City Research Online

City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Opportunistic Conformism

Gary Charness, Michael Naef, and Alessandro Sontuoso

December 12, 2018

Abstract. We study strategic interactions that may be affected by belief-dependent, conformist preferences. Specifically, we propose that beliefs about the behavior of individuals in the same role (i.e., beliefs about "peer behavior") directly affect a player's utility. In examining conformism we propose an experimental design that verifies the presence of the relevant causality direction. Our data reveal "opportunistically conformist" behavior, as subjects are more likely to follow the purported majority if doing so implies an increase in expected material payoff. We provide a general framework that accounts for such a pattern.

KEYWORDS: Conformist preferences; Psychological games; Peers; Trust.

JEL Classification Numbers: C72, C91.

Contact: Gary Charness, Department of Economics, University of California, Santa Barbara, California, 93106-9210, charness@econ.ucsb.edu; Michael Naef, Department of Economics, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, michael.naef@rhul.ac.uk; Alessandro Sontuoso, Philosophy, Politics and Economics, University of Pennsylvania, 249 South 36th St., Philadelphia, Pennsylvania, 19104, sontuoso@sas.upenn.edu.

I. Introduction

Conformism is an element of major importance for economic outcomes, as information about peer behavior has been shown to influence a diverse range of choices, including employees' retirement savings decisions and executives' decisions (Beshears, Choi, Laibson, Madrian, and Milkman, 2015; Shue, 2013).[1] The economic theory of conformism has generally fit into two broad research streams. One such stream presumes that peers' behavior is copied as it reflects private information relevant to the individual's own decision (e.g., Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992). The second stream aims to capture the individual's inherent tendency to identify with a certain class of people (e.g., Bernheim, 1994; Akerlof, 1980). The explanatory power of these frameworks is usually limited to situations with no direct strategic interdependencies between the agents' decisions. In fact, *herding models* imply that the predecessor's observed choice influences one's action via a belief revision, but the decision problem entails that each individual who chooses the right option will receive a fixed payoff in any case, regardless of others (Banerjee, 1992). Similarly, *esteem-based models* assume that the individual cares about the perception of others regarding her own status, but the actions of these others do not typically enter each individual's utility function (Bernheim, 1994).

In this paper we set out to study strategic interactions that may be affected by belief-dependent, conformist preferences. Specifically, we propose that beliefs about the behavior of individuals in the same role (i.e., beliefs about "peer behavior") directly affect a player's utility. In examining conformism we introduce an experimental design that verifies the presence of the relevant causality direction. We do so by exogenously varying beliefs about peer behavior in sequential trust games (Berg, Dickhaut, and McCabe, 1995).

In particular, we investigate the social-psychology notion of conformism, that is, one's tendency to follow the modal behavior and beliefs of one's peers (Cialdini and Trost, 1998); this

---

[1] Peer effects have been found in field environments (e.g., Mas and Moretti, 2009) and in the laboratory (e.g., Falk and Ichino, 2006). Herbst and Mas (2015) show that laboratory estimates of a parameter for peer effects are quantitatively very similar to those from field studies. See also Charness and Fehr (2015) for discussion.

attitude is often characterized as driven by a desire to fit in or "band together" with others in a similar role (Cialdini and Goldstein, 2004). In order to pin down some behavioral predictions that are informed by this notion, we operationalize such a tendency by assuming that the utility function of a conformist player is the sum of a material payoff and a "psychological bonus": the latter captures the player's intrinsic *utility from fitting in*, and varies with the player's beliefs about peer behavior. For the purpose of generating additional predictions against which to analyze our experimental data, we survey alternative models that may entail a relationship between one's behavior and one's beliefs about peer behavior. We find that the best explanation of the data is consistent with our specification of conformism.

It should be stressed that the identification of endogenous peer influences in non-controlled environments has traditionally presented some challenges (Manski, 1993). Moreover, it has been observed that the agents' (first- and higher-order) beliefs about their peers' actions often match what the agents themselves end up doing in that same situation: this correlation between own actions and beliefs about others in the same role could be explained by two mechanisms. *(i)* A tendency to adjust one's behavior in order to fit in with the group ("social conformity"); that is, from the agent's viewpoint, *I do what it is thought most others would do* (Cialdini and Goldstein, 2004).[2] *(ii)* A tendency to overestimate the extent to which others are like oneself, and hence to project one's own action onto others ("false consensus effect"); that is, *I believe the others do what I myself would do* (Ross, Greene, and House, 1977).

In this connection, we note that belief-dependent motivations are generally inferred from experimental evidence of a belief-behavior correlation. (An exception is provided by Costa-Gomes, Huck, and Weizsäcker, 2014, who create an artificial instrumental variable to estimate

---

[2] In the first tests for conformity Asch (1956) had subjects determine the relative lengths of lines. All but one of the participants in each session were confederates of the experimenter, and had beforehand been instructed to give wrong answers in unanimity: as a result, approximately 35% of subjects gave the same incorrect answer as the misleading majority. While in Asch's experiments subjects exhibited a tendency to match the others' *observed* behavior, further research extended the notion of conformism to include an inclination to adapt to the others' *presumed* behavior as well as to their values and norms (Deutsch and Gerard, 1955).

the causal effect of first-order beliefs; see also Andreoni and Sanchez, 2014.) In the context of trust games, a correlation between a particular class of beliefs and actions has been interpreted as "guilt aversion" (Dufwenberg and Gneezy, 2000; Charness and Dufwenberg, 2006) or as "trust responsiveness" (Guerra and Zizzo, 2004; Bacharach, Guerra, and Zizzo, 2007). That is, an individual i adapts her behavior to the beliefs (about her behavior) held by her matched participant j, in order to avoid guilt from letting down j's expectations (Battigalli and Dufwenberg, 2007). Specifically, the presence of guilt aversion has been inferred from a correlation between own behavior and second-order beliefs (about own behavior), where such beliefs are elicited by asking subjects what they think their "opponents" (i.e., people in the other role) expect from them. However, it has been suggested that any such observed correlation may be due to consensus effects, which involve the opposite causal direction: people who are inclined to cooperate might infer from their own inclination that people in general are cooperative (Ellingsen, Johannesson, Tjotta, and Torsvik, 2010).[3,4]

Here we enter the debate on causality by examining a class of motivations – conformist preferences – that involve the same direction of causality as with guilt aversion (in the sense that beliefs cause behavior). More precisely, we test the hypothesis that beliefs about the *behavior of individuals in the same role* directly affect the player's utility, in such a way to reflect a desire to fit in with the purported majority of peers. Our experimental design involves a standard two-player binary trust game in which we inform each subject of the behavior that other same-role participants expect of same-role participants. To that end, we first elicit each subject's belief

---

[3] Ellingsen *et al.* (2010) tested for a causal effect of the matched participant's expectations on individual behavior. They did so by transmitting the respective Trustor's first-order belief to each Trustee, thereby generating an "induced" second-order belief (about own behavior) per Trustee. Having found no evidence for guilt aversion, they conjectured that the correlation (between own behavior and second-order beliefs about own behavior) observed in previous experiments might have been driven by consensus effects. So, in Ellingsen *et al.*'s experiment the existence of consensus effects has been presumed from the absence of evidence for causal effects in the other direction.

[4] Charness, Rigotti, and Rustichini (2017) provide evidence that consensus effects have some bite in experimental behavior. A very strong relationship between beliefs and behavior is observed on the aggregate level when beliefs are elicited from subjects who have already made choices in a prisoner's dilemma; when these beliefs are elicited from subjects who didn't play the game, the relationship is still positive and significant but substantially less strong.

about the behavior of participants in the same role (i.e., one's first-order belief about peer behavior); then, some of these beliefs are averaged and transmitted to other participants in the same role, providing subjects with an "induced" second-order belief about peer behavior.

Our data show that Trustees holding a first-order belief of *predominant cooperation* (on the part of other Trustees) were significantly affected by the inducement of a second-order belief of predominant non-cooperation. Conversely, Trustees with a first-order belief of predominant non-cooperation were unaffected by the inducement of a second-order belief of predominant cooperation. The effect of the exogenous information on Trustors' behavior is symmetrical, that is, an increase in the transmitted belief has a significant positive effect only on the group of Trustors who held a first-order belief of *predominant non-cooperation* (on the part of other Trustors). Such data patterns suggest "opportunistic" conformism, since subjects are more likely to follow the purported majority if doing so implies an increase in expected material payoff.

We stress that the above establishes the presence of the causality implied by conformist preferences for two reasons. Firstly, by informing each subject about others' first-order beliefs, we bring about an *exogenously-generated* second-order belief: thus, any correlation between such beliefs and behavior cannot be attributed to consensus effects. Secondly, those exogenously-generated beliefs involve the behavior of participants in the *same role*; that is, in contrast to previous studies we focus on expectations about the behavior of one's peers, so that one's payoff is not directly affected by those beliefs. To the best of our knowledge, we provide the first evidence for the causal effect of such beliefs on behavior: whereas previous research has shown evidence of an effect of manipulating first-order beliefs about peer behavior,[5] our study crucially accounts for the effect of second-order beliefs (about peer behavior) on one's behavior.

---

[5] The importance of social comparison on relative performance has been studied within the realm of labor economics. Kandel and Lazear (1992) and Huck, Kübler, and Weibull (2012) address the question of whether peer pressure improves performance in situations where payoffs are based on team incentives (i.e., an agent's effort task is defined so as to create a *positive externality* on other team members' payoffs). See Costa and Kahn (2013) for a field experiment on environmental nudges. For some early work, see Festinger's (1954) social-comparison theory; for a model of distributional concerns among peers, in the context of ultimatum games, see Ho and Su (2009).

In a nutshell, we examine strategic interactions that may be affected by belief-dependent, conformist preferences. We operationalize such preferences by assuming that a player's utility is the sum of her material payoff and a psychological bonus. The latter varies with the player's beliefs about peer behavior, in such a way that a player gains a higher psychological utility from following a *more popular* (i.e., purportedly frequent) behavior. We further assume that the extent to which a player derives such intrinsic utility from popular strategies varies with her *degree of conformism* (i.e., an individual-specific constant).

When peers' actual behavior is unobservable, as in our experiment, our framework suggests that individuals use the induced second-order beliefs to revise their first-order beliefs about peer behavior (a predictor of a strategy's popularity). Given that, our framework implies that – all else equal – one is likely to switch to the modal strategy indicated by the exogenous information especially *if* doing so increases one's expected material payoff as well.

The remainder of the article is organized in this manner: section II introduces the experimental design; section III discusses the notion of conformism and other relevant theories; sections IV and V present the experimental results, and section VI concludes.

II. Experimental design and procedures
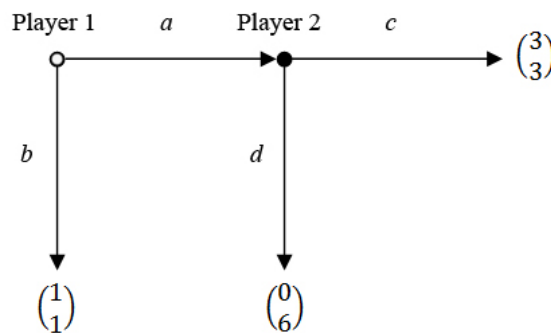
Consider the following binary "trust" game.

Figure 1 - The game

At the initial node player 1 (referred to as "Participant A" in the lab) chooses either *a* or *b*: when *b* is chosen the game ends and each subject receives 1 payoff unit. If player 1 chooses *a*,

player 2 (referred to as "Participant B" in the lab) in turn decides on *c* or *d*, the consequences of which are shown in the respective payoff vectors (the monetary payoff for player 1 is on top).

1. *The design*

 Our main treatment induces an exogenous variation in second-order beliefs about peer behavior by showing subjects the average guess (about the strategy taken by same-role participants) made by a sample of other same-role participants. Each experimental session consisted of the following stages: Introduction Stage; Play Stages I, II; Payment Stage.

Introduction Stage. Subjects were randomly allocated to terminals and given the paper instructions. They were then told that in Part I they would be assigned one of two roles, and we explained the decisions involved in each role. (We shall see that every subject in each part was privately assigned the same role, so the matching of subjects would be implemented only at the end of Part II.) After going through the paper instructions, each subject was asked to answer a set of control questions. A summary of the instructions was finally read aloud by the experimenter.

Play Stage, Part I. All plays were conducted using the strategy method.[6] The order of subsequent tasks was as reported below.

 (i) Subjects were (privately) assigned the role of Participant B.
 (ii) Each subject was asked to guess how many of the other B participants in the same session would choose either *c* or *d*, which – in the lab – were labeled as *"share"* and *"keep"*, respectively. (In what follows such stated beliefs will be formally denoted by $\gamma_2(\cdot)$.) Subjects entered their guess by positioning a slider to the desired percentage.[7]
 (iii) Subjects were invited to wait until all participants had entered their guesses, after which each subject was given feedback about some other B's guesses. (Using our formal notation, $\bar{\gamma}_2(\cdot)$.)

---

[6] While the strategy method is considered controversial by some, Brandts and Charness (2011) find that there are typically no differences regarding treatment effects in the literature. Furthermore, they find that a treatment effect observed using the strategy method is always also observed with the more traditional direct-response method.
[7] The slider was initially positioned at a value of 50%. Subjects had to enter a guess by moving the slider toward a higher share rate (i.e., toward a value of 100%) or a higher keep rate (i.e., toward a value of 0%). Subjects could not leave the slider in the initial position: thus, they had to take a stance and express a belief about the modal behavior.

(iv) Each subject was asked to choose either *c* or *d*, namely *"share"* and *"keep"*, respectively.

Play Stage, Part II. Subjects were told that Part II involved exactly the same steps as Part I, but that one would have a different role and be matched with a different person than in Part I. After they had been given a brief reminder of the instructions, both on-screen and orally, subjects were privately assigned the role of Participant A. Steps *(i)-(iv)* of Part II had the same structure as above, except that each subject's guess, transmitted information, and decision were about *a* or *b*, which – in the lab – were labeled as *"in"* and *"out"*, respectively.

Payment Stage. The payment mechanism consisted of two parts:
- each subject received a £3 show-up fee;
- each subject was paid (in pound sterling) according to the outcome for both Part I and Part II.

Note that the order of the decisions (reversed with respect to the natural sequence *A, then B*) was made possible by the strategy method. It should be stressed that in Part II each subject was matched with a different participant than her match in Part I, and was so informed. In any case, subjects did not know about the tasks to be undertaken in Part II until the end of Part I. Also note that subjects were not told how much they had earned until the end of the experiment.

2. *The belief-transmission mechanism*

The information provided at step *(iii)* consisted of the average guess made by a sample of other participants in the same role, in the same session. Note that, when entering their guesses – at step *(ii)* – subjects did not know that those guesses would be pooled and transmitted to other participants at step *(iii)*.[8] The information was shown in the lower part of the same screen in which subjects were asked to enter their guesses: the message was phrased in such a way as to

---

[8] It is certainly the case that – in Part I – subjects could not know that those guesses would be combined and passed on to other participants at step *(iii)*. It is however possible that subjects in Part II might have imagined that their guesses would be passed on. In section V below we verify that participants did not strategically misreport their beliefs.

7

look like the outcome of an informal opinion survey; the font style and size were the same as those of the other messages, in order not to make the information prominent. In Part I the message read: "A sample of other B participants in this session expects on average that [x]% will transfer half the money, whereas [100-x]% will keep all the money." In Part II the message read: "A sample of other A participants in this session expects on average that [x]% will opt in, whereas [100-x]% will opt out." Experimental instructions and screenshots are in Appendix B.

We used a computerized sampling method for selecting the subjects whose guesses would be pooled and passed on to other participants, in order to collect enough data to conveniently test for our key hypotheses. Specifically, in each part of a session we randomly assigned participants to receive information about an average belief of either *predominant non-cooperation* or *predominant cooperation*. More precisely, the samples of subjects (hence, guesses) were selected in a way such that the transmitted averages were close to either 0.25 or 0.75 (i.e., using our formal notation: $\bar{\gamma}_i(share) \sim 0.25$ or $\bar{\gamma}_i(share) \sim 0.75$ in Part I; $\bar{\gamma}_i(in) \sim 0.25$ or $\bar{\gamma}_i(in) \sim 0.75$ in Part II).

We stress that the reason why the sampling algorithm was devised in a way to select samples of subjects such that all $\bar{\gamma}_i(\cdot)$ approached either 0.25 or 0.75 was simply to obtain two distributions of transmitted beliefs for each part of a session; that is, the "low transmitted belief" and the "high transmitted belief" distributions. We note that one could have chosen any other value: we chose 0.25 and 0.75 only because each is the central value of a range of beliefs about low cooperation [0.00-0.49] or high cooperation [0.51-1.00].[9]

---

[9] It should be stressed that we phrased the on-screen message reporting the average of others' stated beliefs in such a way as to avoid deception. As mentioned above, the message explicitly stated that the reported information referred to *a sample of other participants* (see Appendix B). We believe that our design complies with the norm of experimenter honesty in that we did not lie to the subjects. The reason we chose not to transmit mean beliefs from random samples was simply as a means to have some subjects receive generic information about an average belief of low cooperation and have some others receive generic information about high cooperation, in a symmetric fashion. If we had used random samples, we would have needed hundreds of observations to obtain the symmetric distribution of transmitted beliefs necessary to conduct the analysis below. Finally, note that our explanation of the results does not rest on the assumption that subjects believed the information to be representative.

III. Theoretical predictions

To formulate hypotheses in the context of our experimental game, one has to make two sets of assumptions. The first concerns the nature of the *belief revision* that is triggered by the treatment manipulation. The second concerns the nature of individual *preferences* or, more precisely, the arguments of a subject's utility function. These two sets of assumptions are closely interrelated, as a belief revision may directly or indirectly affect the arguments of a subject's utility function.

We begin by stressing that our experimental manipulation entails the transmission of other participants' first-order beliefs about peer behavior (thereby inducing an exogenous variation in second-order beliefs about peer behavior). In this connection, we note that the fact that one has stated some belief about peer behavior does not necessarily convey information about one's own behavior (e.g., a selfish Trustee may well think that other Trustees are not selfish). Hence, our experimental manipulation cannot be characterized as a "signal" in the sense of a standard Bayesian model. Given the abundant evidence of heterogeneity of behavior and beliefs in social-dilemma experiments (Cooper and Kagel, 2013), specific assumptions about individuals' preferences and belief-revision processes are necessary, if one is to draw predictions about the impact of the present treatment manipulation.

So, we proceed to review theories that – more or less implicitly – entail a relationship between one's behavior and one's beliefs about peer behavior. Below we denote by $\gamma_i(s_i)$ "player i's prior belief about peer behavior" (i.e., the *belief about same-role players stated by i*), whereas we denote by $\bar{\gamma}_i(s_i)$ the "mean prior belief held by the members of a sample of peers" (i.e., the *exogenous belief about same-role players transmitted to i*). For ease of exposition, we often refer to $\gamma_i(s_i)$ as the prior/stated belief, and to $\bar{\gamma}_i(s_i)$ as the exogenous/transmitted belief. Furthermore, we respectively denote by $\alpha_i(s_j)$ and $\beta_i(s_i)$ the first- and second-order beliefs involving the opponent (i.e., the counterpart in a matched pair).

False consensus effects (Ross *et al.*, 1977; Dawes, 1989; Engelmann and Strobel, 2000). This theory entails a cognitive bias – not a social preference – whereby individuals overestimate the extent to which their actions are representative of the actions of others in the same role. This

9

means that whenever i is prompted to estimate her peers' behavior (i.e., the behavior of participants who are assigned the same role), i will project her own choice behavior onto others. So, the consensus effect entails a causal relationship from own behavior s_i to own belief about peer behavior $\gamma_i(s_i)$. As such, this theory does not imply any impact of the exogenous information $\bar{\gamma}_i(s_i)$ on the behavior of a subject who is affected by this cognitive bias.

Intention-based models of social preferences (e.g., Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006; Charness and Rabin, 2002; etc.). These theories assume that individuals exhibit a concern for a variously-defined notion of social welfare, such that a player tends to be fair to opponents who are believed to be fair (relative to some notion of social welfare). To that end, these models allow for the player's *beliefs about the opponents* to enter her utility function; as such, the models do not directly involve any beliefs about peers. The theories therefore remain indeterminate insofar as our experimental manipulation is concerned, unless one makes some ad hoc assumptions about the relationship between the exogenous information, $\bar{\gamma}_i(s_i)$, and i's belief about opponent j's behavior, $\alpha_i(s_j)$. For example, let's assume that there is a causal relationship $\bar{\gamma}_i(s_i) \rightarrow \alpha_i(s_j)$, such that i uses $\bar{\gamma}_i(s_i)$ to inform $\alpha_i(s_j)$.[10] As our experimental design does not measure prior or posterior beliefs $\alpha_i(s_j)$ in each role,[11] for the purposes of the present discussion we take the empirical assumption above to imply *on average* that: "if i is informed that i's peers believe that most other peers will cooperate, then i believes

---

[10] It is unclear if this empirical assumption could be normatively justified in the context of a social-preference theory. In fact, models of social preferences assume heterogeneity in players' degrees of concern for the social welfare. Now, even if the experimental instructions made the structure of the game commonly known among participants, clearly no information about the opponents' or peers' preferences is provided by the experimenter. Hence, the game played in the lab would be best described as one with incomplete information about the others' social preferences. This means that in our experiment – in order for, say, a Trustor to draw inferences about the respective Trustee's type *on the basis of our exogenous information* – one would need to assume that peers' beliefs are reflective of the statistical distribution of Trustees' social-preference types.

[11] In this regard, we note that in Part I of the experiment we elicit Trustees' beliefs about peer behavior. Thus – from the viewpoint of a subject making a decision in Part II – the beliefs we elicited in Part I may be treated as Trustors' priors about the opponents' behavior. Our regression analysis below will account for such beliefs.

that opponent j will most likely cooperate as well." Such an assumption in turn implies that the behavior of other-regarding subjects should be positively affected by the exogenous information. More explicitly, this means that if subjects receive a high exogenous belief, then they will be more likely to cooperate; on the other hand, if subjects receive a low exogenous belief, then they will be more likely to defect. (It is arguable whether this prediction would apply to our Trustees as well, given that our design involves the strategy method.) As will soon be clear, this model is inconsistent with the data, as our regressions do not show a significant impact of the main-effects variable $\bar{\gamma}_i(s_i)$ on behavior s_i in either role (instead, we find that the impact of $\bar{\gamma}_i(s_i)$ on s_i is actually conditional on prior belief $\gamma_i(s_i)$, suggesting that the latter may directly enter the subject's utility function).[12]

Guilt aversion (Battigalli and Dufwenberg, 2007; Charness and Dufwenberg, 2006). A related explanation for behavioral regularities in experimental trust games is guilt aversion, whereby individuals experience a utility loss if they believe they let their opponent down. The theory of guilt aversion entails a causal relationship from *second-order belief about own behavior $\beta_i(s_i)$* to own behavior s_i; as such, it does not directly involve any beliefs about peers. Again, this model remains indeterminate insofar as our experimental manipulation is concerned, unless one makes some ad hoc assumptions about the relationship between the exogenous information, $\bar{\gamma}_i(s_i)$, and i's second-order belief about own behavior, $\beta_i(s_i)$. So, here let's assume that there is a causal relationship $\bar{\gamma}_i(s_i) \rightarrow \beta_i(s_i)$, such that i uses $\bar{\gamma}_i(s_i)$ to inform $\beta_i(s_i)$.[13] As our experimental design

---

[12] It would seem a bit farfetched to apply the social-preference models above to our game setting, by making the additional assumption that i revises first-order belief $\gamma_i(s_i)$ on the basis of the exogenously induced second-order belief $\bar{\gamma}_i(s_i)$. Even if one wished to make such an assumption, it would be unclear in the context of those theories what would justify the revision in the absence of a conformist attitude.

[13] As in the case of social preferences, it is unclear if this empirical assumption could be normatively justified in the context of a model of guilt aversion, which allows for heterogeneity in players' guilt sensitivity. In this case, the game played in the lab would be best described as one with incomplete information about the others' guilt sensitivity. This means that in our experiment – in order for, say, a Trustor to draw some inferences on the basis of our exogenous information – one would need to assume that peers' beliefs are reflective of the statistical distribution of Trustees' guilt and epistemic types. (See the related discussion in Attanasi, Battigalli, and Manzoni, 2016.)

does not measure prior or posterior beliefs $\beta_i(s_i)$, we take this empirical assumption to imply *on average* that: "if i is informed that i's peers believe that most other peers will cooperate, then i believes that opponent j will most likely expect i to cooperate as well." Such an assumption in turn implies that the behavior of guilt averse subjects should be positively affected by the exogenous information. (As noted above, this is inconsistent with our data.)

Conformism. We conclude this section by formulating hypotheses that are informed by the attitude to which social psychologists refer as "conformism". This entails that individuals have a tendency to follow the modal behavior and beliefs of their peers (Cialdini and Trost, 1998). In particular, the social-psychology notion of conformism is often characterized as driven by a desire to fit in or "band together" with others in a similar role (Cialdini and Goldstein, 2004). In order to pin down some behavioral predictions, we operationalize such a desire by assuming that the utility function of a player with conformist preferences is the sum of a material payoff m_i and a "psychological bonus" f_i. The latter captures the player's intrinsic utility from fitting in with the crowd, and is defined by

$$f_i = g_{s_i} \cdot C, \tag{1}$$

with g_{s_i} denoting the relative popularity (i.e., the frequency among i's peers) of the player's chosen strategy s_i; further, C denotes a non-negative, individual-specific (and role-independent) constant measuring the extent to which the player is a conformist. Now – since in our game a subject does not observe how popular each strategy is among her peers – we assume that i uses her belief about peer behavior $\gamma_i(s_i)$ to estimate g_{s_i}. So i's expected utility from strategy profile $s = (s_i, s_{-i})$, given belief γ_i, is defined by

$$E_{\gamma_i}[U_i] = m_i + \gamma_i(s_i) \cdot C, \tag{2}$$

where the first and second term respectively denote i's material payoff and her (anticipated) psychological bonus. Before proceeding we note that, in our binary trust game, expression (2) implies that a conformist player will choose the strategy s_i that maximizes her material payoff or her psychological bonus (or both). This explains for example why – absent any exogenous information – cooperation is the rational course of action for those Trustees who believe "share" to be the most popular strategy (i.e., $\gamma_i(share) > 0.5$) and, at the same time, feature a sufficiently large parameter C.

Next, we move on to define the belief revision triggered by the experimental manipulation. In short, we assume that i revises her first-order belief about peer behavior on the basis of the exogenously induced second-order belief $\bar{\gamma}_i(s_i)$. In Appendix A we specify a formal belief-revision process, but for the time being it suffices to state that such a revision simply entails a positive correlation between the exogenous information and the subject's estimate of g_{s_i}. (We stress that this assumption is empirically justified by the social-psychology definition of conformism, that is, one's tendency to follow the modal behavior and beliefs of one's peers.[14])

What are the behavioral implications of such a belief revision? When the prior and the induced second-order belief coincide directionally (in the sense that they imply the same modal strategy), an individual naturally does not change her view; so, we expect no behavioral change.

More interestingly, consider the case in which a subject's prior belief and exogenous information "conflict" in that the former implies that most peers take strategy s_i, whereas the latter implies a different modal strategy s_i'. Given some heterogeneity in subjects' degrees of conformism, C, we expect a subject to switch to the newly-learned modal strategy especially *if* doing so increases her expected material payoff as well (relative to the payoff i would have gotten had she followed the modal behavior indicated by her prior, all else equal). That is, i is more likely to switch to the newly-learned modal strategy s_i' if $m_i(s_i', \cdot) \geq m_i(s_i, \cdot)$.

> *Hypothesis*. Trustees holding a prior of predominant cooperation (i.e., $\gamma_i(share) > 0.5$) are affected by the inducement of a second-order belief of predominant non-cooperation; furthermore, Trustors holding a prior of predominant non-cooperation (i.e., $\gamma_i(in) < 0.5$) are affected by the inducement of a second-order belief of predominant cooperation.

What justifies this prediction? As noted above, in our binary trust game, expression (2) implies that a conformist player will choose the strategy that maximizes her material payoff or her

---

[14] In Appendix A we formalize the belief-revision process by assuming that i calculates a weighted average between i's own prior $\gamma_i(s_i)$ and the prior belief held by the members of a sample of peers, in such a way to attach equal weight to each individual. We then show that, given some idiosyncratic noise in individual priors, the proposed revision process is normatively justified in that it leads to more accurate beliefs about peer behavior.

psychological bonus (or both). However, since the experimenter does not observe the individual parameter C – and hence cannot quantify the change in psychological bonus resulting from a shift in g_{s_i} (i.e., a shift in the belief about peer behavior) – the prediction above focuses on cases in which the newly-learned modal behavior potentially implies a relative increase in material payoff. There, we predict an effect of $\bar{\gamma}_i(s_i)$ on s_i, conditional on $\gamma_i(s_i)$. We refer to this particular pattern as "opportunistic" conformism.

Examples illustrating the hypothesis above. Consider those Trustees who would have shared, absent the conflicting information (i.e., Trustees holding a high prior $\gamma_i(share)$ and a sufficiently high degree of conformism C). In a nutshell, upon receiving a low exogenous belief, many such Trustees will come to believe that "keep" is actually the more popular strategy: as a result of the belief revision, these Trustees can improve both their material payoff m_i and overall expected utility $E_{\bar{\gamma}_i}[U_i]$ by not sharing (we therefore predict a treatment effect; for a formal account see Observation 1 in Appendix A). Furthermore, consider those Trustors who would have opted out, absent the conflicting information (e.g., Trustors holding a low prior $\gamma_i(in)$ and a sufficiently high degree of conformism C). In a nutshell, upon receiving a high exogenous belief, many such Trustors will come to believe that "in" is actually the more popular strategy: as a result of the belief revision, these Trustors can improve both their expected material payoff and overall utility by opting in, provided that they expect more than a third of Trustees to share (hence, there should likely be a treatment effect; see Observation 4 in Appendix A).

Examples illustrating scenarios as to which we are agnostic. Consider those Trustees who would have not shared, absent the conflicting information (e.g., Trustees holding a low prior $\gamma_i(share)$). Upon receiving a high exogenous belief, many such Trustees will come to believe that "share" is actually the more popular strategy: these Trustees will not improve their material payoff by sharing, so whether we should see a significant effect here depends on the eventuality that such Trustees have a sufficiently high degree of conformism (Observation 2 in Appendix A). Similarly, consider those Trustors who would have opted in, absent the conflicting information (e.g., Trustors holding a high prior $\gamma_i(in)$). Upon receiving a low exogenous belief, many such Trustors will come to believe that "out" is actually the more popular strategy: these Trustors will not improve their expected material payoff by opting out if they expect more than a third of

14

Trustees to share; so, whether we should see a significant effect here depends on the eventuality that such Trustors have a sufficiently high degree of conformism (Observation 3 in Appendix A).

In summary – based on the hypothesis above – in Part I of the experiment we predict an effect of the exogenous information on Trustees who hold a prior of predominant cooperation, i.e., $\gamma_i(share) > 0.5$. By contrast, in Part II there should likely be an effect of the exogenous information on Trustors who hold a prior of predominant non-cooperation, i.e., $\gamma_i(in) < 0.5$. (As shall be seen, both predictions are supported by our data, thereby confirming the hypothesis above.) We are agnostic as to the other scenarios.

IV. Experimental results

1. *Summary statistics and regression models*

Table 1 summarizes the data for the decision to cooperate and the stated belief. Note that the decision to cooperate is binary, with value 1 when a subject chooses *c* (i.e., the Trustee *shares*, in Part I) or *a* (i.e., the Trustor opts *in*, in Part II), and with value 0 otherwise. Most Trustees and Trustors cooperated (58 percent and 68 percent, respectively). On the other hand, Trustees expected on average that a minority of other Trustees would cooperate, whereas Trustors expected on average that most other Trustors would cooperate (44 percent and 59 percent, respectively).[15] For histograms, see Appendix A.

---

[15] The data further show a positive correlation between one's decisions across roles (i.e., *i*'s decision as Trustee and *i*'s decision as Trustor), with a correlation coefficient of 0.370 ($p = 0.000$). In a series of trust games where subjects played both roles, Blanco, Engelmann, Koch, and Normann (2014) found evidence of a similar correlation. Blanco *et al.* went on to show that the preference parameters typically assumed by models with *inequity-averse* (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) or *reciprocal* players (Dufwenberg and Kirchsteiger, 2004) predict a negative correlation for one's choices across roles. Thus, Blanco *et al.* (2014, p. 127) ruled out those formulations of inequity aversion and reciprocity as possible explanations for their data patterns.

Variable	Obs.	Mean	Std. Dev.
$share$	110	0.582	0.496
$\gamma_i(share)$	110	0.444	0.249
in	110	0.682	0.468
$\gamma_i(in)$	110	0.593	0.273

Table 1 - Summary statistics: $\gamma_i(share)$ and $\gamma_i(in)$ denote a subject's stated guess about the percentage of other participants that will choose *share* and *in*, respectively.

Before presenting the regression results, let $d^H\bar{\gamma}_i(\cdot)$ denote a dummy for the exogenous information, taking on value 1 or 0 if subject i's induced second-order belief about peer behavior implies predominant cooperation or non-cooperation, respectively. More explicitly, $d^H\bar{\gamma}_i(share)$ takes on value 1 if individual i received a belief $\bar{\gamma}_i(share) \sim 0.75$, whereas $d^H\bar{\gamma}_i(share) = 0$ if $\bar{\gamma}_i(share) \sim 0.25$. Similarly, $d^H\bar{\gamma}_i(in)$ takes on value 1 if individual i received a belief $\bar{\gamma}_i(in) \sim 0.75$, whereas $d^H\bar{\gamma}_i(in) = 0$ if $\bar{\gamma}_i(in) \sim 0.25$.

We begin to explore the relationship between behavior and beliefs (about peer behavior) by presenting the probit regressions in column 1 of Table 2 below, where:

- the model in the top panel of column 1 consists of the Trustee's decision as the dependent variable, and of the Trustee's stated belief $\gamma_i(share)$ as a predictor for Part I;
- the model in the bottom panel of column 1 consists of the Trustor's decision as the dependent variable, and of the Trustor's stated belief $\gamma_i(in)$ as a predictor for Part II.

In short, column 1 reveals a strongly significant relationship between prior beliefs and behavior. While the result is consistent with a conformist tendency, it does not demonstrate a causal effect of beliefs on behavior. We therefore move on to the analysis of exogenously-generated beliefs.

Now, the model in the *top panel* of column 2 has the Trustee's decision as the dependent variable, and the Trustee's transmitted belief dummy $d^H\bar{\gamma}_i(share)$ as predictor. Similarly, the model in the *bottom panel* of column 2 has the Trustor's decision as the dependent variable, and the Trustor's transmitted belief dummy $d^H\bar{\gamma}_i(in)$ as predictor. Both panels of column 2 show a non-significant effect of the exogenous information, which suggests that second-order beliefs about peer behavior do not directly affect individuals' utility.

Next, the model in the top panel of column 3 consists of the Trustee's decision as the dependent variable, and of the following predictors: *(i)* the Trustee's stated belief, *(ii)* the transmitted belief dummy, and *(iii)* their interaction, i.e., $\gamma_i(share) * d^H \bar{\gamma}_i(share)$. (Note that we shall later address the bottom panel of column 3.) While the transmitted belief dummy remains non-significant, both the stated belief and the interaction variable are now significant at the 5% level. Thus the exogenous information has an effect on Trustees' behavior conditional on their priors. The result is consistent with our assumption that second-order beliefs about peer behavior affect individuals' utility via their first-order beliefs. In the next sections we shall see that the observed pattern reveals opportunistically conformist behavior (i.e., subjects are more likely to switch to the modal strategy indicated by the exogenous information, if doing so entails an increase in expected material payoff).

Comparing the values of the Akaike information criterion – AIC – for models 1-3 of Table 2 (top panel only), we conclude that the best specification is that of model 3, having the smallest value (132.523). (Note that AIC is a standard measure for estimating the quality of a model relative to each of the competing ones, whereby the model with the smallest AIC value is considered the best one.) We stress that unlike other selection criteria, such as the *adjusted R^2*, AIC is not biased by the sheer number of predictors used by each model (i.e., the model selected by the adjusted R^2 is typically the one with more variables). On the contrary, AIC is a more reliable selection criterion in that it penalizes models containing variables that do not significantly improve fit, as assessed by the likelihood function (Burnham and Anderson, 2002).

We turn to discuss models predicting Trustors' behavior. The model in the bottom panel of column 3 has the Trustor's decision as the dependent variable, and the Trustor's stated belief, transmitted belief dummy and their interaction as predictors. Both the transmitted belief and the interaction variable are significant (here the interaction variable is significant at the 10% level). The result confirms a conditional effect of the exogenous information, which is consistent with a conformist attitude: we shall expand on this in the next subsection.

	(1)	(2)	(3)	(4)	(5)
Share					
$\gamma_i(share)$: prior about Trustees	0.023*** (0.005)		**0.015**** (0.006)		
$d^H\bar{\gamma}_i(share)$: belief transmitted Pt I		0.077 (0.242)	**-0.830** (0.557)		
$\gamma_i(share) \cdot d^H\bar{\gamma}_i(share)$: interact. I			**0.029**** (0.013)		
Constant	-0.777*** (0.265)	0.166 (0.173)	**-0.586** (0.361)		
Pseudo R2	0.128	0.000	**0.167**		
AIC	134.398	153.429	**132.523**		
Obs.	110	110	**110**		

	(1)	(2)	(3)	(4)	(5)
In					
$\gamma_i(in)$: prior about Trustors	0.029*** (0.005)		0.039*** (0.008)		**0.040**** (0.009)
$d^H\bar{\gamma}_i(in)$: belief transmitted Pt II		0.190 (0.251)	1.257* (0.672)		**1.303*** (0.756)
$\gamma_i(in) \cdot d^H\bar{\gamma}_i(in)$: interact. II			-0.019* (0.011)		**-0.021*** (0.011)
$\gamma_i(share)$: prior about Trustees				0.014*** (0.005)	**0.018**** (0.006)
$d^H\bar{\gamma}_i(share)$: belief transmitted Pt I					**0.638**** (0.321)
Constant	-1.142*** (0.314)	0.382** (0.171)	-1.821*** (0.514)	-0.138 (0.254)	**-2.903**** (0.751)
Pseudo R2	0.242	0.004	0.268	0.054	**0.346**
AIC	108.180	141.023	108.660	134.077	**101.886**
Obs.	110	110	110	110	**110**

Table 2 - Probit regression coefficients: in brackets are robust standard errors (*, **, and *** indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively, for the relevant Z-statistic, two-tailed tests). The top panel refers to the Trustee's decision (Part I), whereas the bottom panel refers to the Trustor's decision (Part II).

In order to rule out a reasonable alternative to conformism as an explanation for our data patterns, we proceed to discuss the model in column 4. This model captures the reasoning of subjects with standard, self-centered preferences: if Trustors are standard utility-maximizers, then their behavior should be well predicted by their beliefs about the opponents' behavior. Recall that in Part I of the experiment we elicited Trustees' beliefs about the behavior of other Trustees: it follows that – from the viewpoint of a subject making a decision in Part II – the beliefs we elicited in Part I now represent Trustors' expectations about the opponents' behavior. Column 4 suggests a strongly significant positive effect of (prior) beliefs about the opponents' behavior on own behavior. However, comparing the values of the Akaike information criterion for models 1-4 of Table 2 (bottom panel only) suggests that model 4 should be discarded. We stress that, unlike the adjusted R^2, the AIC selection technique does not penalize a model for having fewer predictors than another; thus, the unfavorable AIC exhibited in column 4 provides evidence that Trustors' behavior is not best explained by the subject's prior about the opponents.

Finally, the model in column 5 regresses a Trustor's decision on: *(i)* the stated belief about other Trustors' behavior; *(ii)* the transmitted belief dummy about other Trustors' behavior; *(iii)* the interaction between *i* and *ii* (i.e., $\gamma_i(in) * d^H \bar{\gamma}_i(in)$); *(iv)* the stated belief about Trustees' behavior; *(v)* the transmitted belief dummy about Trustees' behavior. The transmitted belief dummy about Trustees' behavior is significant at the 5% level (while the rest of the coefficients remain very close to those of models 3 and 4). Note that the fact that the exogenous information from Part I is a significant predictor of behavior in Part II is remarkable in itself, as it confirms that individuals use others' beliefs to revise their own first-order beliefs (i.e., in this case their *beliefs about the opponent*). Comparing the values of the Akaike information criterion across all (bottom) models of Table 2 we see that the best specification is model 5, having the smallest value (101.886).

2. *Interpretation and marginal effects*

In what follows we comment on the size and interpretation of the treatment effects. Before doing so, recall that our "opportunistic conformism" hypothesis says that one is more likely to switch to the modal strategy indicated by the exogenous information, if doing so increases one's expected material payoff. As noted in section III above, this implies that in Part I

of the experiment there should be a likely effect of the exogenous information on Trustees who hold a prior of *predominant cooperation*. By contrast, in Part II there should be an effect of the exogenous information on Trustors who hold a prior of *predominant non-cooperation* (see Appendix A for a formal account). To assess the hypothesis above, we now take a closer look at the models singled out by the Akaike information criterion: we first examine the impact of the exogenous information on Trustees' behavior, using model 3 of Table 2 above. Since our specification of conformism says that the impact of the belief revision varies with Trustees' priors, using model 3 we compute the discrete change in predicted sharing (for the transmitted low/high belief $d^H \bar{\gamma}_i(share)$) at each value of the stated belief $\gamma_i(share)$. Figure 2 below graphs such differences, along with 95% confidence intervals: note that the normalized/baseline level indicates the "low treatment" (i.e., the effect of the low transmitted belief).
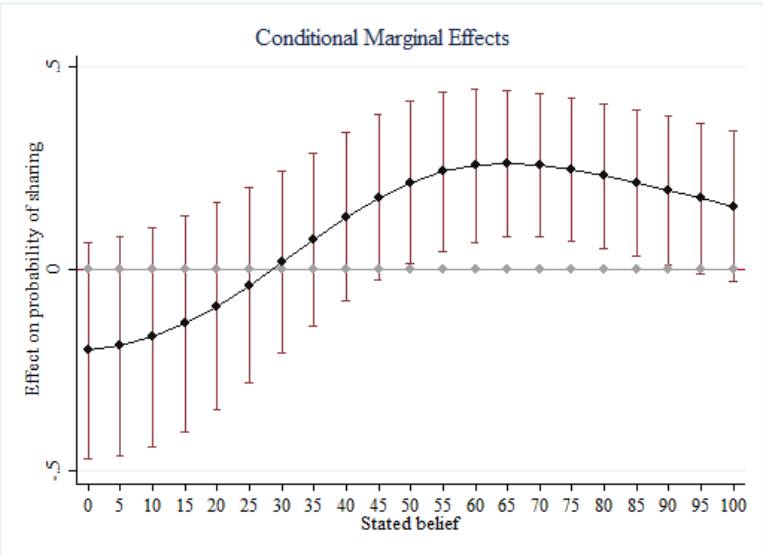

Figure 2 - Analysis of the Trustee's decision. The horizontal axis measures values of the stated belief $\gamma_i(share)$ in 5 percent increments. The vertical axis measures the discrete change from the *baseline level* (i.e., from $d^H \bar{\gamma}_i(share) = 0$) with 95% confidence intervals.

Figure 2 shows that for values of the prior belief greater than 50 percent (i.e., for those Trustees who believed that most other Trustees would cooperate), the marginal effect of the treatments was significant at the 5% level (one exception is that the effect was significant at the 10% level for individuals whose prior beliefs approached unity). On the other hand, for values of the prior belief less than 50 percent the marginal effect was non-significant.[16] While such data patterns appear to confirm our opportunistic conformism hypothesis, we stress that it is not possible to infer from this exercise whether the significant effects were driven by the "low treatment" rather than the "high treatment": this point is crucial for verifying the above predictions, and we shall address it in section V below.

We next examine the impact of the exogenous information on Trustors' behavior, using model 5 of Table 2 above. Again, it is useful to compute the discrete change in predicted opting-in (for the low/high transmitted belief $d^H \bar{\gamma}_i(in)$) at each value of the stated belief $\gamma_i(in)$, while keeping the other predictors at their mean values. Figure 3 graphs such differences, along with 95% confidence intervals. We note that, to a remarkable extent, Figure 3 resembles a mirror image of Figure 2. The data reveal that for those Trustors who believed that few other Trustors would cooperate (i.e., for values of the prior belief less than 30 percent), the marginal effect of the treatments was significant at the 10% level (two-tailed test). On the other hand, for relatively high values of the prior belief the marginal effect was non-significant. Once again this appears to confirm our opportunistic conformism hypothesis, even though it is not possible to infer from this exercise whether the effects were driven by the low rather than the high treatment. We will address this point in the next section.

---

[16] A somewhat related pattern is reported by Thöni and Gächter (2014). In a gift-exchange game with a principal and two employees, the authors find that – when an employee learns that her co-player has provided lower effort than she did – the employee revises her effort downwards; however, the employee hardly increases her effort when a co-player has provided higher effort than she did. Whereas our framework deals with *second-order* beliefs about unmatched participants, Thöni and Gächter's design involves a three-player game whereby either employee's *first-order* belief about the other employee is directly manipulated. See also Gächter, Nosenzo, and Sefton (2013).
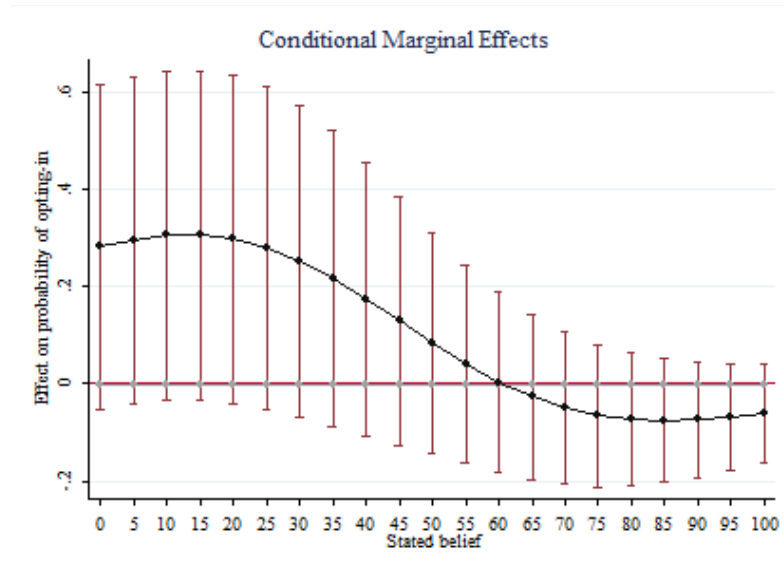
Figure 3 - Analysis of the Trustor's decision. The horizontal axis measures values of the stated belief $\gamma_i(in)$ in 5 percent increments. The vertical axis measures the discrete change from the *baseline level*, along with 95% confidence intervals, while the other predictors are held at their mean values (i.e., the change from $d^H\bar{\gamma}_i(in) = 0$ when $\gamma_i(share) = 0.444$ and $d^H\bar{\gamma}_i(share) = 0.5$).

A final comment: while the marginal effects on Trustors' behavior may appear weaker (in terms of significance) than the respective effects on Trustees, we stress that the size of the treatment effect on Trustors is nevertheless large. Calculating the Pearson's correlation coefficient between *transmitted belief* and behavior – in relation to the "relevant" set of subjects – does confirm the above findings. More specifically, for Trustors with priors below 50% [40 obs. out of 110] the correlation coefficient is 0.32, $p = 0.04$: this confirms that there is an effect of transmitting information on those Trustors.[17]

---

[17] For Trustees with priors above 50% [45 obs. out of 110] the correlation coefficient is 0.16, $p = 0.31$; however, by looking at the sample of Trustees stating priors above 55% [41 obs.] the correlation coefficient gets substantially larger and significant (0.36, $p = 0.02$). In fact, we note that there were some Trustees at the margin (i.e., stating a prior close to 50%) who behaved just like players in the 0-49 percent range of priors typically do. More precisely, three subjects with priors between 51 and 55 percent did not cooperate despite having received a high exogenous belief: this suggests that those three subjects had very low degrees of conformism.

V. Control treatments

1. *Robustness checks*

We designed two control treatments in order to illuminate the above results, as well as to investigate the potential impact of incentivizing beliefs. In short, Treatment T0 was identical to our main treatment above, except that there was no belief transmission at all. Treatment T1 was identical to T0, expect that beliefs were incentivized. Specifically, before entering their decisions each subject was asked to guess how many of the other participants (in the same role) would choose either action, but in T1 subjects were also told that they would receive an additional payment of £2 if their estimate differed by no more than 5 percentage points from the realized value (and would receive nothing otherwise). Subjects were not informed about the correctness of their guesses until the end of Part II. Table 3 summarizes the data for the decision to cooperate and the stated belief with reference to each treatment, T0 and T1.

Treatment T0

Variable	Obs.	Mean	Std. Dev.
share	53	0.509	0.504
$\gamma_i(share)$	53	0.386	0.259
in	53	0.641	0.484
$\gamma_i(in)$	53	0.504	0.309

Treatment T1

Variable	Obs.	Mean	Std. Dev.
share	46	0.586	0.497
$\gamma_i(share)$	46	0.457	0.254
in	46	0.608	0.493
$\gamma_i(in)$	46	0.625	0.258

Table 3 - T0 and T1 summary statistics: $\gamma_i(share)$ and $\gamma_i(in)$ denote a subject's stated guess about the percentage of other participants who will choose *share* and *in*, respectively.

We begin by checking for a correlation between $\gamma_i(s_i)$ and s_i. Unsurprisingly, when focusing on *Trustees* we find a strong correlation between beliefs about peer behavior and own

behavior in both T0 (Pearson's correlation coefficient = 0.485, p = 0.000) and T1 (coeff. = 0.499, p = 0.000). Similarly, when focusing on *Trustors* we find a strong correlation between beliefs and behavior in both T0 (coeff. = 0.587, p = 0.000) and T1 (coeff. = 0.518, p = 0.000).

We now turn to check whether the rate of cooperative behavior or the beliefs about peer behavior differ when such beliefs are incentivized. We first consider beliefs: for the Trustee's decision (i.e., Part I) the null hypothesis is that the mean for $\gamma_i(share)$ is the same for T0 and T1; for the Trustor's decision (i.e., Part II) the null hypothesis is that the mean for $\gamma_i(in)$ is the same for T0 and T1. For Part I, the Wilcoxon-Mann-Whitney test finds no significant difference between the underlying distributions of $\gamma_i(share)$ for T0 and T1, with Z = -1.296, p = 0.195 (two-tailed). For Part II, this test indicates mildly significant evidence against the null hypothesis (Z = -1.827, p = 0.067). Given this result, the analysis of the decision to cooperate (i.e., a test of whether behavior varied when beliefs were incentivized) becomes critical in assessing the potential impact of a belief-incentivization mechanism on our data patterns.

Hence, for the Trustee's decision the null hypothesis is that the mean for *share* is the same for T0 and T1; for the Trustor's decision the null hypothesis is that the mean for *in* is the same for T0 and T1. Regarding Part I, the Wilcoxon-Mann-Whitney test finds no statistically-significant difference across T0 and T1 (Z = -0.769, p = 0.442). Regarding Part II, the Wilcoxon-Mann-Whitney test finds a similar lack of evidence against the null hypothesis (Z = 0.335, p = 0.737). We conclude that paying (or not) for beliefs seems to have not greatly affected behavior or beliefs. Furthermore, if behavior in our controls T0 and T1 – in which subjects are *not* shown any exogenous information – is unaffected by our paying for beliefs, then we can reasonably assume that incentivizing beliefs would not change behavior even when subjects *are* shown the exogenous information. This corroborates the findings from our main treatment.

Our next robustness check verifies whether subjects strategically misreported their beliefs in the *main treatment*. The reason for such a test is the following: in Part I of the main treatment subjects could not likely anticipate that their stated beliefs would be passed on to other participants; it is however possible that, by the time subjects got to Part II, they might have imagined that their guesses would be passed on. So, it is important to verify that subjects did not strategically misreport their beliefs in the hope of influencing the others' behavior. To that end,

we test for the equality of the distribution of stated beliefs across all three treatments. The Kruskal-Wallis test finds no statistically-significant difference among the underlying distributions of $\gamma_i(share)$ across the three treatments ($\chi_2^2 = 2.350$, $p = 0.308$); there is also no statistically-significant difference among the underlying distributions of $\gamma_i(in)$ across the three treatments ($\chi_2^2 = 4.083$, $p = 0.129$). This suggests that subjects did state their true beliefs.

2.   *Final tests*

Our final tests investigate whether the effects observed in the main treatment were driven by the low rather than the high experimental condition. Thus, we separately consider the subsamples for which – in case of conformist preferences – it is more likely to expect a treatment effect (as per our opportunistic conformism hypothesis); i.e., we consider the samples of subjects who may increase both their expected material payoff and overall utility as a result of the belief revision.

Specifically, for Part I we shall focus on Trustees stating priors above 50%: in that sample we test the directional hypothesis that subjects who were provided with a low transmitted belief are *more likely to keep* than subjects who were *not* provided any information at all.[18] We note that only a relatively small number of subjects qualify for such a test (on the basis of their priors, i.e., 61 subjects in total). That said, the left panel of Figure 4 below shows an increase in the percentage of choices to *keep* (from 13.9% to 28.0%); a chi-square test provides mild evidence in support of the hypothesis ($\chi_1^2 = 1.859$, $p = 0.086$, one-tailed).

One might argue that this test may simply reflect the fact that Trustees with high priors are overly sensitive to their peers' beliefs, whether high or low. To test against this hypothesis – in the sample of Trustees stating priors above 50% – we verify whether subjects who were given

---

[18] The subjects who were not provided any information at all are in the pooled T0+T1 sample. The reason we have pooled together the participants from T0 and T1 is twofold: *(i)* we would not have enough observations for the relevant priors without pooling; *(ii)* the analysis presented in section V.1 above shows no meaningful differences in the distribution of stated beliefs or behavior between T0 and T1, justifying the current pooling of subjects.

a high exogenous belief were *less likely to keep* than subjects who were *not* given any information at all (56 subjects in total). This exercise shows no decrease in the percentage of choices to *keep*, which instead rises slightly and non-significantly from 13.9% to 15.0% when respectively comparing the no-information and high-transmitted-belief subsamples ($\chi^2_1 = 0.013$, $p = 0.454$, one-tailed; not shown in Figure 4). This does not support the hypothesis that such Trustees are overly sensitive to peer information, and so corroborates our reading of the results in terms of opportunistic conformism.
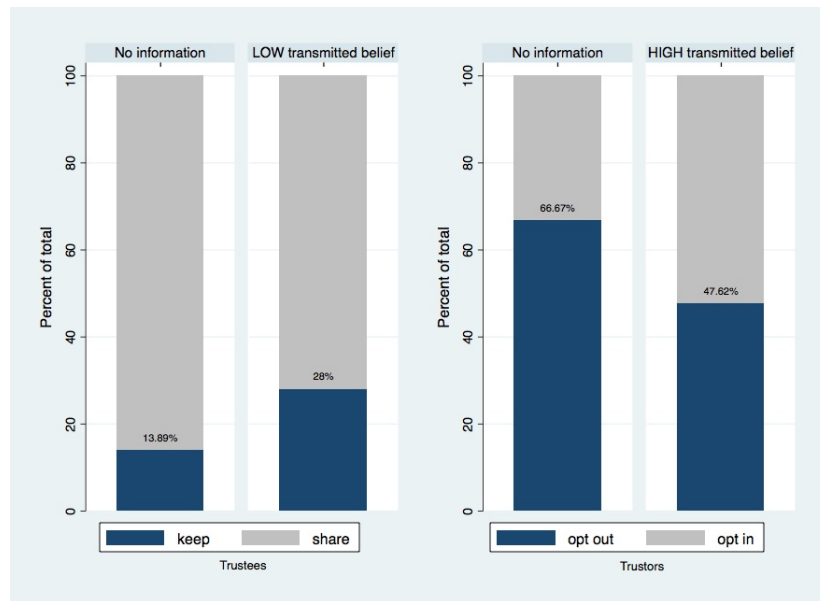

Figure 4 - The left panel graphs the effect of transmitting *low* information on the sample of Trustees who hold high priors. The right panel graphs the effect of transmitting *high* information on the sample of Trustors who hold low priors. (The vertical axis measures the percentage of total choices in the relevant sample.)

For Part II we consider Trustors who stated priors below 50%. We test the directional hypothesis that subjects provided with a high transmitted belief are *less likely to opt out* than subjects who were *not* provided any information at all. Again, we note that only a relatively small number of subjects qualify for this test (60 subjects in total). That said, the right panel of Figure 4 above shows a decrease in the percentage of choices to *opt-out* (from 66.7% to 47.6%); a chi-square test offers mild evidence in support of the hypothesis ($\chi^2_1 = 2.063$, $p = 0.075$, one-tailed).

We complete the analysis by checking whether Trustors who hold low priors are overly sensitive to their peers' beliefs, whether high or low. To test against this hypothesis – for Trustors stating priors below 50% – we verify whether subjects who were given a low exogenous belief were *more likely to opt out* than subjects who were *not* given any information at all (58 subjects in total). This exercise shows a non-significant change in the percentage of choices to *opt-out*, which increases from 66.7% to 78.9% when respectively comparing the no-information and low-transmitted-belief subsamples ($\chi_1^2 = 0.93$, $p = 0.167$, one-tailed; not shown in Figure 4). This does not support the hypothesis that such Trustors are overly sensitive to peer information, and hence corroborates our reading of the results as opportunistic conformism.

VI.   Concluding remarks

This study has investigated conformist preferences. We have operationalized these preferences by assuming that a player's utility varies with her beliefs about peer behavior, in such a way that a player gains a higher psychological utility from following a more popular (i.e., purportedly frequent) behavior. We then presented tests to determine if there are causal effects of beliefs about peer behavior, and found evidence in support of a conformist attitude.

Everyday experience teaches us that some individuals have a tendency to follow the (observed or purported) predominant behavior of the group. Social dilemma-like situations do not often exhibit a clear behavioral pattern as a result of individuals' conflicting motives and expectations. Nevertheless, the intrinsic *desire to fit in* is sometimes sufficiently strong to drive individual choices, even in the absence of implicit normative expectations to conform to peer behavior.[19]

---

[19] The focus of this paper was not on social norms and hence we have limited ourselves to manipulating empirical expectations (i.e., beliefs that a certain behavior *will* be followed), without controlling for normative beliefs (i.e., beliefs that a certain behavior *ought to* be followed). We are therefore agnostic as to the presence of social norms in our data (Sugden, 2000; Bicchieri and Xiao, 2009; Xiao and Bicchieri, 2010; Schram and Charness, 2015). See Bicchieri (2006) for a set of conditions for conformity to a social norm (see also Bicchieri and Sontuoso, 2015).

Here we have provided evidence of such a conformist attitude, having shown that individuals tend to follow the purported modal behavior of their peers, especially when it is "relatively convenient" to do so. Our study suggests that – when individuals are provided with exogenous beliefs about peer behavior – individuals update their own beliefs on the basis of the estimates made by others, and accordingly adjust their behavior. More specifically, our study shows that a subject is more likely to switch to the modal strategy indicated by the exogenous information, if doing so increases her expected material payoff. In particular, this suggests that our subjects generally derived moderate levels of intrinsic utility from following popular strategies, but their degree of conformism (as captured by the individual-specific value C) was *not* extremely large to the point of driving behavior regardless of material payoffs.

In conclusion, we have provided very suggestive evidence regarding the interplay of conformism, belief-revision processes, and self-interest. We hope that researchers besides ourselves will take up this interesting topic, as we suspect that such a relationship can be found in many economic settings.

Appendix A

Distribution of prior beliefs (main treatment)
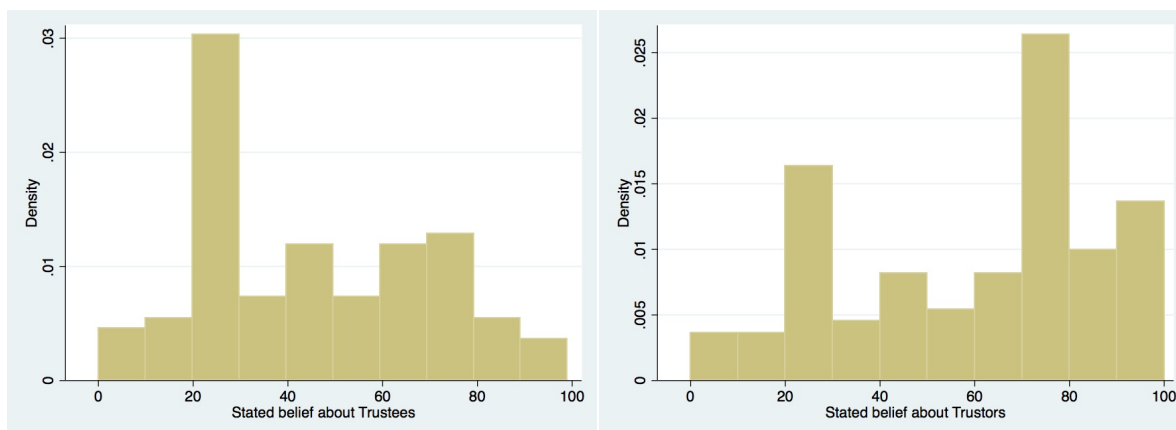
General framework

Preliminaries. In what follows we lay out a general model of conformist preferences, that is applicable to any multiplayer extensive form game. We begin with some notation.

An extensive form game is defined by a structure $G = \langle N, H, P, (\mathrm{I}_i)_{i \in N}, (m_i)_{i \in N} \rangle$, where N is the *set of player-roles*,[20] H is the *set of feasible histories*, P is the *player function*, I_i is player i's *information partition*, and m_i is player i's *material payoff*. Each element of H is a history, which is a finite sequence of actions taken by the players. Let $h(a^\ell)$ compactly denote the sequence $(a^1, ..., a^\ell)$, with a^ℓ being the ℓ-th action chosen along the game tree; a node is identified with the history h leading up to it. We denote by $A_i(h)$ the *set of feasible actions* for player i at history h.

Let Z denote the *set of terminal histories*, with $H \backslash Z$ being the set of non-terminal histories. The player function P associates to each element of $H \backslash Z$ an element of N, with $P(h)$ indicating the player who gets to choose an action after h (i.e., the active player). For each player

---

[20] We later distinguish between a *player-role* (i.e., a member of N) and an *individual participant* in the lab.

$i \in N$ we denote by I_i the information partition of player i (a partition of H), whereas we let $I_i \in I_i$ denote an "information set" of player i (i.e., a cell of the partition); we assume that such information structure satisfies perfect recall. The *set of nodes where i is active* is denoted by H_i, with $H_i := \{h \in H : P(h) = i\}$.

For each player $i \in N$ let S_i denote the *set of pure strategies* of player i, with generic strategy s_i, where $s_i = (a_{i,h})_{h \in H_i}$; note that $a_{i,h} \in A_i(h)$ indicates the action that would be selected by strategy s_i if history h occurred. A strategy profile s is a tuple of strategies, with one strategy per player: let $S = \times_{i \in N} S_i$ define the *set of strategy profiles*, and similarly define $S_{-i} = \times_{j \neq i} S_j$ for players j other than i. We further denote by $S_i(h)$ player i's set of "strategies allowing h" (i.e., strategies leading to node h); the set of strategy profiles allowing h is denoted by $S(h)$. Lastly, we denote by $z(s)$ a terminal history induced by strategy profile s; for each player $i \in N$ material payoffs are defined by functions $m_i : Z \to \mathbb{R}$.

Conformist preferences. Conformist players have a tendency to follow the modal behavior (and beliefs) of their peers. We now present the key ingredients for a simple utility function capturing one's happiness from behaving like the others. We start by modeling an experimental session, where tuples (e.g., pairs, in our trust game) of subjects are randomly formed to play an instance of the game.

Let $\mathcal{N}$ denote the *set of experimental subjects* (i.e., individual participants in a given session), with generic member x. Each individual $x \in \mathcal{N}$ is assigned a player-role $i \in N$ and plays an instance of game $G = \langle N, H, P, (I_i)_{i \in N}, (m_i)_{i \in N} \rangle$, as defined above. Formally, let R denote a surjective function $R : \mathcal{N} \to N$ (i.e., $\forall i \in N$, $\exists x \in \mathcal{N}$, s.t. $R(x) = i$), so that each player-role in N is associated with one or more individuals in $\mathcal{N}$. Given two individuals x and y, we denote by s_i^x and s_i^y the strategies respectively taken by x and y, when assigned role i; similarly, we denote by s_j^x and s_j^y the strategies respectively taken by x and y, when assigned role j. More generally, in what follows we use superscripts when we wish to distinguish among individuals (whereas we use subscripts to distinguish among roles, as usual).

Let $\mathcal{N}_i$ denote the *group of individuals in role i*: this is defined as the subset of $\mathcal{N}$ containing individuals who have been assigned role $i \in N$; that is, $\mathcal{N}_i := \{x \in \mathcal{N} : R(x) = i\}$. Similarly, let $\mathcal{N}_{-i}$ denote the group of individuals with roles j other than i (i.e., $\mathcal{N}_{-i} = \cup_{j \neq i} \mathcal{N}_j$). Furthermore, for each individual $x \in \mathcal{N}_i$ we denote by $\mathcal{N}_{-i}(x)$ the *set of x's co-players*: this is defined as the subset of $\mathcal{N}_{-i}$ containing individuals matched with x to play the same instance of

G. For example, consider a trust game and let $\mathcal{N}_i$ denote the group of individuals in the Participant A (Trustor) role; then, from the viewpoint of an individual $x \in \mathcal{N}_i$, the set $\mathcal{N}_{-i}(x)$ contains the one Participant B (Trustee) who has been matched with x to play an instance of the game.

Each individual $x \in \mathcal{N}_i$ holds a *system of conditional first-order beliefs* $\alpha_i^x = \left(\alpha_i^x(\cdot\,|h)\right)_{h \in I_i}$ about the strategies taken by her co-players (i.e., opponents); formally, $\alpha_i^x(\cdot\,|h) \in \Delta\left(S_{-i}(h)\right)$, with $\Delta\left(S_{-i}(h)\right)$ denoting the set of probability measures over the set of strategy profiles of individuals $y \in \mathcal{N}_{-i}(x)$.[21] Moreover, we denote "individual x's belief that members of $\mathcal{N}_i$ take s_i" by writing $\gamma_i^x(s_i)$: for brevity, this is often referred to as the individual's *belief about peer behavior*.[22]

We assume that the utility function of individual x (in role i) is the sum of a material payoff m_i^x and a psychological bonus f_i^x. The latter captures the individual's intrinsic utility from behaving like the others (without regard to any payoff considerations), and is defined by

$$f_i^x = g_{s_i} \cdot C^x, \tag{A1}$$

with g_{s_i} denoting the relative popularity of the individual's chosen strategy; further, $C^x \in [0, \infty)$ is an individual-specific (and role-independent) constant measuring the extent to which x is a conformist.

Note that we identify g_{s_i} with the number of experimental subjects who have chosen the same strategy as x, expressed as a fraction of $|\mathcal{N}_i|$.[23] However – since a subject does not typically observe how popular each strategy is among her peers – we assume that x uses her belief about

---

[21] Ideally such beliefs are updated conditional on I_i, and satisfy Bayes' rule whenever possible. In a game with perfect information the second-mover is assumed to hold an updated belief $\alpha_i(\cdot\,|h)$ at each $h \in I_i$, whereby she believes that the first-mover has implemented any action leading to h with probability one (for an extended account of conditional beliefs see Battigalli and Siniscalchi, 1999, and Battigalli and Dufwenberg, 2009). Because our design employs the strategy method, such conditional belief systems are not actually updated while the game unfolds.

[22] In the experiment we asked each subject to guess the proportion of other participants taking an action (not including the respondent), in order to make the task cognitively easier. The current specification of $\gamma_i^x(s_i)$, which includes the behavior of the respondent herself, simplifies the analysis but does not drive our results about the impact of transmitting aggregate information.

[23] That is, we identify g_{s_i} with the frequency of s_i among x's peers. Note that $|\mathcal{N}_i|$ denotes the cardinality of set $\mathcal{N}_i$ (i.e., the number of participants in role i).

peer behavior $\gamma_i^x(s_i)$ to estimate g_{s_i}. So x's expected utility from strategy profile $s = (s_i, s_{-i})$, given belief γ_i^x, is defined by

$$E_{\gamma_i^x}[U_i^x] = m_i^x(z(s)) + \gamma_i^x(s_i) \cdot C^x, \tag{A2}$$

where the first and second term respectively denote the individual's material payoff and her (anticipated) psychological bonus. We assume that each individual $x \in \mathcal{N}_i$ selects a pure strategy in order to maximize such expected utility function. (As usual, we assume common knowledge of the functional form of the individual's utility.) We finally note that, in our binary trust game, expression (A2) implies that a conformist individual will choose the strategy s_i that maximizes her material payoff or her psychological bonus (or both). Below are a few examples.

Illustrations (with no exogenous information). Consider our binary trust game, as depicted in Figure 1 above. Suppose that subject x is assigned the role of Trustee (i.e., $i = 2$) and that – for example – she believes that 75% of subjects in her role will share; that is, $\gamma_2^x(c) = 0.75$. In that case x's expected utility from *sharing* (i.e., $s_2^x = c$) equals $u_2^x(c, \gamma_2) = m_2(z(a, c)) + \gamma_2^x(c) \cdot C^x = 3 + 0.75 \cdot C^x$ (note that in what follows we denote by u_i^x the expectation of U_i^x, i.e., $E_{\alpha_i, \gamma_i}[U_i^x] \equiv u_i^x$; also, we often drop superscripts to simplify the exposition). On the other hand, x's expected utility from *keeping* (i.e., $s_2^x = d$) equals $u_2^x(d, \gamma_2) = 6 + 0.25 \cdot C^x$. Hence, x will prefer to share as long as $C^x \geq 6$. By contrast, assume that $\gamma_2^x(c) = 0.25$: in this case x's expected utility from *sharing* equals $u_2^x(c, \gamma_2) = 3 + 0.25 \cdot C^x$; on the other hand, x's expected utility from *keeping* equals $u_2^x(d, \gamma_2) = 6 + 0.75 \cdot C^x$. It follows that x will prefer to share as long as $C^x \leq -6$ (a negative value), and hence will never share.

Now suppose that subject y is assigned the role of Trustor (i.e., $i = 1$) and that she believes that 75% of subjects in her role will opt in; that is, $\gamma_1^y(a) = 0.75$. Further, let $\alpha_1^y(\cdot)$ denote y's first-order belief about her opponent. Here y's expected utility from opting *in* (i.e., $s_1^y = a$) equals $u_1^y(a, \gamma_1, \alpha_1) = \alpha_1^y(c) \cdot m_1(z(a, c)) + \gamma_1^y(a) \cdot C^y = 3 \cdot \alpha_1(c) + 0.75 \cdot C^y$; on the other hand, y's expected utility from opting *out* (i.e., $s_1^y = b$) equals $u_1^y(b, \gamma_1) = 1 + 0.25 \cdot C^y$. It follows that y will prefer to opt in as long as $C^y \geq 2 - 6 \cdot \alpha_1(c)$: most notably, this implies that y will always opt in if $\alpha_1(c) \geq \frac{1}{3}$ (regardless of C^y), capturing a case in which the expected material payoff is large enough to justify one's choice to opt in; also note that y will always opt in if $C^y \geq 2$ (regardless of α_1), which instead captures a case in which the expected psychological bonus is large enough to justify opting in. By contrast, assume that $\gamma_1^y(a) = 0.25$. Here y's expected utility from opting *in* equals $u_1^y(a, \gamma_1, \alpha_1) = 3 \cdot \alpha_1^y(c) + 0.25 \cdot C^y$; on the other hand, y's expected

utility from opting *out* equals $u_1^y(b, \gamma_1) = 1 + 0.75 \cdot C^y$. So y will prefer to opt in as long as $C^y \leq 6 \cdot \alpha_1(c) - 2$: this implies that y will never opt in if $\alpha_1(c) < \frac{1}{3}$; instead, if $\alpha_1(c) \geq \frac{1}{3}$ she will opt in on condition that C^y is weakly smaller than a certain value (i.e., the bonus one forgoes by not conforming to the majority must be less than the material payoff one might so obtain), with such value increasing with $\alpha_1(c)$ and ranging in the interval $[0, 4]$.

A (conformist) belief-revision process. We move on to define the belief revision triggered by our experimental manipulation. Before doing so, a clarification is in order. Later on we will comment on some relevant equilibrium conditions: there, we will follow the traditional justification of equilibrium as a theoretical characterization of a steady state (which is reached as individuals fine-tune their expectations on the basis of experience). Yet, we note that the analysis of equilibrium does not serve as a predictive device in the context of a one-shot game. In fact, *we do not assume that individuals' beliefs are correct in our experiment*. Indeed, for the current purposes (i.e., testing for a conformist attitude) it is not necessary to assume that beliefs are correct. So – in order to formulate our predictions about the impact of the treatment manipulations – we shall simply assume common knowledge of instrumental rationality, given the above-defined utility function and the belief-revision process below.

We begin by supposing some short-term idiosyncratic noise: in particular, we assume that each player starts with some (unexplained) prior belief about peer behavior (see the theory of fictitious play for a similar assumption; Fudenberg and Levine, 1998). Specifically, we assume mutual awareness of such noise in individual beliefs, but make no assumptions about the actual probability of error. We further suppose that – before anyone carries out a strategy – each player receives some aggregate information about the priors held by a sample of peers. Given that, our proposed belief-revision process says that *a player calculates a weighted average between her own prior and the prior held by the members of a sample of peers*, in such a way to attach equal weight to each individual.

We note that the presumed revision process is empirically justified by the social-psychology notion of conformism (i.e., one's tendency to follow the modal behavior and beliefs of one's peers). We also note that, given some short-term idiosyncratic noise in individual priors, the presumed revision process is normatively justified in that it leads to more accurate beliefs about peer behavior. In fact, the literature on the "wisdom of crowds" shows that averaging multiple individuals' predictions leads to better estimates than those of the average

33

individual (Galton, 1907; Clemen, 1989; Surowiecki, 2005): the basic intuition is that averaging cancels individual errors. Below we briefly define the wisdom-of-the-crowd effect for *any* forecasting problem and later apply it to our model.

Consider a group of n individuals (with generic member x), and let v^x denote the individual estimate of some unknown real quantity. Each individual estimate v^x can be defined as the sum of the true value T (i.e., the correct value of the variable to be predicted) and an individual error D^x (i.e., $D^x := v^x - T$). Given that, the wisdom-of-the-crowd effect says that *the average prediction of the group is at least as accurate as the individual prediction of the average group-member* (Manski, 2011). Research on forecasting typically uses absolute error as a measure of accuracy L – that is, $L(v^x, T) = |v^x - T|$ – since this specification equally punishes errors in either direction. When this or any other convex function (e.g., squared error) is used as a measure of accuracy, Jensen's inequality straightforwardly proves the wisdom-of-the-crowd effect (i.e., the convex transformation of a mean is less than or equal to the mean applied after convex transformation; for a proof, see Rudin, 1987). In a nutshell, for any set of real numbers, the absolute value of their mean is weakly less than the mean of their absolute values. Thus, the absolute value of the *mean estimate error* must be weakly less than the mean of the *absolute errors of the individual estimates*, i.e., $\left|\left(\frac{1}{n}\sum_{x=1}^{n} v^x - T\right)\right| \leq \frac{1}{n}\sum_{x=1}^{n}|(v^x - T)|$. This simple finding has long been known in statistical decision theory (see the discussion in Manski, 2011, p. 466).

With regard to our model – assuming some short-term idiosyncratic noise in individual priors – the result above suggests that one combine the exogenous information into a revised belief $\hat{\gamma}_i^x$ such as

$$\hat{\gamma}_i^x = \frac{1}{n} \cdot \gamma_i^x + \frac{n-1}{n} \cdot \bar{\gamma}_i, \qquad\qquad\qquad \text{(A3)}$$

where γ_i^x and $\bar{\gamma}_i$ respectively denote the *prior belief held by x* and the *mean prior belief held by the members of a sample of peers* (note that n denotes the cardinality of the set containing a

sample of peers plus individual x).[24] (For an experimental test of this revision process in a non-strategic environment, see Mannes, 2009.)

Finally – since in our setting a subject is not informed about n – we assume that individual x will try to guess the sample size. We therefore assume that x's estimate of $\hat{\gamma}_i^x$ is given by

$$\tilde{\gamma}_i^x = q^x \cdot \gamma_i^x + (1 - q^x) \cdot \bar{\gamma}_i, \tag{A4}$$

where q^x denotes x's expectation of random variable Q, with $Q \in \left(0, \frac{1}{2}\right]$. Note that the upper bound of Q is $\frac{1}{2}$ (i.e., a subject can attach such a large weight to her prior, only if she thinks that the sample of peers is singleton). It is clear that the revised belief approaches the exogenous information as the purported sample size increases.

Predicting the effect of the treatment manipulations—comparative statics. We turn to analyze the impact of the treatment manipulations on our binary trust game, using (A4) to estimate g_{s_i} in expression (A1) above. To that end, we present some comparative statics observations concerning scenarios where an individual *does* or *does not* face "conflicting" beliefs about peer behavior. Note that we say that γ_i and $\bar{\gamma}_i$ conflict if one implies that most peers take a strategy, while the other implies a different modal strategy. For each player-role $i \in \{1,2\}$ we focus on the following scenarios: *(i)* "scenario HL": $\gamma_i^x(\cdot) = 0.75$ and $\bar{\gamma}_i(\cdot) = 0.25$; *(ii)* "scenario HH": $\gamma_i^x(\cdot) = \bar{\gamma}_i(\cdot) = 0.75$; *(iii)* "scenario LH": $\gamma_i^x(\cdot) = 0.25$ and $\bar{\gamma}_i(\cdot) = 0.75$; *(iv)* "scenario LL": $\gamma_i^x(\cdot) = \bar{\gamma}_i(\cdot) = 0.25$.

Each observation below contrasts the "low treatment" against the "high treatment", holding prior beliefs constant. We begin by considering some individual x in the role of Trustee (i.e., $i = 2$). In particular, Observation 1 analyzes the impact of the exogenous information on a high-prior Trustee.

---

[24] We stress that the result above does *not* rest on the assumption that the sample is representative. Also, with reference to our experimental setting, we note that at the time of the belief transmission neither the experimenter nor any other individual could possibly know the true value to be estimated (because strategies had not been taken yet).

Observation 1. *Under scenario HH a Trustee with a sufficiently large C^x will share, whereas under scenario HL any such Trustee will keep.*

For ease of exposition, we have relegated the proof of this and the next observations to the end of Appendix A. Here is the interpretation. Consider such Trustees who would have shared, absent the conflicting information (i.e., Trustees with a sufficiently large C^x playing under scenario HH). As a result of the belief revision under scenario HL, many such Trustees will come to believe that "keep" is actually the more popular strategy: these Trustees can improve both their material payoff and overall utility by not sharing.[25] Therefore, the effect of the "low treatment" on high-prior Trustees is likely to be significant.

Observation 2. *Under scenario LL a Trustee will keep regardless of C^x; similarly, under scenario LH a Trustee will keep, unless her C^x is very large* (with the minimum threshold sharply increasing with the weight q the individual attaches to her prior).

Consider such Trustees who would have not shared, absent the conflicting information (e.g., Trustees with any value C^x playing under scenario LL). As a result of the belief revision under scenario LH, many such Trustees will come to believe that "share" is actually the more popular strategy: these Trustees will not improve their material payoff by sharing, so whether we should see a significant effect here depends on their degree of conformism. Also, we note that the larger is the weight one assigns to one's own prior, the larger is the psychological bonus one would require to find it attractive to conform to the high information.[26] This suggests that the effect of the "high treatment" on low-prior Trustees is likely to be negligible, unless subjects are extremely conformist or substantially discount their priors.

---

[25] That is true for any prior $0.50 < \gamma_2^x(c) \leq 0.75$, regardless of the Trustee's degree of conformism C^x. Furthermore, for priors $\gamma_2^x(c) > 0.75$, only a Trustee with an extremely large C^x would *not* find it attractive to conform to the low information: for instance, when $\gamma_2^x(c) = 0.76$ only a Trustee with a staggering $C^x \geq 300$ would prefer to share!

[26] For instance a Trustee who attaches weight $q = 0.25$ to her prior $\gamma_2^x(c) = 0.25$, upon receipt of the information $\bar{\gamma}_2(c) = 0.75$, would prefer to share only if $C^x \geq 12$ (i.e., if the bonus is more than twice the maximum attainable material payoff). Further, when $q = \frac{1}{3}$ one would share only if $C^x \geq 18$.

We now consider some individual x in the role of Trustor (i.e., $i = 1$). We begin by discussing the effect of the exogenous information on a Trustor holding a high prior $\gamma_1^x(a)$. (Note that in what follows we denote $\alpha_1^x(c)$ simply by α.)

Observation 3. *Under scenario HH, a Trustor with belief α larger than one third will opt in (Trustors with a large C^x will opt in regardless of α); similarly, under scenario HL, a Trustor with belief α larger than one third will opt in, unless her C^x is very large (i.e., larger than a value that sharply increases with both α and q).*

Consider such Trustors who would have opted in, absent the conflicting information (for instance Trustors holding a high belief α, with <u>any value</u> C^x, playing under scenario HH). As a result of the belief revision under scenario HL, many such Trustors will come to believe that "out" is actually the more popular strategy: these Trustors will not improve their expected material payoff by opting out if they expect more than a third of Trustees to share.[27] So, whether we should see a significant effect here depends on their degree of conformism. In short, the effect of the "low treatment" on high-prior Trustors may be negligible, unless subjects are extremely conformist or substantially discount their priors.[28]

Observation 4. *Under scenario LL, a Trustor will opt in on condition that her belief α is larger than one third and – at the same time – her C^x is quite small; by contrast, under scenario LH a Trustor will opt in if either α or C^x are sufficiently large.*

Consider such Trustors who would have opted out, absent the conflicting information (for instance Trustors holding a high belief α, and with a large value C^x, playing under scenario LL). As a result of the belief revision under scenario LH, many such Trustors will come to believe

---

[27] As we have assumed common knowledge of the functional form of the individual's utility, we do not make any particular assumption about the relationship between $\bar{\gamma}$ and α (e.g., a Trustor might think that her peers' decision to opt out is driven by their psychological bonus, and not by their expected material payoff). That said, in our experimental setting (where each participant first plays as a Trustee), subjects may well use the exogenous belief *from Part I* so as to inform $\alpha_1^x(c)$ *in Part II*. This in fact appears to be confirmed by model 5 of Table 2, where the beliefs from Part I are a significant predictor of behavior in Part II.

[28] For instance a Trustor with $\alpha = \frac{2}{3}$, who attaches weight $q = 0.25$ to her prior $\gamma_2^x(a) = 0.75$ upon receipt of the information $\bar{\gamma}_2(a) = 0.25$, will prefer to opt in as long as $C^x \leq 4$. Further, when $q = \frac{1}{3}$ one will opt in if $C^x \leq 6$.

that "in" is actually the more popular strategy: as a result of the belief revision, these Trustors can improve both their expected material payoff and overall utility by opting in, provided that they expect more than a third of Trustees to share. Since the conditions for a Trustor to opt in are less stringent when moving from LL to LH, this exercise suggests that the effect of the "high treatment" on low-prior Trustors is likely to be significant.

A note on equilibrium analysis. Equilibrium conditions are typically interpreted as theoretical characterizations of a steady state, that is reached as individuals fine-tune their expectations on the basis of experience. Experience may be thought of as being acquired by playing multiple instances of the game with participants from the same population. (So, if players observe everyone's strategies at the end of each game and eventually gather a great many observations, then beliefs will satisfy the relevant consistency conditions.) It is clear that such a characterization does not apply to our experiment. However, if one wished to explain behavior that has stabilized in a recurrent game, then equilibrium analysis would indeed be the relevant mode of inquiry. In that case, we note that an appropriate notion of "psychological equilibrium" could involve extending the standard requirements of the sequential equilibrium, by imposing that *each individual's belief about peer behavior be correct* (i.e., $\gamma_i^x(\cdot)$ must be derived from the frequency distribution of strategies chosen by members of $\mathcal{N}_i$; this implies that any two players will have the same prior and conditional beliefs about peers).[29] Another avenue for future research could involve extending the notion of incomplete-information rationalizability (Battigalli and Siniscalchi, 2003), in such a way to allow the analysis of behavior under incomplete information about the others' degrees of conformism.

---

[29] Battigalli and Dufwenberg's (2009) specification of psychological sequential equilibria extends the standard consistency requirement by imposing that higher-order beliefs at each information set be correct for all players and histories. We note that this condition is not necessary here, as second-order beliefs do not directly enter our utility function.

Proofs

Observation 1. Consider some individual x in the role of Trustee; also, suppose that $\gamma_2^x(c) = p$ and $\bar{\gamma}_2(c) = 0.75$. In this case x's expected utility from *sharing* (i.e., $s_2^x = c$) equals $u_2^x(c, \gamma_2) = 3 + C \cdot (p \cdot q + 0.75 \cdot (1 - q))$; on the other hand, x's expected utility from *keeping* (i.e., $s_2^x = d$) equals $u_2^x(d, \gamma_2) = 6 + C \cdot \left(1 - (p \cdot q + 0.75 \cdot (1 - q))\right)$. It follows that x will be indifferent between her two actions if $C = \frac{3}{2 \cdot p \cdot q - 1.5 \cdot q + 0.5}$. By contrast, assuming that $\bar{\gamma}_2(c) = 0.25$ a little algebra shows that in this case x will be indifferent if $C = \frac{3}{2 \cdot p \cdot q - 0.5 \cdot q - 0.5}$.

      Suppose that $\gamma_2^x(c) = 0.75$ and that x is informed that $\bar{\gamma}_2(c) = 0.25$ ("scenario HL"): in this case x would prefer to share if $C^x \geq \frac{3}{q - 0.5}$, given the constraint $q \in \left(0, \frac{1}{2}\right]$; note that these conditions cannot be jointly satisfied, which implies that x will prefer to keep regardless of C^x. Next, suppose that $\gamma_2^x(c) = 0.75$ and that x is informed that $\bar{\gamma}_2(c) = 0.75$ ("scenario HH"): in this case x will prefer to share as long as $C^x \geq 6$. *Therefore, under scenario HH a Trustee with a sufficiently large C^x will share, whereas under scenario HL any such Trustee will keep.*

Observation 2. Suppose that $\gamma_2^x(c) = 0.25$ and that x is informed that $\bar{\gamma}_2(c) = 0.75$ ("scenario LH"): in this case x will prefer to share if $C^x \geq \frac{3}{0.5 - q}$, given the constraint $q \in \left(0, \frac{1}{2}\right]$; that is, x will prefer to share on condition that C^x is sufficiently large, with the minimum threshold sharply increasing with q and ranging in the interval $(6, \infty)$. Next, suppose that $\gamma_2^x(c) = 0.25$ and that x is informed that $\bar{\gamma}_2(c) = 0.25$ ("scenario LL"): in this case x will prefer to keep regardless of C^x. *Therefore, under scenario LL a Trustee will keep regardless of C^x; similarly, under scenario LH a Trustee will keep, unless her C^x is very large* (with the minimum threshold sharply increasing with q).

Observation 3. Consider some individual x in the role of Trustor; also, suppose that $\alpha_1^x(c) = \alpha$, $\gamma_1^x(a) = p$, and $\bar{\gamma}_1(a) = 0.75$. In this case x's expected utility from opting *in* (i.e., $s_1^x = a$) equals $u_1^x(a, \gamma_1, \alpha_1) = 3 \cdot \alpha + C \cdot (p \cdot q + 0.75 \cdot (1 - q))$; instead, x's expected utility from opting *out* (i.e., $s_1^x = b$) equals $u_1^x(b, \gamma_1) = 1 + C \cdot \left(1 - (p \cdot q + 0.75 \cdot (1 - q))\right)$. It follows that x will be indifferent between her two actions if $C = \frac{1 - 3 \cdot \alpha}{2 \cdot p \cdot q - 1.5 \cdot q + 0.5}$. By contrast, assuming that $\bar{\gamma}_1(a) = 0.25$ a little algebra shows that in this case x will be indifferent if $C = \frac{1 - 3 \cdot \alpha}{2 \cdot p \cdot q - 0.5 \cdot q - 0.5}$.

      Suppose that $\gamma_1^x(a) = 0.75$ and that x is informed that $\bar{\gamma}_1(a) = 0.25$ ("scenario HL"): in this case x will prefer to opt in as long as $C^x \leq \frac{3 \cdot \alpha - 1}{0.5 - q}$, given the constraint $q \in \left(0, \frac{1}{2}\right]$. That is, if $\alpha < \frac{1}{3}$ these conditions cannot be jointly satisfied, which implies that x will prefer to opt out regardless of C^x; instead, if $\alpha \geq \frac{1}{3}$ she will opt in on condition that C^x is weakly smaller than a certain value, with such value sharply increasing with both α and q (and approaching infinity as q approaches $\frac{1}{2}$). Next, suppose that $\gamma_1^x(a) = 0.75$ and that x is informed that $\bar{\gamma}_1(a) = 0.75$ ("scenario HH"): here x will prefer to opt in as long as $C^x \geq 2 - 6 \cdot \alpha$; most notably, this implies that if $\alpha < \frac{1}{3}$ she will opt in on condition that C^x is sufficiently large; instead, if $\alpha \geq \frac{1}{3}$ she will opt in regardless of C^x. *Therefore, under scenario HH, a Trustor with belief α larger than one third will opt in (Trustors with a large C^x will opt in regardless of α); similarly, under scenario HL, a Trustor with belief α larger than one third will opt in, unless her C^x is very large* (i.e., larger than a value that sharply increases with both α and q).

Observation 4. Suppose that $\gamma_1^x(a) = 0.25$ and that x is informed that $\bar{\gamma}_1(a) = 0.75$ ("scenario LH"): here x will prefer to opt in as long as $C^x \geq \frac{3 \cdot \alpha - 1}{q - 0.5}$, given the constraint $q \in \left(0, \frac{1}{2}\right]$. That is, if $\alpha < \frac{1}{3}$ she will opt in on condition that C^x is sufficiently large, with the minimum threshold decreasing with α and increasing with q; instead, if $\alpha \geq \frac{1}{3}$ she will opt in regardless of C^x. Next, suppose that $\gamma_1^x(a) = 0.25$ and that x is informed that $\bar{\gamma}_1(a) = 0.25$ ("scenario LL"): here x will prefer to opt in if $C^x \leq 6 \cdot \alpha - 2$; this implies that if $\alpha < \frac{1}{3}$ she will prefer to opt out regardless of C^x; instead, if $\alpha \geq \frac{1}{3}$ she will opt in on condition that C^x is weakly smaller than a certain value, which such value ranging in the interval $[0, 4]$. *Therefore, under scenario LL, a Trustor will opt in on condition that her belief α is larger than one third and – at the same time – her C^x is quite small; by contrast, under scenario LH a Trustor will opt in if either α or C^x are sufficiently large.*

Appendix B

Procedure, experimental instructions and screenshots

The experiment was run with *zTree* (Fischbacher, 2007) in the ExpReSS Lab at Royal Holloway, University of London, between February and May 2012; subjects were recruited via emails forwarded across all faculties at Royal Holloway. A total of 209 subjects participated in the experiment; each session consisted of one of the three treatments (no subject could participate in more than one session). Each session took around 45 minutes and average earnings were £8 (including a £3 show-up fee), with minimum and maximum payments being £4 and £14, respectively. Paper instructions and transcripts of *zTree* screenshots are shown below (the main treatment is labeled as "T2").

Thank you for participating in this study.

Please note that it is prohibited to communicate with other participants during the experiment. If you have a question once the experiment has begun, please raise your hand and an assistant will come to your desk to answer it. Violation of this rule leads to immediate exclusion from the study and from all payments.

You will never learn the identity of the other participants, neither before nor after the study; and not one of the other participants will learn anything about your identity. Also, no other participant will learn what you earn during the experiment: upon completion of the session, the amount of money you will have earned will be paid out individually and privately. Hence, no other participant will know your choices and how much money you earn in this experiment.

You will receive £3 for participating in this session; additionally you also receive money depending on the decisions made (as described in the next paragraphs).

The experiment consists of two parts ("Part I" and "Part II"), each involving one simple decision task; your payment at the end of the session will be calculated as follows.

Your payment

= £3 (show-up fee) + any amount earned in Part I + any amount earned in Part II

In what follows we describe the procedure for Part I.

Part I

There are two types of participants, participants "A" and participants "B".

You will be assigned a type and paired with **one other participant** who was assigned another type than you.

This part consists of two steps, which you will perform with the particular participant you are paired with.

Step 1: Participant A must choose between the following two options. The <u>first</u> option ("OUT") gives a payout of £1 to both participants. The <u>second</u> option ("IN") is to instead transfer both pounds (i.e. £2 in total) to participant B and leave further decisions to him/her. If participant A transfers the 2 pounds to participant B, they will be tripled and participant B will receive 3 x 2 = 6 pounds.

Step 2: Only if participant A chooses the second option ("IN"), participant B will then decide if he/she transfers £3 back to participant A and keeps £3 for himself/herself OR if participant B keeps all the £6 for himself/herself.

Procedure for the two steps

Step one: Decision of participant A

It is up to participant A to choose one of the 2 options (OUT or IN): EITHER both participants receive £1 each OR the money and further decisions are transferred to participant B.

If participant A chooses the option OUT, both of you will receive £1. In this case participant B cannot change the payout allocation and the first part ends.

As a result

At the end of step one, there are two possible situations.

- If participant A has transferred the £2 to participant B (option IN), participant B has £6 and participant A has nothing.
- If participant A has chosen the option OUT, both of you have £1.

Step two: Decision of participant B

If participant A has transferred the money to participant B (option IN), then B receives £6 and it is now up to participant B to decide about the distribution of the £6 between the two participants. Participant B can EITHER:

- transfer £3 back to participant A and keep £3 for himself/herself

OR

- keep all the £6 for himself/herself and leave nothing to participant A.

After participant B's decision this part is completed and the earnings for both participants will be determined according to B's decision.

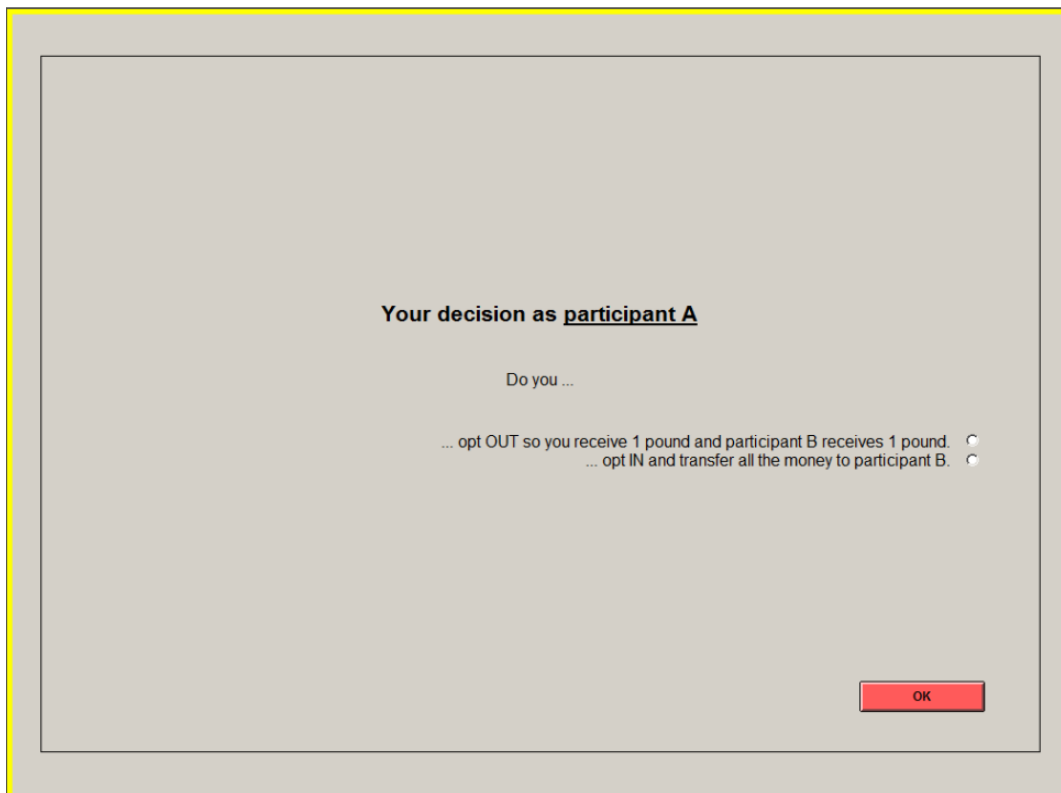The above information is summarised in the following table:

		A's income	B's income
A chose OUT		£1	£1
A chose IN	B keeps all	£0	£6
	B transfers half	£3	£3

Specific procedure and on-screen instructions for Part I

You are assigned the role of participant B

Note that you will complete the above-described two steps only once.

Step 1: Participant A decides by entering his/her choice on the screen shown below.

Your decision as participant A

Do you ...

... opt OUT so you receive 1 pound and participant B receives 1 pound. ○
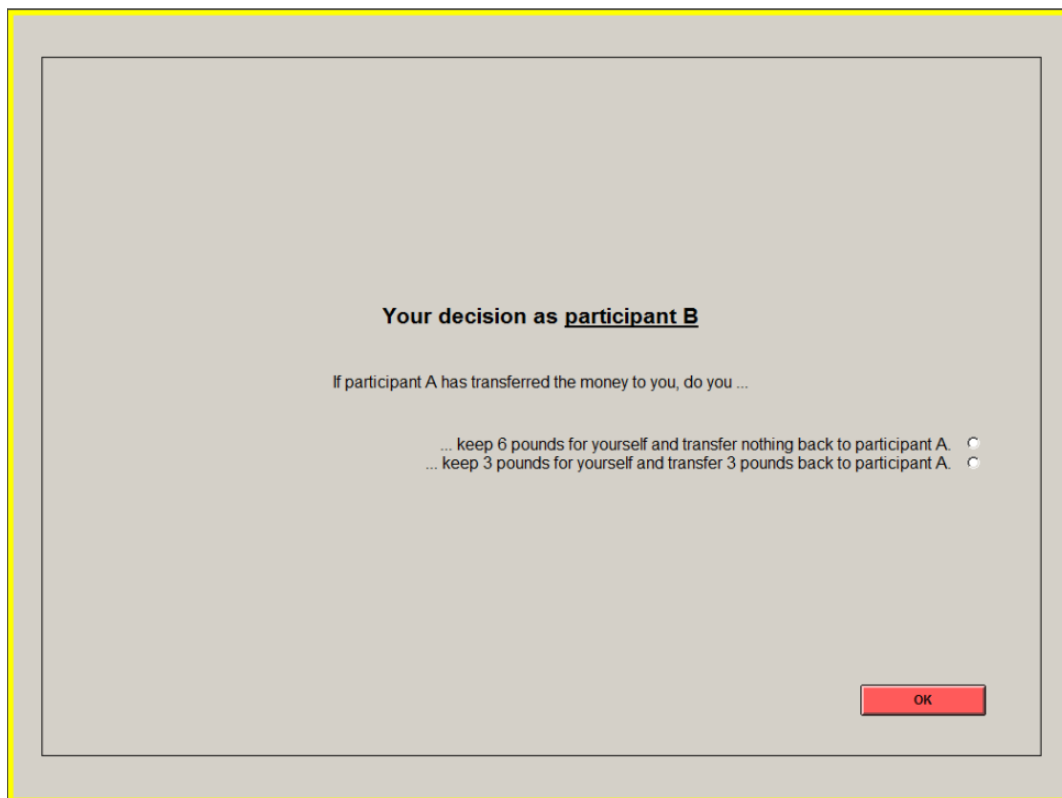... opt IN and transfer all the money to participant B. ○

OK

<u>Step 2</u>: We will ask <u>you</u> (participant B) how you would like to divide the £6 between participant A and yourself. Note that your answer will have an effect only if participant A does choose to transfer the money to you (option IN).

Participant A will not know your decision when he/she submits his/her own decision.

As explained above, **you decide on whether to transfer half the money to participant A or keep all the £6 for yourself.**

You will enter your choice on the following screen:

Your decision as <u>participant B</u>

If participant A has transferred the money to you, do you ...

... keep 6 pounds for yourself and transfer nothing back to participant A. ○
... keep 3 pounds for yourself and transfer 3 pounds back to participant A. ○

OK

Control questions

Please answer the following control questions. Please contact the study organizer if you have any questions.

1. Participant A has chosen IN. You then choose to transfer half the money back to participant A.
   What is the income of participant A? …………….
   What is the income of participant B (yourself)?…………….

2. Participant A has chosen IN. You then choose to keep all the money for yourself.
   What is the income of participant A? …………….
   What is the income of participant B (yourself)?…………….

3. Participant A has chosen OUT.
   What is the income of participant A? …………….
   What is the income of participant B (yourself)?…………….

Please feel free to ask questions at any point if you feel you need some clarification. Please do so by raising your hand.

We will start with Part I once the instructions are clear to everyone. Are there any questions?

Part II

We are now ready to undergo the last part of the study. This part has exactly the **same two-step procedure as in Part I**.

The payouts are the same as before and are summarised in the following table:

		A's income	B's income
A chose OUT		£1	£1
A chose IN	B keeps all	£0	£6
	B transfers half	£3	£3

The only difference is that you are assigned a different type in this part than in the previous part.

You are now assigned the role of <u>participant A</u>.

Again, you will be paired with one other participant. **This other participant will be a different person than the one you were paired with in Part I**.

Please refer to your paper handout or ask an assistant if you need reminding of the procedure.

[Transcript of on-screen messages]

<u>Treatments T0-T1</u>

Screen 1 (Part I)

<u>You are assigned the role of participant B</u>

Prior to entering your decision as participant B, we would like to know what you think of the other participants who have been assigned the same type as you (i.e. participants B).

In other words, we ask you to guess how many of today's participants B (excluding yourself) will choose to transfer half the money back, and how many of today's participants B will keep all the money for themselves.

Please enter your guess by positioning the below slider to the desired percentage.

[The below line is only for treatment T1.]

Note: You can earn some additional income if your guess is correct. If your guess differs by no more than 5 percentage points from the realized value, at the end of the study you will receive an additional payment of £2. Otherwise, you do not receive an additional income.

Screen 2 (Part I)

Enter 2nd mover decision.

Screen 3 (Part II)

Insert instructions for Part II here.

Screen 4 (Part II)

<u>You are assigned the role of participant A</u>

Prior to entering your decision as participant A, we would like to know what you think of the other participants who have been assigned the same type as you (i.e. participants A).

In other words, we ask you to guess how many of today's participants A (excluding yourself) will choose IN, and how many of today's participants A will choose OUT.

Please enter your guess by positioning the below slider to the desired percentage.

[The below line is only for treatment T1.]

Note: You can earn some additional income if your guess is correct. If your guess differs by no more than 5 percentage points from the realized value, at the end of the study you will receive an additional payment of £2. Otherwise, you do not receive an additional income.

Screen 5 (Part II)

Enter 1st mover decision.

Screen 6

Outcome.

Treatment T2

Screen 1 (Part I)

You are assigned the role of participant B

Prior to entering your decision as participant B, we would like to know what you think of the other participants who have been assigned the same type as you (i.e. participants B).

In other words, we ask you to guess how many of today's participants B (excluding yourself) will choose to transfer half the money back, and how many of today's participants B will keep all the money for themselves.

************first lower part of screen 1*****

Please enter your guess by positioning the below slider to the desired percentage.

************second lower part of screen 1 [to appear after subjects have entered their guesses]*****

A sample of other participants B in this session expects on average that <x>% will transfer half the money, whereas <100-x>% will keep all the money.

TRANSFER HALF: x%

KEEP: (100-x)%

Screen 2 (Part I)

Enter 2nd mover decision.

Screen 3 (Part II)

Insert instructions for part II here.

Screen 4 (Part II)

You are assigned the role of participant A

Prior to entering your decision as participant A, we would like to know what you think of the other participants who have been assigned the same type as you (i.e. participants A).

In other words, we ask you to guess how many of today's participants A (excluding yourself) will choose IN, and how many of today's participants A will choose OUT.

************first lower part of screen 4*****
Please enter your guess by positioning the below slider to the desired percentage.

************second lower part of screen 4[to appear after subjects have entered their guesses]*****
A sample of other participants A in this session expects on average that <x>% will OPT IN, whereas <100-x>% will OPT OUT.
IN: x%
OUT: (100-x)%

Screen 5 (Part II)

Enter 1st mover decision.

Screen 6

Outcome.

VII. References

Akerlof, George A. 1980. "A Theory of Social Custom, of Which Unemployment May Be One Consequence" *Quarterly Journal of Economics*, 94(4): 749-775.

Andreoni, James and Alison Sanchez. 2014. "Do Beliefs Justify Actions or Do Actions Justify Beliefs? An Experiment on Stated Beliefs, Revealed Beliefs, and Social-Image Manipulation" *NBER Working Paper*, No. 20649.

Asch, Solomon E. 1956. "Studies of Independence and Conformity: A Minority of One Against A Unanimous Majority" *Psychological Monographs*, 70(9): 1-70.

Attanasi, Giuseppe, Pierpaolo Battigalli and Elena Manzoni. 2016. "Incomplete-Information Models of Guilt Aversion in The Trust Game" *Management Science*, 62(3): 648-667.

Bacharach, Michael, Gerardo Guerra and Daniel J. Zizzo. 2007. "The Self-Fulfilling Property of Trust: An Experimental Study" *Theory and Decision*, 63(4): 349-388.

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

Battigalli, Pierpaolo and Martin Dufwenberg. 2007. "Guilt in Games" *American Economic Review P&P*, 97(2): 170-176.

Battigalli, Pierpaolo and Martin Dufwenberg. 2009. "Dynamic Psychological Games" *Journal of Economic Theory*, 144(1): 1-35.

Battigalli, Pierpaolo and Marciano Siniscalchi. 1999. "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games" *Journal of Economic Theory*, 88(1): 188-230.

Battigalli, Pierpaolo and Marciano Siniscalchi. 2003. "Rationalization and Incomplete Information" *Advances in Theoretical Economics*, 3(1).

Berg, Joyce, John Dickhaut and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History" *Games and Economic Behavior*, 10(1): 122-142.

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

Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian and Katherine L. Milkman. 2015. "The Effect of Providing Peer Information on Retirement Savings Decisions" *Journal of Finance*, 70(3): 1161-1201.

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

Bicchieri, Cristina and Alessandro Sontuoso. 2015. "I Cannot Cheat on You After We Talk: Communication and Norms in Mixed-Motive Games" in *The Prisoner's Dilemma*, ed. M. Peterson. Cambridge University Press.

Bicchieri, Cristina and Erte Xiao. 2009. "Do The Right Thing: But Only if Others Do So" *Journal of Behavioral Decision Making*, 22(2): 191-208.

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

Blanco, Mariana, Dirk Engelmann, Alexander Koch and Hans-Theo Normann. 2014. "Preferences and Beliefs in A Sequential Social Dilemma: A Within-Subjects Analysis" *Games and Economic Behavior*, 87(C): 122-135.

Bolton, Gary E. and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition" *American Economic Review*, 90(1): 166-193.

Brandts, Jordi and Gary Charness. 2011. "The Strategy Versus The Direct-Response Method: A First Survey of Experimental Comparisons" *Experimental Economics*, 14(3): 375-398.

Burnham, Kenneth P. and David R. Anderson. 2002. *Model Selection and Multimodel Inference*. New York: Springer-Verlag.

Charness, Gary and Martin Dufwenberg. 2006. "Promises and Partnership" *Econometrica*, 74(6): 1579-1601.

Charness, Gary and Ernst Fehr. 2015. "From The Lab to The Real World" *Science*, 350(6260): 512-513.

Charness, Gary and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests" *Quarterly Journal of Economics*, 117(3): 817-869.

Charness, Gary, Luca Rigotti and Aldo Rustichini. 2017. "Social Surplus Determines Cooperation Rates in The One-Shot Prisoner's Dilemma" *Games and Economic Behavior*, 100: 113-124.

Cialdini, Robert B. and Noah J. Goldstein. 2004. "Social Influence: Compliance and Conformity" *Annual Review of Psychology*, 55(1): 591-621.

Cialdini, Robert B. and Melanie R. Trost. 1998. "Social Influence: Social Norms, Conformity and Compliance" in *The Handbook of Social Psychology*, ed. D. T. Gilbert, S. T. Fiske and G. Lindzey. Boston: McGraw-Hill.

Clemen, Robert T. 1989. "Combining Forecasts: A Review and Annotated Bibliography" *International Journal of Forecasting*, 5(4): 559-583.

Cooper, David J. and John H. Kagel. 2016. "Other-Regarding Preferences" in *The Handbook of Experimental Economics, Vol. 2*, ed. J. Kagel and A. Roth. Princeton University Press.

Costa, Dora L. and Matthew E. Kahn. 2013. "Energy Conservation 'Nudges' and Environmentalist Ideology: Evidence from A Randomized Residential Electricity Field Experiment" *Journal of the European Economic Association*, 11(3): 680-702.

Costa-Gomes, Miguel A., Steffen Huck and Georg Weizsäcker. 2014. "Beliefs and Actions in The Trust Game: Creating Instrumental Variables to Estimate The Causal Effect" *Games and Economic Behavior*, 88: 298-309.

Dawes, Robyn M. 1989. "Statistical Criteria for Establishing A Truly False Consensus Effect" *Journal of Experimental Social Psychology*, 25(1): 1-17.

Deutsch, Morton and Harold B. Gerard. 1955. "A Study of Normative and Informational Social Influences upon Individual Judgment" *Journal of Abnormal and Social Psychology*, 51(3): 629-636.

Dufwenberg, Martin and Uri Gneezy. 2000. "Measuring Beliefs in an Experimental Lost Wallet Game" *Games and Economic Behavior*, 30(2): 163-182.

Dufwenberg, Martin and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity" *Games and Economic Behavior*, 47(2): 268-298.

Ellingsen, Tore, Magnus Johannesson, Sigve Tjøtta and Gaute Torsvik. 2010. "Testing Guilt Aversion" *Games and Economic Behavior*, 68(1): 95-107.

Engelmann, Dirk and Martin Strobel. 2000. "The False Consensus Effect Disappears if Representative Information and Monetary Incentives Are Given" *Experimental Economics*, 3(3): 241-260.

Falk, Armin and Urs Fischbacher. 2006. "A Theory of Reciprocity" *Games and Economic Behavior*, 54(2): 293-315.

Falk, Armin and Andrea Ichino. 2006. "Clean Evidence on Peer Effects" *Journal of Labor Economics*, 24(1): 39-57.

Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation" *Quarterly Journal of Economics*, 114(3): 817-868.

Festinger, Leon. 1954. "A Theory of Social Comparison Processes" *Human Relations*, 7(2): 117-140.

Fischbacher, Urs. 2007. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments" *Experimental Economics*, 10(2): 171-178.

Fudenberg, Drew and David K. Levine. 1998. *The Theory of Learning in Games, Vol. 2*. Cambridge, MA: MIT Press.

Gächter, Simon, Daniele Nosenzo and Martin Sefton. 2013. "Peer Effects in Pro-Social Behavior: Social Norms or Social Preferences?" *Journal of the European Economic Association*, 11(3): 548-573.

Galton, Francis. 1907. "Vox Populi (The Wisdom of Crowds)" *Nature*, 75(7): 450-451.

Guerra, Gerardo and Daniel J. Zizzo. 2004. "Trust Responsiveness and Beliefs" *Journal of Economic Behavior & Organization*, 55(1): 25-30.

Herbst, Daniel and Alexandre Mas. 2015. "Peer Effects on Worker Output in The Laboratory Generalize to The Field" *Science*, 350(6260): 545-549.

Ho, Teck-Hua and Xuanming Su. 2009. "Peer-Induced Fairness in Games" *American Economic Review*, 99(5): 2022-49.

Huck, Steffen, Dorothea Kübler and Jörgen Weibull. 2012. "Social Norms and Economic Incentives in Firms" *Journal of Economic Behavior & Organization*, 83(2): 173-185.

Kandel, Eugene and Edward P. Lazear. 1992. "Peer Pressure and Partnerships" *Journal of Political Economy*, 100(4): 801-817.

Mannes, Albert E. 2009. "Are We Wise About The Wisdom of Crowds? The Use of Group Judgments in Belief Revision" *Management Science*, 55(8): 1267-1279.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem" *Review of Economic Studies*, 60(3): 531-542.

Manski, Charles F. 2011. "Interpreting and Combining Heterogeneous Survey Forecasts" in *The Oxford Handbook of Economic Forecasting*, ed. M. P. Clements and D. F. Hendry. Oxford University Press.

Mas, Alexandre and Enrico Moretti. 2009. "Peers at Work" *American Economic Review*, 99(1): 112-145.

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

Ross, Lee, David Greene and Pamela House. 1977. "The 'False Consensus Effect': An Egocentric Bias in Social Perception and Attribution Processes" *Journal of Experimental Social Psychology*, 13(3): 279-301.

Rudin, Walter. 1987. *Real and Complex Analysis*. New York: McGraw-Hill.

Schram, Arthur and Gary Charness. 2015. "Inducing Social Norms in Laboratory Allocation Choices" *Management Science*, 61(7): 1531-1546.

Shue, Kelly. 2013. "Executive Networks and Firm Policies: Evidence from The Random Assignment of MBA Peers" *Review of Financial Studies*, 26(6): 1401-1442.

Sugden, Robert. 2000. "The Motivating Power of Expectations" in *Rationality, Rules and Structure*, ed. J. Nida-Rümelin and W. Spohn. Amsterdam: Kluwer.

Surowiecki, James. 2005. *The Wisdom of Crowds*. New York: Anchor Books.

Thöni, Christian and Simon Gächter. 2014. "Peer Effects and Social Preferences in Voluntary Cooperation" *CESifo Working Paper Series* 4741, CESifo Group Munich.

Xiao, Erte and Cristina Bicchieri. 2010. "When Equality Trumps Reciprocity" *Journal of Economic Psychology*, 31(3): 456-470.