inequality may also make donations strategic substitutes: 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 – making donations strategic substitutes 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 where thousands of registered workers are commonly employed in tasks that require human intelligence, which is becoming increasingly popular for conducting economic experiments (DellaVigna and Pope, 2016). Some people on this platform queue up well paying tasks that they plan to do later during the day: a 90 seconds time-out in the registration phase of the experiment screens out people that begin the experiment without making progress. 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, 2016).[18] In our experiment, subjects that complete the study earn 1.20 USD participation fee plus bonus pay. Tokens constitute the experimental currency at the exchange rate of 1 token=0.005 USD. Completing the experiment took participants 17.06 minutes 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*. Participation in the experiment is allowed only once, and no retakes are allowed for subjects that accidentally drop out of the study.

[17] We collect data over more than one session to minimize the risks that server overload could cause the app that we developed using oTree (Chen et al., 2016) to crash.

[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.

3. Experimental Results

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 this task, we measure an average individual oneness score towards the partner of 2.8 (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.

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

In this task, subjects have four paintings and they need to agree on the correct artist to associate from a list of five artists for each painting. Through a chat box, subjects can make contact to discuss the correct answers and strategies to solve the task.[20] The fact that subjects on average score 40 of the 80 available points indicates significant coordinated effort to solve the common puzzles: random click-through from both subjects only grants 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

[19] Studying behavioral mechanisms that are modulated by social interactions is methodologically complex. Some papers in the literature leverage existing social relationships and identities, while others have induced the formation of social relationships and identities within the experiment (Goette et al. (2012) and Chen et al. (2014) include reviews of this literature). For our investigation, to avoid contaminating the conformity channel of social influence with other forms of social influence deriving from the prospects of future interactions, we choose to take the approach of building social relationships among randomly and anonymously matched strangers.

[20] To look up the paintings they do not know, many of the subjects realize that they can use Google image search, and they tend to split up paintings to search with their partner.

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			None			Moderate			High			
Incentives to other			None	Moderate	High	None	Moderate	High	None	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.

information and contact with the other player in the JPS task facilitate the formation of perceived social proximity.

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. Remember that the oneness scale ranges from 1 to 7 with possible realization every half unit. In other studies, 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 interquartile 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. Predictably, a large share of subjects exhibits no social proximity to a stranger partner. More interestingly, notice that at least half of the sample exhibits social proximity towards a stranger, with whom they have recently made contact to solve puzzles, similar to social proximity that other studies observe towards acquaintances. This is an indication of the role played by contact opportunities in the JPS for harnessing social proximity.

Table 2: 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.

In Table 2, 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 Amazon Mechanical Turk 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.[21] 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)) between the regression residuals of the two players in each pair is an indication of the role played in this experiment by the opportunity to make contact for the development of a social connection.

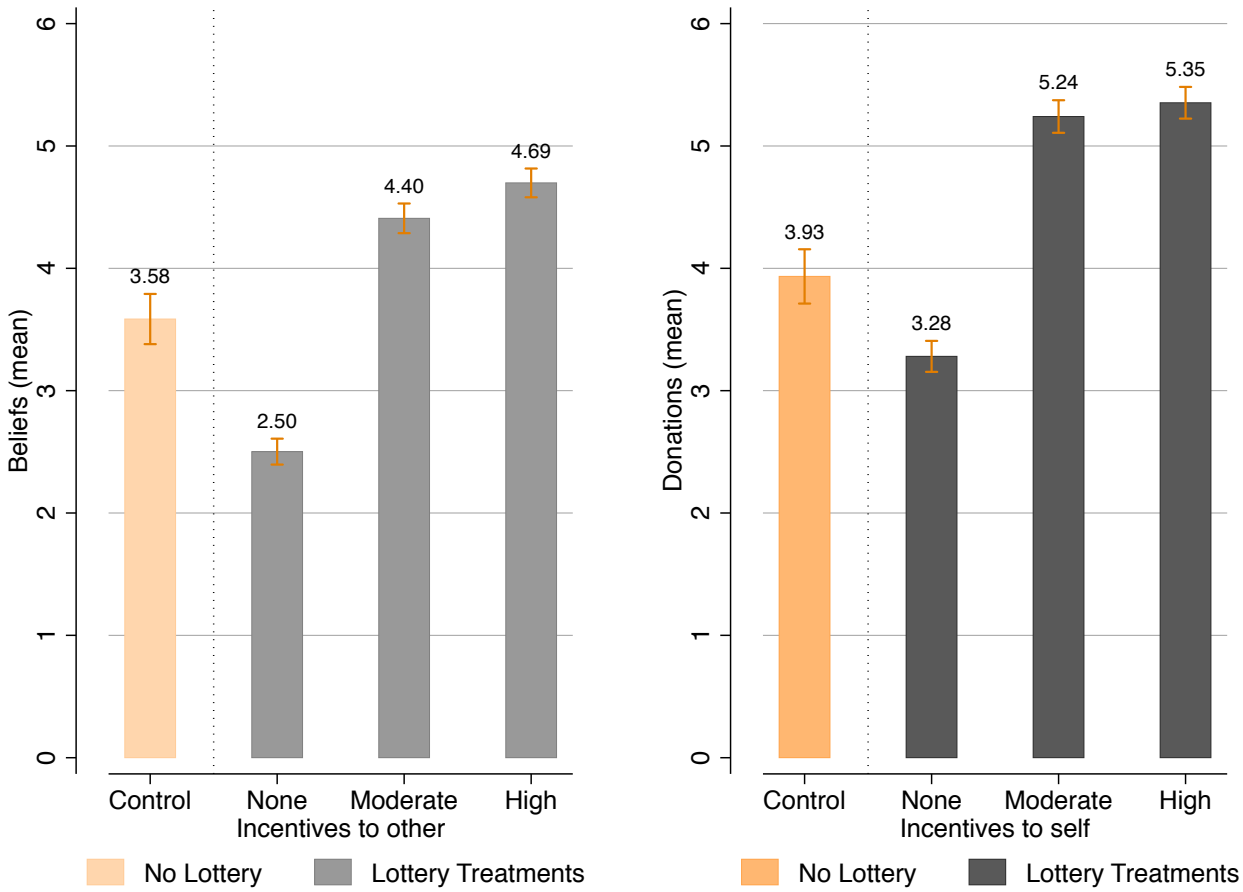
3.2. Beliefs and Donations

Before discussing social influence, at least four points are worth illustrating about the direct (and predicted direct) effects of monetary incentives on donations (Figure 2). First, we notice that beliefs pretty accurately capture the comparative statics of how partners react to incentives, but underestimate the generosity of other.[22] Second, we find a predictable *disappointment effect*: offering lottery incentives while providing information to the subject of an unlucky lottery draw (incentives are *none*) frustrates donations compared to the payoff equivalent *no lottery* control treatment in which subjects were not told that they might be provided incentives (rank-sum test: $p < 0.001$). Subjects also internalize the *disappointment effect* when they express their belief on the donation behavior of their partner. Third, information that doubles the expected piece rate for generating donations (i.e. from *moderate* to *high*) has no significant effect on donations (rank-sum test: $p = 0.373$). Fourth, the moderate incentive alone, leads to 1.31 additional donations compared to the *no lottery* treatment. This result is quantitatively

[21] 80.4% of subjects used the chatbox to get in touch with their partner, and 64.6% of pairs managed to have a conversation (*contact* = 1). In these conversations, subject share their knowledge about the four paintings, share relevant personal information and considerations (for example, 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 jointly 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.

[22] This evidence is consistent with previous 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)).

Figure 2: Direct Incentive Effects



Notes: 2914 observations. Each of the gray bars pools the three treatments that maintain constant either partner's incentives (left panel) or player's incentives (right panel). Error bars indicate standard errors of the mean.

relevant because it indicates a crowding-in of donations: providing an incentive to volunteer time and efforts resulted in more resources transferred to charity than were spent to incentivize participants.[23] This result adds to an existing literature that documents the effectiveness of monetary incentives to act prosocially (Landry et al., 2006; Karlan and List, 2007). We find that lottery incentives can *crowd-in* donations. Taken together, these descriptive findings can be summarized in the following result:

Result 1 (Incentive Effects):

[23]Back of the envelope calculations help illustrate that payoffs to the charity were calibrated such that one hour of time volunteered for the charity would generate about 10 USD for the charity and an expected monetary pay to participant of 1.67 USD. In fact, each donation required an average of 86.81 seconds and generated 0.25 USD for the charity for sure and 0.25 USD with probability $\frac{1}{6}$ in the moderate incentive treatments.

- *The availability of incentives to act prosocially increases charitable efforts, but doubling the stakes of incentives does not increase donations significantly.*
- *People that are told they may receive an incentive to donate, but are randomized out of incentives, decrease donations.*
- *The crowd predicts correctly the comparative statics of incentive effects, but systematically underestimates partners' prosocial behavior.*

3.3. Conformity as Social Influence

We now move to considering how the economic environment faced by the partner affects individual donations. Table 3 summarizes beliefs and donations across treatments and highlights the non-monotonic pattern that we will try to explain in this section. Beliefs will turn out very useful to understand this pattern. As previously noted, subjects expect their partner to monotonically increase donations with incentives. Turning to donations, besides the main incentive effects, notice how partner's incentives affect individual donations: systematically, individual donations *increase* when incentives of the partner go from none to moderate and *decrease* when incentives of the partner go further up–from moderate to high. Increases in donations due to moderate incentives for the partner come along with strong increases in expected donations from the partner, but also drops in donations due to high incentives to the partner are associated with increased expected donations from the partner. In the rest 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.

Table 3: Donations and Beliefs about Partner's Donations Across Treatments (Means and Standard Errors)

		Beliefs			Donations		
Incentives offered							
No (control)		3.585			3.934		
		(0.205)			(0.222)		
Yes (3x3 treatments)							
		Incentives to other			Incentives to other		
		None	Moderate	High	None	Moderate	High
Incentives to self	None	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)

As a first step, we estimate in a regression framework the effect of changes in partner's (j's) incentives on self's (i's) donations. As a second step, we also estimate the effect of changes in j's incentives on i's beliefs about j's donations. Assessing how j's incentives affect i's beliefs over j's behavior and i's behavior allows to evaluate the predictions of the conformity model and identify the conditions under which donation behavior reveals the strategic complementarities of social influence.

For the first step, we use regression equation (3.1) to estimate the indirect effect of j's incentives on i's donations. This regression model includes treatment indicators for the direct effects of i's incentives on i's donations and indicators for indirect effects of j's incentives on i's donations, exploits the *no lottery* control treatment to isolate the *disappointment effect* discussed above, and controls for observable characteristics of both players in the pair. Models of (im)pure altruistic giving predict that β_4 and β_5 in this regression – the effects of j's incentives on i's donations – should be non-positive. As illustrated in prediction 2.1 instead, positive values of β_4 and β_5 are consistent with the conformity mechanism of social influence.

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

The second step takes us closer to directly identifying the strategic complementarities of social influence, by additionally estimating the mirror least squares regression

(3.2) for individual predictions of the donation of the social reference. If donations exhibit strategic complementarity, we should observe expected increases in donations of the social reference j in response to j's incentives (δ_1 and δ_2) to move along with increases in i's donations.

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

The estimation of regression models (3.1) and (3.2) is presented in panels (a) and (b) of Table 4, respectively. The regression framework replicates the direct incentive effects presented in the previous section. The control *no lottery* treatment isolates the *disappointment effect* of being offered a lottery incentive and being randomized out of incentives, which is estimated to decrease donations by 0.831 units (column (2)). Positive expected pay for generating donations increases donations by roughly two units, with no significant difference between the moderate expected pay and the high expected pay. Beliefs about donations of the partner increase monotonically in her incentives to donate.

Consider now the effects of partner's incentives on beliefs and donations. Compared to no incentives baseline, when the partner receives moderate incentives, we find that both beliefs and donations increase by 1.962 ($p < 0.001$) and 0.356 ($p = 0.055$) units, respectively. 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 unaffected by 0.357 ($p = 0.046$) units. Below we summarize and interpret these effects.

Result 2 (Effects of Incentives to the Social Reference):

For individuals in our sample, all else equal,

1. *increasing incentives for their partner to donate from none to moderate leads to an increase in both their donations and their expectations about their partner's donations;*

2. *further increasing incentives for their partner from moderate to high, leads to a decrease in their donations, and an increase in expectations about partner's donations.*

To interpret this result, remember that in a standard model of prosocial behavior, with concave (pure or impure) altruistic utility term $u(.)$, incentives are expected to have a positive direct effect on individual donations. When incentives to donate are heterogeneous, altruistic donors give *less* if others are expected to give *more*. We instead find, in line with our conformity framework of social influence, that the introduction of moderate incentives shifts both belief about partner's donations and donations of the player whose incentives are not affected in the same direction.

A further increase in partner's incentives produces effects that are harder to interpret. If we ignore the effect of moderate incentives to the partner, the evidence that high incentives to the partner lead to increased beliefs about partners' giving but do not affect individual donations is consistent with a model in which the marginal benefit of giving is not a function of partner's giving – this could be a model of impure altruism where all the variation in donation behavior comes from the warm-glow utility term and pure altruism has no bite. However, warm-glow giving alone would not explain that donations increase when incentives to the partner decrease from high to moderate.

Through the lenses of our conformity framework, there are two natural ways of interpreting the second part of Result 2. The first is that because the strategic substitution of (im)purely altruistic preferences and the strategic complementarities of conformity can pull donations in opposite directions, it is possible that at moderate incentives to the partner strategic complementarities happen to dominate while at high incentives to the partner strategic substitutions dominate. The second plausible interpretation 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. We can use our theoretical framework to guide the analysis that can shed light on this dilemma.

Our theoretical framework predicts that the behavior of a closer social reference should have a stronger social influence effect on one's donations. Therefore, we can hope to gain better interpretation of our data by investigating how social influence spreads among people with high and low social proximity.

We partition the sample at the median score of our measure of social proximity. This partition of the sample has the realized advantage that subjects in the two groups

exhibit significant differences in their social connection to their partner, but do not systematically differ in baseline levels of prosociality. Average oneness in the bottom half of the oneness distribution of our sample is at 1.4 – very close to the lower bound of the spectrum and far below oneness measured in the literature towards acquaintances, while average oneness in the top half of the oneness distribution is at 4.0 – similar to oneness measured towards a non-close friend (Gächter et al., 2015). Using the control treatment, we can 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 ($p = 0.861$).

Confident that differences in the reaction to incentives among subjects with high and low social proximity to their partner are due to their social connections, we estimate regression equations (3.1) and (3.2) for each of these two sample partitions.

Consistent with Prediction 2.1, columns (3) and (4) of Table 4 show that social influence is stronger for subjects with higher social proximity to the partner. The behavior of subjects with low social proximity to their partner follows the standard model of prosocial behavior: these 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. If anything, 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 neither of these point estimates are 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 none 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. As in the full sample, when partner's incentives further increase from moderate to high donations *decrease* by 0.667 ($p = 0.007$) units. Differently from what we observe in the full sample however, increasing partner's incentives from moderate to high cause no significant shift in beliefs about her donations ($p = 0.750$). Below we summarize these results and discuss the interpretation.

Result 3 (Social Proximity Shapes the Effects of Incentives to the Social Reference): *For individuals that are strongly connected to their partner, all else equal,*

(a) increasing incentives for the partner to donate from none to moderate leads to an increase in both their expectations about partner's donations and their own donations

(b) *further increasing incentives for the partner from moderate to high does not affect their expectations about partner's donations, but leads to a drop in their own donations.*

For individuals that are weakly connected to their partner, holding their incentives constant,

(c) *increasing incentives for the partner to donate monotonically shifts expectations about partner's donations upward, but their donations are not affected.*

From our theoretical framework, we should expect the strategic complementarities of social influence to be mitigated by weakness of the social connection to the social reference. Consistent with our theoretical framework, we observe patterns evoking social influence for individuals that are socially close to their partner but not for those who are socially distant. Individuals with a weak social connection to their partner do not increase donations when the partner is expected to increase donations in response to incentives, and a model of concave altruistic utility is sufficient to explain their behavior. Besides the fact that the theoretical framework proposed can explain the differential effects of partner's incentives presented in Result 3, this result is important because it allows us to make progress in shedding light on the dilemma left open in the interpretation of Result 2. For the full sample we said that our theoretical setup offers two possible interpretations for the drop in individual giving when partner's incentives increase from moderate to high that cannot be distinguished. The first is that because the strategic substitution of concave altruistic preferences and the strategic complementarities of conformity can pull donations in opposite directions, strategic complementarity happens to dominate at moderate incentives while at high incentives to the partner strategic substitutions dominate. The second is that norm adherence (λ) is weakened when the behavior of a social reference is more strongly influenced by monetary self interest. Through part (b) of Result 3, we can rule out the first explanation in favor of the second one. In fact, because beliefs about the donations of socially close partners are nearly identical when partners get moderate and when partners get high incentives, substitution due to a lower marginal benefit from altruistic giving utility cannot explain why people give less when their partner's incentives increase. Our theoretical framework leads us to conclude that the strategic complementarities of social influence, in particular the extent to which people care to conform to the behavior of their partner (norm adherence), are malleable to the economic environment faced by the social reference.

This idea is reminiscent of the influential paper from Gneezy and Rustichini (2000a,b) and the more recent paper from Fuster and Meier (2009). From their experiments,

Table 4: OLS for the Effect of Partner's Incentives on Donations and Beliefs (Coefficient Estimates and Standard Errors in Parentheses)

(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: None)*					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: None)*					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 *None* = *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: None)*					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: None)*					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 *None* = *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.

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 seem to simply shut down adherence to social norms: in fact, the size of incentives matters. Relatively small incentives to act prosocially can preserve a certain level of norm adherence and produce social influence.[24] 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 prosocial behavior.

Result 3 is also important for methodological reasons. 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.[25] 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)). Result 3 is key for illustrating a violation of the exclusion restriction that would be necessary for using incentives to the social reference as an instrument for beliefs.[26]

3.4. 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

[24] Although the absence of social interactions after the donation does not allow us to explain our findings with the same behavioral channel, the theory of Dutta et al. (2018) of ostracism and endogenous social norms demonstrates that sufficiently large incentives are needed for the collapse of norm following.

[25] The concern that beliefs on others behavior more closely reflect the response function of the *observer* rather than the *observed*.

[26] 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 size of incentives, without being able to test the validity of the exclusion restriction.

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 a plausible restriction for globally explaining the experimental data.

The restrictions on average donation levels posed by the *main diagonal condition* are summarized in Table 5. Average donations in the nine incentivized treatments of our experiment can be treated as a nine-dimensional normal distribution with means $\boldsymbol{\mu}_{p_i,p_j}$ (which we treat as unknown) and diagonal covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\sigma^2}_{p_i,p_j}\mathbb{I}$ (which we treat as known). For the joint test, we use maximum likelihood to determine the vector $\hat{\boldsymbol{\mu}}_{p_i,p_j}$ that best fits the nine dimensional vector of sample means $\overline{\boldsymbol{Donation}}_{p_i,p_j}$ - with and without the inequality constraints imposed by the *main diagonal condition* that we summarize in Table 5. A Likelihood Ratio test from the constrained and unconstrained likelihood functions is used to assess the joint validity of these constraints. The test statistic is $\chi^2_{(d)}$ distributed with degrees of freedom d equal to the number of binding inequality constraints.

Table 5: Inequalities in Average Donations between Incentivized Treatments Predicted by the Main Diagonal Condition

		Incentives to other				
		None		Moderate		High
Incentives to self	None	$\mu_{n,n}$	>	$\mu_{n,m}$	>	$\mu_{n,h}$
	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 6 (columns (1) to (3)) that the main diagonal condition appears violated when incentives to self are both none and high. As already noted, this is because incentives to the partner non-monotonically affect individual donations. Also notice, that when we look at average donations from subjects with strong social proximity to their partner separately from subjects with weak social proximity to their partner violations of the main diagonal condition of incentive inequality are 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 explain the behavior of subjects with a strong connection to their partner.

Moving to a test of the significance of these patterns, Table 6 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)). This indicates that we cannot reject the null hypothesis that donation behavior in the sample is explained by the social influence effect 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, is highly heterogeneous among subject 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 need for conformity to explain the effects of partner's incentives on donations exhibited by individuals with above median social proximity towards their partners.

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

Table 6: 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	
		None (1)	Moderate (2)	High (3)	None (4)	Moderate (5)	High (6)		p-value (7)
Incentives to self	None	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	
		None (1)	Moderate (2)	High (3)	None (4)	Moderate (5)	High (6)		p-value (7)
Incentives to self	None	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	
		None (1)	Moderate (2)	High (3)	None (4)	Moderate (5)	High (6)		p-value (7)
Incentives to self	None	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.

result is clearly inconsistent with the morale effects of incentive inequality, and is also inconsistent with a concave altruistic utility of giving.[27] 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 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.

In the appendix, Table A.7 also illustrates consistent patterns of rejection of the joint test when we impose the *main diagonal condition* on expectations that people form about the behavior of the social reference. This indicates that the social influence effect that people anticipate when forming beliefs about the behavior of a close social reference is unlikely to be mainly driven by inequity effects.

Taken together, the results of the analyses in this section indicate that the morale effects of inequality can explain the non-monotonic effects of partner's incentives on donation behavior when incentives to self are moderate and violations of this inequity aversion framework are not sufficiently strong to reject the theory in our sample. However, we also showed that there are important heterogeneities. While morale effects explain very well the behavior of socially distant subjects, they miss the strong non-monotonic effects of partner's incentives observed among socially close subjects. These patterns can be explained by the conformity framework of social influence.

3.5. 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

[27] A more standard framework of inequity aversion (Fehr and Schmidt, 1999), in which subjects suffer from differences in realized payoff difference, seems inappropriate for this framework because subjects do not get to see and compare their realized payoffs with the partner's and is generally not tractable because realized payoff differences depend of the actions of both players in a pair. Such a framework only makes the clear prediction that partner's incentives should not affect individual donations when either of the two players 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 6 panel (c) provides some evidence against this prediction.

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 the state of the world, 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 the case of a binary donation decision 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, is silent on whether the behavior of more or less socially close individuals should induce the complementarities of behavior. For this reason, 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.[28]

4. Conclusion

This study proposes a novel experiment to study social influence independently of social learning. In our experiment pairs of players collaborate in 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 none 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

[28] 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.

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, we interpret the non-monotonic effects of partner's incentives on individual 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 their motives for donating to charity, and individuals are less prone to adhere to behavior 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 presence of incentives is not a binary switch of norm adherence. The stake of incentives matters.

An implication of this result 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 cops 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 to 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, but more studies should be conducted to quantify the economic significance of conformity to 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.

_ **and Dylan Walker**, "Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks," *Management Science*, August 2011, *57* (9), 1623–1639.

_ **and** _ , "Tie Strength, Embeddedness, and Social Influence: A Large-Scale Networked Experiment," *Management Science*, April 2014, *60* (6), 1352–1370.

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.

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.

Bruhin, Adrian, Lorenz Goette, Simon Haenni, and Lingqing Jiang, "Spillovers of Prosocial Motivation: Evidence from an Intervention Study on Blood Donors," SSRN Scholarly Paper ID 2554419, Social Science Research Network, Rochester, NY January 2015.

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.

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, Yan and Sherry Xin Li, "Group identity and social preferences," *American Economic 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 empathy?altruism 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 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.*

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

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 Cooperation," *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.

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.

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.

Karlan, Dean and John A. List, "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment," *The American Economic Review*, December 2007, *97* (5), 1774–1793.

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.

Lacetera, Nicola, Mario Macis, and Angelo Mele, "Viral Altruism? Charitable Giving and Social Contagion in Online Networks," *Sociological Science*, April 2016, *3*, 202–238.

Landry, Craig E., Andreas Lange, John A. List, Michael K. Price, and Nicholas G. Rupp, "Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, May 2006, *121* (2), 747–782.

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.

Rabin, Matthew, "Incorporating Fairness into Game Theory and Economics," *The American Economic Review*, 1993, *83* (5), 1281–1302.

Roberts, Russell D., "A Positive Model of Private Charity and Public Transfers," *Journal of Political Economy*, February 1984, *92* (1), 136–148.

Shang, Jen and Rachel Croson, "A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods," *The Economic Journal*, October 2009, *119* (540), 1422–1439.

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.

Vesterlund, Lise, "The informational value of sequential fundraising," *Journal of Public Economics*, March 2003, *87* (3-4), 627–657.

Warr, Peter G., "Pareto optimal redistribution and private charity," *Journal of Public Economics*, October 1982, *19* (1), 131–138.

A. Additional Tables

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	
		None	Moderate	High	None	Moderate	High	Test statistic	p-value
Incentives to other	None	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	
		None	Moderate	High	None	Moderate	High	Test statistic	p-value
Incentives to other	None	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	
		None	Moderate	High	None	Moderate	High	Test statistic	p-value
Incentives to other	None	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.3: Joint Problem Solving Task Software Interface

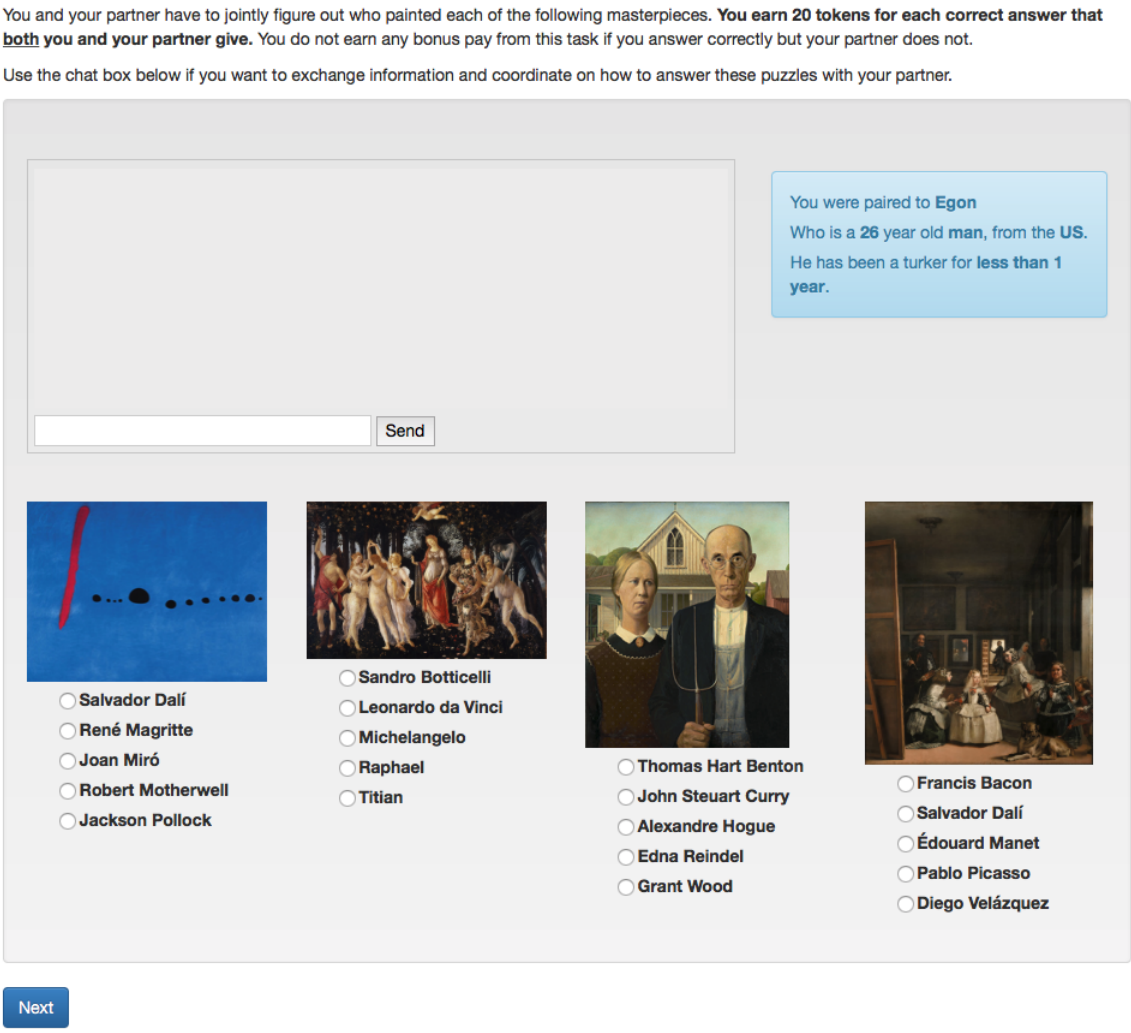
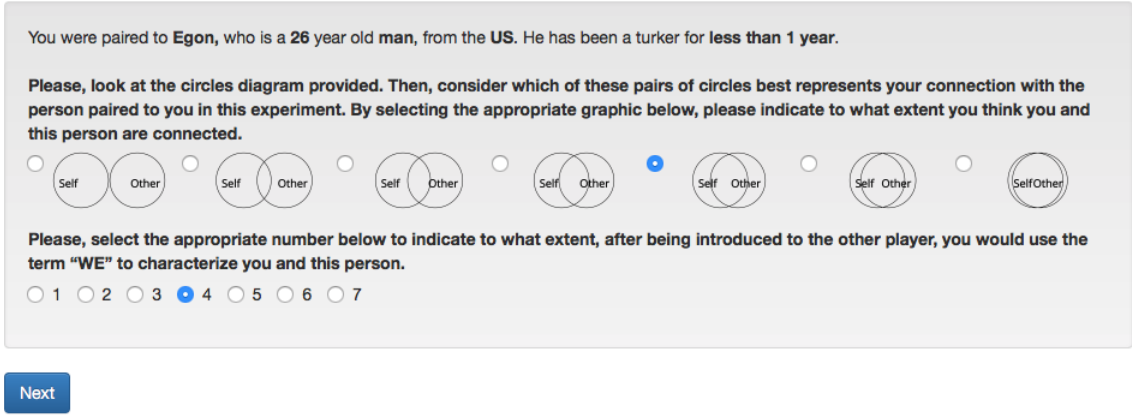


Figure B.4: Elicitation of the IOS (top) and WE (bottom) Scales



Next

Figure B.5: 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**. You can generate up to ten donations 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.

Egon is being **lucky**. He picked number 2. His winning number is between 1 and 3. He has **one chance in three to win the 50 tokens prize** for engaging in a donation, and has been informed of that.

You may be **lucky**! You picked number 5 and your winning number is between 4 and 6. You have **one chance in three to win the 50 tokens prize** for engaging in a donation.

You were paired to **Egon**, who is a **26** year old **man**, from the **US**. He has been a turker for **less than 1 year**.

How many donations would you expect Egon 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 three to earn 50 tokens for himself)**
- ◯ **2 Donations (100 tokens to DWB , and one chance in three to earn 100 tokens for himself)**
- ◯ **3 Donations (150 tokens for DWB , and one chance in three to earn 150 tokens for himself)**
- ◯ **4 Donations (200 tokens for DWB , and one chance in three to earn 200 tokens for himself)**
- ◯ **5 Donations (250 tokens for DWB , and one chance in three to earn 250 tokens for himself)**
- ◯ **6 Donations (300 tokens for DWB , and one chance in three to earn 300 tokens for himself)**
- ◯ **7 Donations (350 tokens for DWB , and one chance in three to earn 350 tokens for himself)**
- ◯ **8 Donations (400 tokens for DWB , and one chance in three to earn 400 tokens for himself)**
- ◯ **9 Donations (450 tokens for DWB , and one chance in three to earn 450 tokens for himself)**
- ◯ **10 Donations (500 tokens for DWB , and one chance in three to earn 500 tokens for himself)**

How many donations would you like to generate yourself?

- ◯ **0 Donations (0 tokens for DWB)**
- ◯ **1 Donation (50 tokens for DWB , one chance in three to earn 50 tokens for yourself**
- ◯ **2 Donations (100 tokens for DWB , one chance in three to earn 100 tokens for yourself**
- ◯ **3 Donations (150 tokens for DWB , one chance in three to earn 150 tokens for yourself**
- ◯ **4 Donations (200 tokens for DWB , one chance in three to earn 200 tokens for yourself**
- ◯ **5 Donations (250 tokens for DWB , one chance in three to earn 250 tokens for yourself**
- ◯ **6 Donations (300 tokens for DWB , one chance in three to earn 300 tokens for yourself**
- ◯ **7 Donations (350 tokens for DWB , one chance in three to earn 350 tokens for yourself**
- ◯ **8 Donations (400 tokens for DWB , one chance in three to earn 400 tokens for yourself**
- ◯ **9 Donations (450 tokens for DWB , one chance in three to earn 450 tokens for yourself**
- ◯ **10 Donations (500 tokens for DWB , one chance in three to earn 500 tokens for yourself**