



## City Research Online

### City, University of London Institutional Repository

---

**Citation:** Bicchieri, C. & Sontuoso, A. (2015). I cannot cheat on you after we talk. In: The Prisoner's Dilemma. (pp. 101-114). Cambridge, UK: Cambridge University Press. ISBN 9781107044357

This is the accepted version of the paper.

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

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/34898/>

**Link to published version:**

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

---

---

City Research Online:

<http://openaccess.city.ac.uk/>

[publications@city.ac.uk](mailto:publications@city.ac.uk)

---

# I Cannot Cheat on You after We Talk

Cristina Bicchieri and Alessandro Sontuoso

*University of Pennsylvania*

July 1, 2015

Experimental evidence on pre-play communication supports a “focusing function of communication” hypothesis. Relevant communication facilitates cooperative, pro-social behavior because it causes a shift in individuals’ focus towards strategies dictated by some salient social norm. After reviewing the formal foundations for a general theory of conformity to social norms, we provide an original application illustrating how a framework that allows for different conjectures about norms is able to capture the focusing function of communication and to explain experimental results.

## 1. Introduction

The experimental literature on social dilemmas has long documented the positive effect of communication on cooperation. Sally (1995), in a meta-analysis spanning thirty-five years of Prisoner’s Dilemma experiments, shows that the possibility of communicating significantly increases cooperation. Social psychologists have explained such a finding by hypothesizing that the act of communicating contributes to promoting trust by creating empathy among participants. See Loomis (1959), Desforges et al. (1991), Davis and Perkowski (1979). Bicchieri (2002, 2006), in a different perspective, puts forward a *focusing function of communication* hypothesis, according to which communication can focus agents on shared rules of behavior and – when it does focus them on pro-social ones – generates a normative environment which is conducive to cooperation. More specifically, when individuals face an unfamiliar situation, they need cues to understand how to best act and, for this reason, they check whether some behavioral rule *they are aware of* applies to the specific interaction. The effect of communication is to make a behavioral rule *situationally salient*, that is, communication causes a shift in an individual’s focus towards the strategies dictated by the now-salient rule. In doing so, communication also coordinates players’ mutual expectations about which strategies will be chosen by the parties. In other words, (under some conditions) communication elicits social norms.

While a large proportion of studies on the effect of pre-play communication focuses on Prisoner’s Dilemma games, Bicchieri, Lev-On and Chavez (2010) examine behavior in

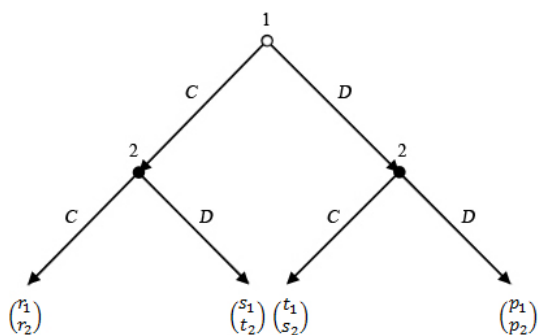
sequential trust games. In what follows we shall look at those findings, discuss an interpretation based on the above hypothesis, and suggest a theoretical application that can account for it. Given that our analysis equally applies to the Prisoner's Dilemma, this essay contributes to the broad literature on social dilemmas by proposing an application for dynamic interactions.<sup>1</sup>

Bicchieri (2002) provides the basis for the focusing function of communication argument: when a rule of behavior becomes situationally salient, it causes a shift in an individual's focus, thereby generating empirical and normative expectations that direct one's actions. Before we elaborate on the argument, it is convenient to summarize Bicchieri's conditions for a social norm to exist and be followed.<sup>2</sup> A social norm exists and is followed by a population if two conditions are satisfied: First, every individual must be aware that she is in a situation in which a particular rule of behavior applies ("contingency" clause). Second, every individual prefers to conform to it, on the double condition ("conditional preference" clause) that: (i) she believes that most people conform to it (i.e. *empirical expectations* condition), and (ii) she believes that most people believe she ought to conform to it (i.e. *normative expectations* condition).

In order to develop an equilibrium model that can capture more precisely the variables the experimenter manipulates in a laboratory environment, Bicchieri's model can be integrated with *psychological game theory* (Geanakoplos et al. 1989, Battigalli and Dufwenberg 2009). Such integration allows to explicitly formalize the impact of the above

---

<sup>1</sup> It should be noted that a dichotomous Trust Game can be thought of as a special version of the sequential Prisoner's Dilemma where payoffs are given as follows:



with  $r_1 > t_1 = p_1 > s_1$  and  $t_2 > r_2 > p_2 = s_2$ . (Instead, in a Prisoner's Dilemma payoffs satisfy the following inequality:  $t_i > r_i > p_i > s_i, \forall i \in \{1,2\}$ .)

<sup>2</sup> Bicchieri (2006: 11).

conditions on a player's utility by incorporating conjectures about norms into the expected utility of a conformist player (Sontuoso 2013).

This essay draws on the above theoretical treatments and provides an application illustrating how a formal framework that allows for different conjectures about norms is able to capture the focusing function of communication and to explain experimental results. In sum, the core of the argument is that *communication can focus people on some behavioral rule that is relevant to the specific interaction*. In so doing, it coordinates players' mutual expectations about which strategies will be played. So – if the aforementioned contingency condition holds – one may assume that making a behavioral rule salient through communication will result in greater compliance with the rule (Cialdini et al. 1991).

The remainder of the essay is organized in this manner: Section 2 discusses Bicchieri, Lev-On and Chavez's (2010) experimental results on the effect of communication in trust games; Section 3 briefly reviews models of norm compliance; Section 4 provides an application accounting for the focusing function of communication, and Section 5 draws conclusions.

## **2. Experimental evidence**

Bicchieri, Lev-On and Chavez (2010) study two features of pre-play communication in trust games: *content relevance* of the message (i.e. relevant vs. irrelevant communication) and *media richness* (i.e. face-to-face, "FtF", vs. computer-mediated communication, "CMC").

Consider the following trust game: the investor (first-mover) receives \$6 and can choose to send any discrete amount  $x$  to the trustee (second-mover); the amount the trustee receives is tripled by the experimenter, so that the trustee can then send any discrete amount  $y$  in the interval  $[0, 3x]$  back to the investor.

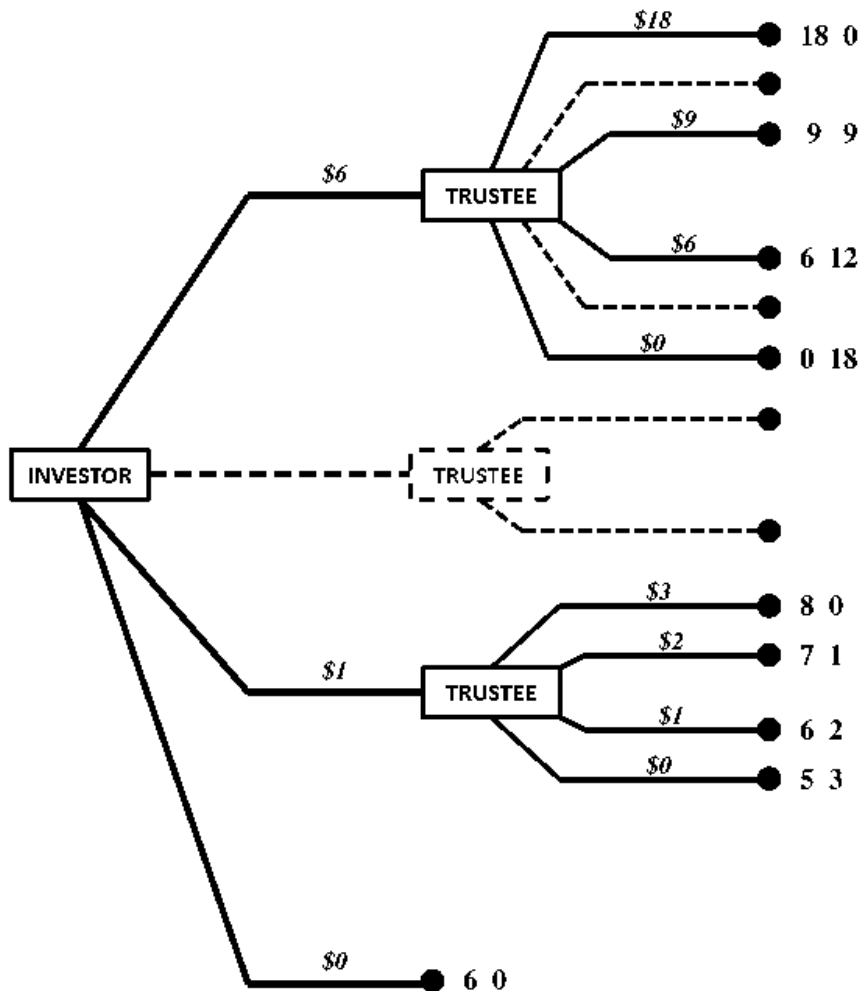


Figure 1. The trust game of Bicchieri, Lev-On and Chavez (2010)

In the experiment participants were paired randomly and played three variants of the above game, each time with a different subject, in the following order: (i) no-communication game [i.e. the baseline condition just described]; (ii) irrelevant or relevant CMC communication game; (iii) irrelevant or relevant FtF communication game.<sup>3</sup> Investors did not receive feedback on the amount that the trustee returned until the end of the experimental session; also, after making their decision – in each variant of the game – investors were asked to report their expectation about the amount returned by the trustee. The authors were interested in three dependent variables: trust (defined as the amount of dollars sent by the

<sup>3</sup> In the *CMC* conditions subjects could communicate via computer-based text chat for five minutes, whereas in the *FtF* conditions subjects communicated face-to-face for two minutes. In the *irrelevant* conditions subjects were instructed that they could discuss only the questions given by the experimenter (about a completely irrelevant topic), whereas in the *relevant* conditions they could discuss any topic except their identities. (Roughly half of the experimental sessions featured the relevant conditions while the remaining sessions featured the irrelevant conditions.)

investor), reciprocity (the amount returned by the trustee), and expectation (the amount the investor expected to get back).

Table 1 shows the participants' average responses across the five combinations of relevance and medium: a first look at the table reveals that both relevance and medium had large, positive effects on the three dependent variables.

|                                 | Control<br>(N=32) | Ftf-Relevant<br>(N=14) | CMC-<br>Relevant<br>(N=14) | FtF-Irrelevant<br>(N=18) | CMC-<br>Irrelevant<br>(N=18) |
|---------------------------------|-------------------|------------------------|----------------------------|--------------------------|------------------------------|
| <b>Trust</b>                    | 2.63 (0.36)       | 5.57 (0.46)            | 5.14 (0.57)                | 4.17 (0.49)              | 3.28 (0.61)                  |
| <b>Reciprocity</b>              | 1.92 (0.48)       | 7.57 (0.96)            | 5.14 (1.33)                | 3.33 (1.05)              | 1.94 (0.78)                  |
| <b>Expected<br/>reciprocity</b> | 3.54 (0.53)       | 8.36 (0.69)            | 7.43 (0.96)                | 5.56 (0.91)              | 4.28 (0.93)                  |

Table 1. Mean (SEM) of trust, reciprocity, and investor's expectations by communication relevance and medium (N = 96). FtF = Face-to-face; CMC = Computer-mediated communication.

Note that *relevant face-to-face* communication had the largest effects on all three variables while *relevant computer-mediated* communication had the second largest effects. Figure 2 below illustrates the distribution of trust across the five conditions.

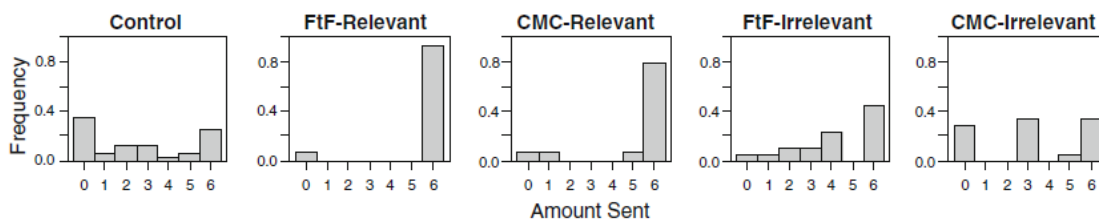


Figure 2. Distribution of trust by communication medium and relevance.

As shown above, investors were most trusting in both *relevant* communication conditions (where the majority sent their entire \$6 endowment).

The authors' further data analysis discloses the following key points: (i) the behavior of investors was strongly determined by their *expectations of trustees' reciprocation*;<sup>4</sup> (ii) the variable most conducive to creating such expectations was not the medium, but rather the *content relevance* of the message (i.e. investments were significantly higher following

<sup>4</sup> Note, however, that those expectations were rarely met, since expected reciprocation was significantly higher than the actual reciprocation across conditions (except when \$6 was invested).

unrestricted communication than restricted or no communication and – whenever communication was restricted to irrelevant topics – there were no significant differences between the amounts sent in the CMC, FtF, and control conditions); (iii) reciprocity significantly increased with *trust*, *content relevance* of the message, and *medium* (more precisely, reciprocity was higher in the CMC condition for lower amounts of trust but became higher in the FtF condition for higher amounts of trust).<sup>5</sup>

How do such results relate to the two explanations for the effect of communication on social dilemmas that we mentioned in the introduction (i.e. one explanation maintains that communication enhances cohesion and group identity while the other asserts that communication elicits social norms)? The data seem to provide evidence in favor of the latter explanation: in fact, if the former were valid, then one should not find an effect of content relevance on the expected reciprocation (point (ii) above). On the other hand, Bicchieri’s focus theory of norms is consistent with the data, since it predicts an effect of the message content relevance on the expected reciprocation. Specifically, Bicchieri (2002, 2006) hypothesizes that, when participants are allowed to talk about the strategic situation at hand, the discussion on “how to appropriately behave” will lead the participants to become aware of the fact that the current interaction is one to which some default rule of behavior applies.<sup>6</sup> Hence, focusing subjects on a rule of behavior generates and coordinates empirical and normative expectations.

### 3. Theoretical foundations

In what follows we review two formal, theoretical treatments of norms and subsequently draw on them to develop an application accounting for the experimental results.

Bicchieri (2006: 52) proposes a general utility function based on norms. Considering an  $n$ -player normal form game, let  $S_i$  denote the strategy set of Player  $i$  and  $S_{-i} = \prod_{j \neq i} S_j$  be the set of strategy profiles of players other than  $i$ . A norm  $N_i$  is defined as a (set-valued)

---

<sup>5</sup> For instance, as the amount that the investor sent approached zero, the odds that the trustee returned each available dollar were over seven times higher in CMC than in FtF. With each additional dollar that the investor sent, however, the odds that the trustee reciprocated increased more rapidly in FtF conditions. In other words, the probability of returning each available dollar increased with the amount invested, but increased more rapidly for the FtF and control conditions than for CMC.

<sup>6</sup> See Lev-On et al. (2010) for the effect of group (vs. dyadic) communication in trust games.



function from one's expectation about the opponents' strategies to the "strategies one ought to take", that is,  $N_i: L_{-i} \rightarrow S_i$ , with  $L_{-i} \subseteq S_{-i}$ .<sup>7</sup> A strategy profile  $s = (s_1, \dots, s_n)$  is said to *instantiate* a norm for Player  $j$  if  $s_{-j} \in L_{-j}$  (i.e. if  $N_j$  is defined at  $s_{-j}$ ), and to *violate* a norm if, for some  $j$ , it instantiates a norm for  $j$  but  $s_j \neq N_j(s_{-j})$ . Player  $i$ 's utility function is a linear combination of  $i$ 's material payoff  $\pi_i(s)$  and a component that depends on norm compliance:

$$(1) \quad U_i(s) = \pi_i(s) - k_i \max_{s_{-j} \in L_{-j}} \max_{m \neq j} \{ \pi_m(s_{-j}, N_j(s_{-j})) - \pi_m(s), 0 \}$$

where  $k_i \geq 0$  represents  $i$ 's sensitivity to the norm and  $j$  refers to the norm violator. The norm-based component represents the maximum loss (suffered by players other than the norm violator  $j$ ) resulting from all norm violations: the first maximum operator takes care of the possibility that there might be multiple rule-complying strategy profiles; the second maximum operator ranges over all the players other than the norm violator  $j$ . Bicchieri's utility function makes it possible for the experimenter to test whether subjects' behavior is consistent with preferences for conformity to a social norm, given that the above-mentioned conditions for the existence of a social norm are satisfied (i.e. contingency, and preferences conditional on the relevant empirical and normative expectations; see Bicchieri 2006). Specifically, this utility function captures conformist preferences in case a norm exists. Hence, the norm-based component of the utility function represents the maximum loss resulting from all violations of an *established* norm.

Bicchieri's utility function makes very sharp predictions in cases where there is no ambiguity about subjects' expectations as to what the norm prescribes. In order to explicitly represent conditionally conformist preferences in dynamic games where multiple rules of behavior may apply, Sontuoso (2013) extended Bicchieri's framework to a "psychological" utility function and a belief-based formulation of her conditions for norm existence (which are directly reflected into the player's utility function). Given that in Bicchieri's theory of norms *expectations are crucial to compliance*, having a model of how subjects derive their

---

<sup>7</sup> For example, in an  $n$ -player Prisoner's Dilemma a shared norm may be to cooperate: in that case,  $L_{-i}$  includes the *cooperate* strategies of all players other than  $i$ . Note that in the case where – given the others' strategies – there is not a norm prescribing how Player  $i$  should behave, then  $N_i$  is not defined at  $L_{-i}$ .

conjectures about norms may be useful for interpreting the experimental results of Bicchieri, Lev-On and Chavez (2010).

Before outlining such a model of norm compliance, we shall introduce some notation on dynamic games: let an extensive form game be given by  $\langle N, H, P, (I_i)_{i \in N} \rangle$ , where  $N$  is the set of players,  $H$  is the set of feasible histories,  $I_i$  is the information partition of Player  $i$ . Further, let  $Z$  denote the set of terminal histories, with  $H \setminus Z$  being the set of non-terminal histories; given that, let  $P$  denote the player function (which assigns to each element of  $H \setminus Z$  an element of  $N$ ), and let  $A_i(h)$  denote the set of feasible actions for Player  $i$  at history  $h$ .<sup>8</sup> The *material payoffs* of players' strategies are described by functions  $m_i: Z \rightarrow \mathbb{R}$  for each player  $i \in N$ . Then, denote the set of Player  $i$ 's pure strategies allowing history  $h$  as  $S_i(h)$ ; strategy profiles allowing history  $h$  are defined as  $S(h)$ , and  $S_{-i}(h)$  for all players  $j$  other than  $i$ ; given that, let  $z(s)$  indicate a terminal history induced by some strategy profile  $s \in S$ . Battigalli and Dufwenberg (2009) provide a framework for the analysis of *dynamic psychological games* where conditional higher-order systems of beliefs influence players' preferences: as in their model, here it is assumed that (at each history) every player holds an updated system of first-order beliefs  $\alpha_i = (\alpha_i(\cdot | h))_{h \in H_i}$  about the strategies of all the co-players;<sup>9</sup> at each history Player  $i$  further holds a system of second-order beliefs  $\beta_i$  about the first-order belief system of each of the opponents.<sup>10</sup>

The model of norm compliance we employ (Sontuoso 2013) assumes there exists a set of default rules of behavior, where each rule specifies strategies appropriate to generic (mixed-motive) games; players have a subset of such rules stored in their minds and derive from them "norm-conjectures" (i.e. expectations as to which rule-driven strategies apply to the current game). Therefore, if an individual  $j$  is a norm-driven player – and presumes that her co-players are norm-driven too – she can form her first-order belief  $\alpha_j$  by assuming her co-players' behavior to be consistent with some rule. A "*behavioral rule*" is defined as a set-valued function  $r$  that assigns to every non-terminal history  $h$  one or more elements from

---

<sup>8</sup> Note that a node of the game tree is identified with the history leading up to it (i.e. a path in the game tree) as in Osborne and Rubinstein (1994).

<sup>9</sup> For example, in a game with perfect information, at each  $h \in H_i$  Player  $i$  holds an updated belief  $\alpha_i(\cdot | h)$  such that she believes that all players have chosen all the actions leading to  $h$  with probability 1.

<sup>10</sup> It is assumed that players' beliefs at different information sets must satisfy Bayes' rule and common knowledge of Bayesian updating.

the set of strategy profiles  $S(h)$ .<sup>11</sup> The *set of behavioral rules* is denoted by  $R$ , and the *behavioral rule subset of Player  $i$*  by  $R_i$  (with  $R_i \subseteq R$ ), with  $R_i$  representing the set of rules  $i$  is aware of. Further, given a game  $G$  and some rule  $\hat{r}$ , one can derive the *set of strategy profiles dictated by  $\hat{r}$*  (e.g. the set of strategy profiles dictated by  $\hat{r}$ , when evaluated at the initial history, is denoted by  $\hat{r}(h^0)$ ) and, given  $R_i \subseteq R$ , one can derive the *set of Player  $i$ 's "rule-complying" actions at history  $h$*  (e.g. denoted by  $A_{i,h}(R_i(h^0))$ ), which depicts the set of actions prescribed – by any of the rules  $r \in R_i$  – to Player  $i$  at history  $h$ .

Finally, a “*norm-conjecture*” of Player  $i$  is defined as a collection of independent probability measures  $\rho_i = (\rho_i(\cdot | h))_{h \in H \setminus Z}$  (with  $\rho_i(a|h)$  being the probability of action  $a$  at history  $h$ ) such that *the support of  $\rho_i$  is a subset of the rule-complying actions of the active player at history  $h$* . Conditionally conformist preferences are represented by an expected utility function given by a linear combination of the player’s material payoff and a component representing some disutility arising from deviations from the presumed norm.<sup>12</sup> Formally, a norm-driven individual has conditionally conformist preferences characterized by a utility function  $u_i^C$  of the form

$$(2) \quad u_i^C(z, s_{-i}, \alpha_j) = m_i(z) - k_i d_i^C d_i^E \left( 1 + \sum_{j \neq i} \max \left\{ 0, E_{\rho_i, s_j, \alpha_j} [m_j | h^0] - m_j(z) \right\} \right),^{13}$$

with  $s_{-i} \in S_{-i}(z)$ ,  $k_i \in [0, \infty)$  and where:  $k_i$  is Player  $i$ 's *sensitivity to the presumed norm*;  $d_i^C$  is a dummy variable equal to one if  $i$  is aware of one or more behavioral rules applicable to the

---

<sup>11</sup> For instance, consider a rule that prescribes behavior minimizing payoff-inequality among players: when one evaluates this rule at the initial history, the rule will dictate those strategy profiles that minimize the difference in payoffs among players, considering that every terminal node can be reached; instead if one of the players deviates along the play, when evaluating this behavioral rule at a history following such a deviation, the rule will dictate strategy profiles that minimize the difference in payoffs among players, conditional on the terminal nodes that can still be reached (Sontuoso 2013).

<sup>12</sup> More precisely, the anticipated disutility is a function of any positive difference between the “initially expected payoff to  $j$ ” and the payoff  $j$  would get in the event of a rule violation. Note that it is assumed that if Player  $j$  is a norm-driven individual – and presumes that her co-players are norm-driven too – she can form her first-order belief  $\alpha_j$  by assuming her co-players’ behavior to be consistent with some norm-conjecture  $\rho_j = \rho_i$  (hence, with some rule  $r$ ).

<sup>13</sup>  $E[X]$  denotes the expected value of  $X$ .

given game, equal to zero otherwise;  $d_i^E$  is a dummy variable equal to one if  $i$  believes that every  $j \neq i$  is aware and will also adhere to some  $r \in R$ , equal to zero otherwise.<sup>14</sup>

In the next section we shall provide an application illustrating how a formal framework that can allow for different (conjectures about) norms is able to capture the focusing function of communication and to explain the experimental results of Bicchieri, Lev-On and Chavez (2010).

#### 4. An application accounting for the focusing function of communication

First, we shall define some specific behavioral rules reflecting principles which are usually assumed to regulate behavior in social dilemmas and which, one may assume, could apply to the trust game of Bicchieri, Lev-On and Chavez (see Figure 1 above). It is useful to recall here the above definition of *behavioral rule*, i.e. a correspondence dictating the strategy profiles most “appropriate” – according to a certain principle – for each node of the given mixed-motive game. Also, given a set of potential rules  $R$ , we assume that each player’s *culture identifies* a subset  $R_i$  (stored in  $i$ ’s memory) which contains default rules of behavior that the player is aware of (Sontuoso 2013).

Recall that the trust game of Bicchieri, Lev-On and Chavez (2010) was defined as follows: the investor receives \$6 and can choose to send any discrete amount  $x$  to the trustee; the amount the trustee receives is tripled by the experimenter, so that the trustee can then send any discrete amount  $y$  in the interval  $[0,3x]$  back to the investor. Given that, here are some rules applicable to the experimental trust game.

- “Inequality-Reducing” rule,  $r^F$ : any strategy profile such that the investor chooses an amount  $x$  (other than \$0), and the trustee chooses an amount  $y$  that minimizes the difference in payoffs.
- “Pareto-Efficiency” rule,  $r^P$ : any strategy profile such that the investor chooses \$6, and the trustee chooses any amount.
- “Reciprocity” rule,  $r^C$ : any strategy profile such that the investor chooses an amount  $x$  (other than \$0), and the trustee chooses an amount  $y \geq x$ .

---

<sup>14</sup> After deriving a belief-based formulation of Bicchieri’s conditions for a social norm to exist and be followed, Sontuoso (2013) proposed a notion of “*Social Sequential Equilibrium*” allowing for belief-dependent conformist motivations (by refining Battigalli and Dufwenberg’s specification of the sequential equilibrium concept of Kreps and Wilson 1982).

Before applying the above rules, note that in what follows we denote an *action* by the amount sent; for example, if the investor (*I*) chooses action \$6, then the set of the trustee's (*T*) feasible actions at history  $h = \$6$  is given by  $A_T(\$6) = \{\$18, \$17, \$16, \dots, \$2, \$1, \$0\}$ . It follows that if the investor chooses, say, \$1 and then the trustee chooses also \$1, the payoff profile induced by the reduced strategy profile  $(\$1, \$1 \cdot)$  is given by the pair  $(\$6, \$2)$ .

Now, considering for example  $r^F$ , one can derive the *set of strategy profiles dictated by*  $r^F$  which – when evaluated at the initial history – contains five elements, that is: (i) the strategy profile where the investor chooses \$6 and the trustee chooses  $\$9_{if \$6} \$7_{if \$5} \$5_{if \$4} \$3_{if \$3} \$1_{if \$2} \$0_{if \$1}$ , which yields the payoff profile  $(\$9, \$9)$ ; (ii) the strategy profile where the investor chooses \$5 and the trustee chooses the same as above, which yields the payoff profile  $(\$8, \$8)$ , etc... In short, using the above notation, the set of reduced strategy profiles dictated by  $r^F$  is given by:

$$r^F(h^0) = \{(\$6, \$9 \cdot), (\$5, \$7 \cdot), (\$4, \$5 \cdot), (\$3, \$3 \cdot), (\$2, \$1 \cdot)\}.$$

Similarly, considering  $r^P$ , one can derive the *set of (reduced) strategy profiles dictated by*  $r^P$  which – when evaluated at the initial history – contains nineteen elements, that is:

$$\{(\$6, \$18 \cdot), (\$6, \$17 \cdot), \dots, (\$6, \$1 \cdot), (\$6, \$0 \cdot)\}.$$

Also, the *set of (reduced) strategy profiles dictated by*  $r^C$  is given by:

$$\left\{ \begin{array}{l} (\$6, \$18 \cdot), (\$6, \$17 \cdot), \dots, (\$6, \$6 \cdot), (\$5, \$15 \cdot), (\$5, \$14 \cdot), \dots, (\$5, \$5 \cdot), \\ (\$4, \$12 \cdot), (\$4, \$11 \cdot), \dots, (\$4, \$4 \cdot), (\$3, \$9 \cdot), (\$3, \$8 \cdot), \dots, (\$3, \$3 \cdot), \\ (\$2, \$6 \cdot), (\$2, \$5 \cdot), \dots, (\$2, \$2 \cdot), (\$1, \$3 \cdot), (\$1, \$2 \cdot), (\$1, \$1 \cdot) \end{array} \right\}.$$

Next, if one assumes that both the investor (*I*) and the trustee (*T*) are aware of all the above behavioral rules (i.e.  $R_I = R_T = \{r^F, r^P, r^C\}$ ) then one can derive, for each player, the *set of rule-complying actions at history*  $h$ , which depicts the set of actions prescribed – by any of the rules  $r \in R_i$  – to Player  $i$  at history  $h$ . For the investor this is given by  $A_{i=I, h=h^0}(R_i(h^0)) = \{\$6, \$5, \dots, \$1\}$  while for the trustee there will be one set of rule-complying actions for each history, i.e.

$$A_{i=T, h=\$6}(R_i(h^0)) = \{\$18, \$17, \dots, \$6\}, A_{i=T, h=\$5}(R_i(h^0)) = \{\$15, \$14, \dots, \$5\}, \dots, A_{i=T, h=\$1}(R_i(h^0)) = \{\$3, \$2, \$1\}.$$

It is now clear that the aforementioned rules, when applied to the experimental trust game, dictate several strategy profiles. It then follows that the support of  $i$ 's *norm-conjecture*

$\rho_i$  (i.e. the set of the active player's rule-complying actions that are assigned positive probability by  $\rho_i$ ) may contain any of the above rule-complying actions. Especially in cases like this, where there are several admissible (i.e. rule-complying) actions – unless players can communicate – it might be difficult for them to engage in a process of mutually consistent belief formation relative to a presumed social norm: this may result in no social norm being followed. Instead, assume that a rule prescribing behavior that minimizes payoff-inequality among players is made salient *through communication*; in this case  $R_I$  and  $R_T$  are still defined as above (i.e.  $R_I = R_T = \{r^F, r^P, r^C\}$ ), but now it is reasonable to assume that players will converge towards a norm-conjecture derived from  $r^F$  only (and, in turn, they will derive their first- and second-order beliefs from such a norm-conjecture). In light of the experimental results discussed in Section 2, it seems reasonable to conclude that – in the experiment of Bicchieri, Lev-On and Chavez (2010) and, specifically, in the *relevant face-to-face* communication condition – the Inequality-Reducing rule  $r^F$  constituted a social norm and was being followed by the experimental subjects.<sup>15</sup> In fact, as shown in Figure 2, the relevant FtF communication game exhibited such a high level of trust that almost every investor contributed her entire \$6 endowment (while the average amount returned by the trustee was almost \$8, and the modal choice was \$9).

More explicitly, note that the norm-conjecture induced by  $r^F$ , for  $\forall i \in N$ , is such that:  $\rho_i$  may take on value 1 for any one of the investor's actions (other than \$0), and takes on value 0 for all of the trustee's actions but \$9<sub>if \$6</sub>, \$7<sub>if \$5</sub>, \$5<sub>if \$4</sub>, \$3<sub>if \$3</sub>, \$1<sub>if \$2</sub>, \$0<sub>if \$1</sub>. Given that, each player  $i$  can form her first-order belief  $\alpha_i$  by assuming her co-player's behavior to be consistent with her norm-conjecture  $\rho_i$ ; for example, the investor's belief  $\alpha_I = (\cdot | h^0)$  will correspond to a probability measure over the strategies of the trustee, with the support of  $\alpha_I$  containing only the opponent's rule-complying strategies. Using formula (2) above, the investor can then calculate the expected utility from each of her actions as follows: first, note that the investor's utility would involve a potential loss (i.e. a “psychological” disutility)

---

<sup>15</sup> A social norm  $r^*$  (exists and) *is followed* by population  $N$  if: every player  $i \in N$  has conformist preferences represented by a utility function  $u_i^C$  (as given in formula (2) above), with  $d_i^C = 1$ ,  $d_i^E = 1$ , and  $k_i > 0$ ; every player  $i$  maximizes her expectation of  $u_i^C$ ; every  $i$  holds correct beliefs about every  $j$ 's ( $j \in N$ , with  $j \neq i$ ) first-order belief and behavior; every player  $i$ 's behavior is consistent with one of the end-nodes yielded by  $r^* \in R_i \cap R_j$  (according to norm-conjectures  $\rho_j = \rho_i$  for  $\forall j \in N$ );  $k_i$  is sufficiently large for every  $i \in N$  (Sontuoso 2013). (See also Bicchieri (2006: 11) for the conceptual distinction between the existence of a norm *versus* its being followed.)

only at  $x = \$0$  while it would be maximized at  $x = \$6$  (conditional on the trustee's sensitivity  $k_T$ ). Yet before considering the latter case, let's look at the game from the trustee's perspective: in order to calculate the optimal action at each history after which she has to move, the trustee will compare her utility from conforming against her utility from deviating from the presumed norm; so, the *trustee's expected utility from deviating* (choosing, say,  $\$0$ ) after the investor has chosen  $x = \$6$  would equal  $u_T^C(z, \rho_i, \beta_T) = 18 - k_T [1 + 9]$ ;<sup>16</sup> instead, the trustee's utility from choosing  $\$9$  after  $x = \$6$  would simply correspond to her material payoff (i.e.  $m_T(z(\$6, \$9)) = \$9$ ). In brief, the trustee's conformist preferences can be expressed compactly as:  $18 - 10k_T \leq 9 \Rightarrow k_T \geq \frac{9}{10}$ . If the investor effectively believes that  $k_T \geq \frac{9}{10}$ , then she will compare her utility from taking a rule-complying strategy against her utility from deviating from the presumed norm: the investor's expected utility from deviating (i.e. choosing  $x = \$0$ ) would equal  $u_I^C(z, \rho_i, \beta_I) = 6 - k_I [1 + \bar{y}]$ , where the "initially expected payoff to the trustee" given the reduced strategy  $s_T = y \cdot$  (i.e.  $E_{\rho_i, s_T, \alpha_T}[m_T|h^0]$ ) is now denoted by  $\bar{y}$  for convenience. Instead, the investor's expected utility from choosing  $x = \$6$  would be given by  $m_I(z(\$6, \$9)) = \$9$ . Hence, the investor's conformist preferences can be expressed compactly as:  $6 - k_I [1 + \bar{y}] \leq 9 \Rightarrow k_I \geq -\frac{3}{1+\bar{y}}$ , which is always satisfied. To conclude, the investor will choose  $\$6$  and the trustee will choose  $\$9$  if  $\$6$   $\$7$  if  $\$5$   $\$5$  if  $\$4$   $\$3$  if  $\$3$   $\$1$  if  $\$2$   $\$0$  if  $\$1$  (whenever  $k_T \geq \frac{9}{10}$ ).

To sum up, if a rule prescribing behavior that minimizes payoff-inequality among players is made salient through communication, the equilibrium path will involve the investor choosing  $\$6$  and the trustee choosing  $\$9$ , so that both players' payoff is  $\$9$ : again, looking at Table 1 above, one may conclude that – in the relevant face-to-face communication condition – subjects did play an equilibrium where norm-conjectures were correct. In other words, communication coordinated players' expectations by making a particular rule salient. Besides, note that moving from *relevant face-to-face* communication to *relevant computer-mediated* communication somewhat reduced both reciprocity and expected reciprocity. Indeed, when relevant communication is allowed, the expectation that trustees will abide by the social norm activated through communication is likely to be *less vivid in a*

---

<sup>16</sup> Note that, for some player  $i$ ,  $u_i^C(z, s_{-i}, \beta_i)$  represents  $i$ 's estimation of  $u_i^C(z, s_{-i}, \alpha_j)$ .

*computer-mediated* than in a face-to-face environment (and coordinating expectations will be more difficult).<sup>17</sup>

## 5. Conclusion

We investigated the focusing (and coordinating) function of communication: drawing on experimental evidence we have argued that – when a behavioral rule becomes situationally salient – it causes a shift in an individual’s focus, thereby generating (empirical and normative) expectations that direct one’s strategies. We presented an application illustrating how a formal framework that allows for different conjectures about norms is able to capture such a focusing function.

Given that this framework allows to compare predictions under *different focal rules of behavior*, it could be of help in designing novel experiments that study the dynamics of the cognitive processes characterizing sequential trust games and, more generally, social dilemmas that involve a stage of non-binding, pre-play communication among participants.

## References

- Battigalli, Pierpaolo and Martin Dufwenberg. 2009. “Dynamic Psychological Games” *Journal of Economic Theory*, 144(1): 1-35.
- Bicchieri, Cristina. 2002. “Covenants without Swords: Group Identity, Norms, and Communication in Social Dilemmas” *Rationality and Society*, 14(2): 192-228.
- . 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.
- Bicchieri, Cristina and Azi Lev-On. 2007. “Computer-Mediated Communication and Cooperation in Social Dilemmas: An Experimental Analysis” *Politics Philosophy Economics*, 6(2): 139-168.
- Bicchieri, Cristina, Azi Lev-On and Alex Chavez. 2010. “The Medium or the Message? Communication Relevance and Richness in Trust Games” *Synthese*, 176(1): 125-147.
- Cialdini, Robert B., Raymond R. Reno and Carl A. Kallgren. 1991. “A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in

---

<sup>17</sup> In this respect see also the discussion in Bicchieri and Lev-On (2007).



Human Behavior" in *Advances in Experimental Social Psychology*, Vol. 24, ed. Mark P. Zanna. San Diego: Academic Press.

Davis, Deborah and William T. Perkowski. 1979. "Consequences of Responsiveness in Dyadic Interaction: Effects of Probability of Response and Proportion of Content-Related Responses on Interpersonal Attraction" *Journal of Personality and Social Psychology*, 37(4): 534-550.

Desforges, Donna M., Charles G. Lord, Shawna L. Ramsey, Julie A. Mason, Marilyn D. Van Leeuwen, Sylvia C. West and Mark R. Lepper. 1991. "Effects of Structured Cooperative Contact on Changing Negative Attitudes toward Stigmatized Social Groups" *Journal of Personality and Social Psychology*, 60(4): 531-544.

Geanakoplos, John, David Pearce and Ennio Stacchetti. 1989. "Psychological Games and Sequential Rationality" *Games and Economic Behavior*, 1(1): 60-79.

Kreps, David M. and Robert Wilson. 1982. "Sequential Equilibria" *Econometrica*, 50(4): 863-894.

Lev-On, Azi, Alex Chavez and Cristina Bicchieri. 2010. "Group and Dyadic Communication in Trust Games" *Rationality and Society*, 22(1): 37-54.

Loomis, James L. 1959. "Communication, the Development of Trust, and Cooperative Behavior" *Human Relations*, 12(4): 305-315.

Osborne, Martin J. and Ariel Rubinstein. 1994. *A Course in Game Theory*. Cambridge, Mass.: MIT Press.

Sally, David. 1995. "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992" *Rationality and Society*, 7(1): 58-92.

Sontuoso, Alessandro. 2013. "A Dynamic Model of Belief-Dependent Conformity to Social Norms" *MPRA Paper 53234*, University Library of Munich, Germany.