I.S.S.N: 1885-6888

ECONOMIC ANALYSIS

WORKING PAPER SERIES

The Power of Words: Why Communication Fosters Cooperation and Efficiency

♦

Raúl López-Pérez

Working Paper 1/2009

UAM
UNIVERSIDAD AUTONOMA
DE MADRID

DEPARTAMENTO DE ANÁLISIS ECONÓMICO:
TEORÍA ECONÓMICA E HISTORIA ECONÓMICA

The Power of Words: Why Communication Fosters Cooperation and Efficiency[*]

Raúl López-Pérez[†]

November 2008

Abstract: We present a game-theoretical model that accounts for abundant experimental evidence from games with non-binding communication ('cheap talk'). It is based on two key ideas: People are *conditionally* averse to break norms of honesty and fairness (i.e., the emotional cost of breaking a norm is low if few people comply), and heterogeneous with regard to their concern for norms. The model explains (a) why cooperation in social dilemmas rises if players can previously announce their intended play, (b) why details of the communication protocol like the number of message senders and the order in which players communicate affect cooperation, (c) why players in sender-receiver games tend to transmit more information than a standard analysis would predict, and (d) why senders of false messages are often sanctioned if punishment is available.

Keywords: Communication, Cooperation, Fairness, Heterogeneity, Honesty, Reciprocity, Social Norms.
JEL classification numbers: C72, D01, D62, D64, Z13.

[*] I am indebted to Urs Fischbacher, Michael Kosfeld, Andreas Leibbrandt, Topi Miettinen, Michael Näf, Christian Zehnder, and participants at several conferences and seminars for helpful comments. Part of this research was conducted while visiting the Institute for Empirical Research in Economics at Zurich, and I would like to thank their members for their great hospitality. I also gratefully acknowledge financial support from the European Union through the ENABLE Marie Curie Research Training Network.

[†] Department of Economic Analysis, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain. Tel.: +34 91 497 6801; Fax: +34 91 497 6930; E-mail: raul.lopez@uam.es

1. Introduction

People communicate with others in many of their daily interactions, and much of that communication takes place by means of costless, non-binding messages (*cheap talk*). Apparently, one reason why people exchange messages is because they can often improve their payoff when they communicate. In fact, there is abundant experimental evidence in line with this, even if the communicators are anonymous subjects playing one-shot games. Thus, Sally (1995) offer a meta-analysis of social dilemma experiments conducted from 1958 to 1992 and report that non-binding promises raise cooperation by 30 percent, thus improving the average player's payoff with respect to the case with no communication.[1]

Social researchers have advanced different hypotheses in order to explain why non-binding communication fosters efficiency, like (1) communication enhances group identity (Orbell et al., 1990), (2) communication acts as a coordinating device (Farrell and Rabin, 1996), (3) communication raises payoff expectations on receivers, and senders feel badly if they let down those expectations (guilt aversion, as in Charness and Dufwenberg, 2006), and (4) communication elicits social norms (Bicchieri, 2002). In this paper, we suggest that social norms are the key factor, and provide a game-theoretical model to account for the effect of communication. Our model posits that people care about social norms (Elster, 1989; Becker, 1996; López-Pérez, 2008) in a *conditional* manner –intuitively, they feel painful emotions like shame when they transgress internalized norms that others respect. Further, we posit that players are heterogeneous,[2] distinguishing between three types of people: (i) Selfish players who do not care at all about norms, (ii) H-players who find binding a norm of honesty, and (iii) EH-players who find binding a norm of fairness and honesty –more precisely, this norm commends to achieve and efficient and egalitarian (E) outcome and to be honest, which explains the acronym EH.

A key message from the paper is that the interaction between these three types of agents is crucial to understand the effect of communication. For instance, the data shows –see Sally (1995)- that (a) some people cooperate (under certain conditions) even if they *cannot* communicate, (b) communication increases cooperation (which suggests that some people cooperate *only* if there is prior communication), and (c) some people never cooperate in social dilemmas, even if pre-play talk is available. These phenomena are

[1] Ledyard (1995, pp. 156-8) and Bicchieri (2002) survey the evidence on communication in public good games and social dilemmas, while Ellingsen and Johannesson (2004) review some related psychological literature. In addition, Crawford (1998) survey the experimental evidence on how communication affects coordination on efficient outcomes.

[2] The idea that agents are heterogeneous in their pro-sociality is consistent with a large body of experimental data. To start, the evidence from social dilemmas without pre-play talk (Croson, 2000; Brandts and Schram, 2001, Fischbacher et al., 2001) points out that some subjects are conditional cooperators who cooperate if they expect others to cooperate as well, while remaining subjects rarely cooperate. In addition, recent lab evidence shows as well that subjects differ in their propensity to tell the truth (Gneezy, 2005; Sanchez-Pagés and Vorsatz, 2007).

respectively explained by the presence of EH-types, H-types, and selfish types. In effect, EH-players cooperate if they expect their co-player to cooperate as well –that is, they are conditional cooperators- and their presence explains why some people cooperate even if they cannot talk. In turn, communication increases cooperation because it allows H-types to make promises and hence commit themselves to cooperate. Is that ever an optimal strategy? Yes, if they believe that the promise *receiver* is an EH-player, that is, the type of person who cooperates conditionally and hence would defect if the sender announced defection. Finally, selfish agents are necessary to explain why cooperation sometimes fails to happen, *even* if communication is available.

In line with the available experimental evidence, the model also points out that the effectiveness of pre-play talk subtly depends on a number of variables, like the content of the messages sent and received, the order of play of the message sender, the number of message senders and receivers, and the expected price of being sincere. In addition, simple extensions of the basic model can explain why players in sender-receiver games tend to transmit more information than a standard analysis would predict and why people are often willing to spend resources to punish cheaters.

Recent theories of other-regarding preferences, which relax the standard assumption that *all* agents are selfish, are closely related to our model.[3] Rabin (1993) model reciprocity in normal-form, two-player games as the idea that people are kind to those who are kind to them, and harm those who harm them. Dufwenberg and Kirchsteiger (2004) extend Rabin's ideas to extensive form games. Levine (1998) assume type-based altruism and spitefulness, and both Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) propose models of inequity-averse players. Finally, Charness and Rabin (2002), Falk and Fischbacher (2006), and Cox et al. (2007) introduce both reciprocity and distributive concerns.

In contrast to our theory, these models cannot account for the effect of communication on efficiency. Models of inequity aversion, for instance, assume that players' utility only depends on the distribution of material payoffs, that is, something that cheap talk cannot shape by definition. As a result, communication has no effect on best-responses and equilibria of the action stage subgame. In order to say something more determinate, however, one might assume that pre-play talk affects players' expectations, thus making some equilibria focal (Schelling, 1960). For instance, Farrell (1987) and Farrell and Rabin (1996) assume that communication acts as a coordinating device.[4] Although this focal point hypothesis shares some predictions with our model, it cannot replicate other predictions that are very consistent with the available evidence. In this regard, our impression is that some experimental phenomena (like the sanctioning of

[3] Camerer (2003) and Fehr and Schmidt (2006) extensively survey this literature. See also López-Pérez (2008) for a detailed comparison between our model and some of these models.

[4] Consult also Aumann (1990) and Rabin (1994). See also Crawford (1998) or Aumann and Hart (2003) for extensive surveys of the literature on cheap talk.

cheaters in one-shot games with a punishment stage, or the effectiveness of pre-play talk in games with a unique equilibrium) are very difficult to clarify unless one assumes that communication shapes utility (and not only expectations).

In this respect, we note that some recent papers analyze how communication shapes utility. Thus, Ellingsen and Johannesson (2004) combine inequity aversion and a *fixed* cost of lying in a hold-up game, and provide experimental evidence in line with this; Miettinen (2005) study a two-player game which is preceded by negotiations and assume that players feel badly if they deviate from the agreement; Demichelis and Weibull (2008) posit that players have a lexicographic preference for honesty (second to the material payoffs in the stage game) and analyze evolutionary stability in coordination games, and Kartik (2008) study sender-receiver games when the sender bears a cost of lying.

The rest of the paper is organized as follows. Next section describes the basic model[5] and section 3 studies how pre-play communication affects cooperation in social dilemmas. Section 4 proposes a number of simple extensions of the basic model to explain over-communication in sender-receiver games and the sanctioning of cheaters. Section 5 concludes by mentioning some possible applications.

2. The model

Consider any n-player, extensive form game of perfect recall. Let N = {1,..., n} denote the set of players, z a terminal node, h an information set, and $M(h)$ the set of available moves at h. Further, let $u_i(z)$ denote player i's utility payoff at z, and $x_i(z)$ denote player i's monetary (material) payoff at z.

A player may be given the opportunity to communicate at some information set of the game. We say that player A communicates with player B if she sends a message to B, and assume for the moment that messages can be used only to announce a player's future actions (i.e., her intentions). Hence, a message is a statement of the type 'I will play action **a** at information set h'. Note well that (the submission of) a message is a formal move in the game tree (for expositional reasons, we term *actions* any other formal moves different than messages). To keep matters simple, and unless noted otherwise, we posit that sending a message is costless, that a communicator can always keep silent if she wishes so (as a matter of convenience, we consider silence a message), that players share a common language, and that communication occurs without noise –i.e., if B receives a message from A, it is common knowledge that B will interpret it as player A does.

[5] This model is an extended version of the model in López-Pérez (2008) in that we allow for the possibility that some players care about honesty. This point was immaterial in López-Pérez (2008) because our focus there was on games where players *cannot* communicate. We showed there that abundant lab evidence from such class of games is consistent with our model, and discussed in detail the psychological intuition behind our hypotheses. The interested reader is therefore directed there for a lengthy discussion of these points.

The key hypothesis of the model is that there exist *three* types of players: *EH, H*, and *Selfish* –the reason for this terminology will be clear later. Unless otherwise noted, we posit that players' types are private information and use Perfect Bayesian Equilibrium (PBE) as a solution concept. We let ρ and μ respectively denote the probability of being an H and an EH-type -of course, the existence of selfish agents requires $\rho + \mu < 1$.

Selfish players are risk neutral money-maximizers with utility function $u_i(z) = x_i(z)$. In contrast, the utility of the other two types depends on the money earned $x_i(z)$ *but also* on norms –i.e., rules indicating how one ought to move-, the intuition being that they feel ashamed or guilty if they violate a binding norm. To formalize this, we first formally define norms:

Definition 1: A norm is a nonempty correspondence ψ: $h \rightarrow M(h)$ applying on any information set, except on Nature's ones.

This concept allows us to introduce some new terms. First, we say that a player *respects* or *complies* with any norm ψ at h if (i) her move at h is consistent with that norm or if (ii) she does not move at h. Otherwise, she *deviates* from ψ. Further, we denote by $\mathrm{R}(\psi, \mathrm{z})$ the set of players who respected ψ in the history of z, and by $\mathrm{r}(\psi, \mathrm{z}) \in [0, 1]$ the overall proportion of players who respected ψ in the history of z. Note well that these two concepts are exogenous ones (as ψ), not equilibrium ones.

EH and H-types are different because they care about different norms: The *EH-norm* and the *H-norm*, respectively, which we precisely define below. However, the utility function of both types has the same structure (in the following expression, ψ denotes the EH-norm if the player is an EH-type and the H-norm if the player is an H-type):

$$u_i(z) = \begin{cases} x_i(\mathrm{z}) - \gamma \cdot \mathrm{r}(\psi, \mathrm{z}) & \text{if } i \notin \mathrm{R}(\psi, \mathrm{z}),\ (0 < \gamma) \\ x_i(z) & \text{if } i \in \mathrm{R}(\psi, \mathrm{z}) \end{cases}$$

In other words, EH and H-types suffer a psychological cost when they deviate from the norm that they find binding –for this reason, we will refer sometimes to both of them as *principled* types. Crucially, the strength of this cost positively depends on the proportion of people who respect the norm.[6] We make five remarks on this hypothesis. First, its intuition has to do with the psychology of shame, an emotion that is strongly correlated with inferiority feelings, as we have argued in López-Pérez (2005). Second, it introduces reciprocal behavior: EH and H-types are more likely to respect their own norms if they expect sufficiently many players to comply as well - abundant lab evidence supports the idea that (some) humans behave in a reciprocal manner; consult Fehr and Gächter (2000).

[6] In López-Pérez (2008) we assumed that this cost depends on the number (and not the proportion) of players who respect the norm. Both specifications render qualitatively similar results in the games analyzed there, but we now believe that our new modeling choice is empirically more valid (especially in multiple-player games). For this reason, we opted for it here.

Third, although we assume for simplicity that the cost increases linearly with $r(\psi, z)$, our results do not depend on this (what is essential here is that the cost is strictly increasing). Fourth, parameter γ can be interpreted as a player's internalization index. Finally, note that the cost is null if nobody complies with the norm: To put it like this, there is no preference for norm compliance per se. Although adding an unconditional cost in our model would be direct, we have chosen not to do that because the available evidence (Fehr and Gächter, 2000) suggests that most cooperative behavior occurs conditionally, hence rejecting the idea of a (significant) fixed cost.

We must be precise about the EH and H-norms in order to obtain determinate behavioral predictions and test the model. Basically, the H-norm is a norm of honesty (this explains its name), while the EH-norm is a norm of distributive justice (which implies cooperation) *and* honesty.

Definition 2 (The H-norm): At any h where the mover can communicate, this correspondence selects *any* available message. At any other h, it selects action $\mathbf{a} \in M(h)$ if the mover announced **a** previously with a message, and the whole set $M(h)$ otherwise.

In other words, this norm affirms that anyone who sends a message about her *future* intentions ought to honor her word later and act as announced. Since the H-norm restricts behavior only when a player can communicate, it follows that the utility function of an H-type *who cannot communicate* collapses to that of a selfish player. We pass now to describe the EH-norm, for which we need to introduce an additional concept.

Definition 3: Let t_0 denote any initial decision node of the game –i.e., any node immediately following Nature's moves (if any)- and $X(t_0)$ denote the set of all allocations of *monetary* payoffs that succeed t_0 (we assume it to be a compact set). Allocation $x = (x_1, ..., x_n) \in X(t_0)$ is an (Efficient and Egalitarian) *E-allocation* of t_0 if it maximizes function ($0 < \delta < 1$)

$$F^E(x) = \sum_{i \in N} x_i - \delta(\max_{i \in N} \{x_i\} - \min_{i \in N} \{x_i\}) \qquad (1)$$

over $X(t_0)$. A path connecting node t_0 and one of its E-allocations is an *E-path* of the game. An *E-action* is an action that belongs to at least one E-path.

Assumption $0 < \delta < 1$ implies that any E-allocation is necessarily Pareto efficient (this can be easily proved by contradiction). Hence, one can see an E-allocation as a Pareto efficient allocation that, in comparison with other available monetary allocations, is not too unequal and socially inefficient. Note that we use for simplicity a very rudimentary measure of inequity (the largest distance between players' incomes), but more sophisticated measures could be easily introduced.

Definition 4 (The EH-norm): At any h where a player can communicate, this norm selects silence and any message announcing an E-action. At any other h, the norm selects (i) action **a** $\in M(h)$ if the mover announced **a** previously, and (ii) any E-action of h otherwise -if there is no E-action, the norm selects the whole set $M(h)$.

There are a number of ideas buried in this definition. To start, this norm commends to achieve an E-allocation *and* to be honest. For this reason, the EH-norm and the H-norm are rather different: The EH-norm is not only a norm of honesty but also a norm of distributive justice (because it commends to achieve a fair allocation). Additionally, the EH-norm asks for moral coherence in that it only allows announcements of E-actions, a point that seems natural: If E-actions are the 'right' actions and moreover announcements are morally binding, it does not make sense to announce something different.

We finish with two remarks. *First*, the reader may wonder why we consider the EH and the H-norm and not other norms. Observe that the complexity of the model increases with the number of norms. This motivated us to limit that number to the minimum possible - in any case, the model is flexible enough to easily include additional heterogeneity. Furthermore, both norms are relatively simple and they embody assumptions that happen to be consistent with much experimental evidence. For instance, the idea present in the EH-norm that both efficiency *and* payoff equality are basic ingredients of fairness can explain a very good deal of the experimental evidence from games without communication, as the results in López-Pérez (2008) attest. *Second*, the model assumes that both EH and H types care about honesty, whereas only the EH-types care about fairness (or social preferences). As we show throughout the paper, this hypothesis is in line with the data from numerous experimental studies, including a within-subject study by Hurkens and Kartik (forthcoming) (see section 4.1). However, why are principled players heterogeneous? Intuitively, many of us feel badly if we deviate from a binding norm and we are unable to morally justify that behavior (Elster, 1999). In this regard, we conjecture that finding excuses for lying might be relatively more difficult than finding excuses for not cooperating,[7] and that might account for heterogeneity. In any case, additional within-subjects studies are required to further clarify this point.

3. Communication in Social Dilemmas

This section studies how communication affects cooperation in social dilemmas. To organize the exposition, we present a number of predictions of the model. For each prediction, we provide a simple example to illustrate it and (when available) some supporting experimental evidence.[8] To clarify the *net* effect of communication on

[7] This might be especially true in lab experiments, which are often anonymous settings: For instance, since it is uncertain whether the other participants are richer or less needy, some subjects might feel entitled not to cooperate and hence behave as a selfish person would do. In contrast, this same people might have no excuse to cheat others.

[8] To organize the exposition, we finish the discussion on each prediction with a black square (▪).

cooperation, we first analyze some determinants of cooperation when communication is *not* available.

Prediction 1: Some players may cooperate in equilibrium even if they cannot communicate. The actual rate of cooperation depends on the constellation of monetary payoffs and the proportion of principled types.

Example: Table 1 depicts *monetary* payoffs in the Prisoner's Dilemma (PD) lab game. In this game, each player chooses between cooperation (action C) and defection (action D). Both earn c monetary units if they cooperate and d if they defect. Further, a unilateral defector gets a 'temptation' payment of t while a unilateral cooperator gets a normalized payoff of zero. Payoffs satisfy $t > c > d > 0$ so that defection strictly dominates cooperation in monetary terms, and $2c > t$ so that mutual cooperation is socially efficient.

Assume that the PD players make their choices simultaneously and that they cannot communicate with each other. To get utility payoffs, we note two things: (i) As players cannot communicate, the utility of an H-type coincides with that of a selfish type, (ii) condition $2c > t$ implies that (c, c) is the only E-allocation of this game and cooperation the only E-action. With this in mind, table 2 illustrates players' utility payoffs if Row is selfish (or an H-type) and Column is an EH-type (other cases are direct from this). Trivially, Row's payoffs coincide with her own pecuniary payoffs. On the other hand, Column gets some disutility if he deviates unilaterally from the EH-norm, while he feels no disutility if both players defect.

	C	D
C	c, c	0, t
D	t, 0	d, d

Table 1: Monetary payoffs in the PD

	C	D
C	c, c	0, t - γ /2
D	t, 0	d, d

Table 2: Utility Payoffs if Row (Column) is selfish (EH)

Inspection of table 2 indicates that a selfish player (or an H-player) never cooperates in equilibrium. In contrast, there exists a perfect Bayesian equilibrium in which the EH-players cooperate if the following inequality holds (this indicates that cooperation is optimal for them):

$$\mu \cdot c \geq \mu(t - \gamma/2) + (1 - \mu) \cdot d \ , \tag{2}$$

that is, if μ is larger than

$$\mu^{sim} = \frac{d}{d - t + c + \gamma/2} . \tag{3}$$

The intuition behind this equilibrium is straightforward. Selfish and H-players defect to maximize their monetary payoff. In contrast, EH-types cooperate if the probability μ that the co-player is also an EH type is large enough, as defection is likely to be unilateral in this case and hence entail a psychological cost. Of course, this cost should be large enough in order to sustain cooperation -observe in this regard that $1 \geq \mu^{sim}$ requires $\gamma \geq 2 \cdot (t-c)$ so that no cooperation is expected if $\gamma < 2 \cdot (t-c)$. Finally, we note that there exists an additional equilibrium in which all types of players defect (this equilibrium exists for any value of μ). In this sense, cooperation also requires that the EH-types find the former, cooperative equilibrium more intuitive than this latter equilibrium. We implicitly assume this in what follows.

Experimental evidence for prediction 1: The model predicts that some players (the EH-types) cooperate, which is consistent with abundant evidence –consult Sally (1995) for a meta-analysis. The model also predicts that cooperation requires at least a mass μ^{sim} of EH-types. This means that EH-types are conditional or reciprocal cooperators: They cooperate only if the co-player is likely to cooperate as well, a prediction well supported by the data from numerous experiments –see again Sally (1995) or Croson (2000).

Finally, note that μ^{sim} decreases with c and increases with t and d. Since cooperation is hindered as μ^{sim} increases, the model consequently forecasts that cooperation depends directly on c, and indirectly on t and d, which is again consistent with the lab evidence –see Rapoport and Chammah (1965, pp. 36-39), and Clark et al. (2001). A possible interpretation of this result is that cooperation respects *the law of demand*: Cooperation decreases when its price increases –to understand this, observe that the expected price of cooperation $\mu \cdot (t-c) + (1-\mu) \cdot d$ depends negatively on c and positively on t and d.▪

While prediction 1 indicates that the constellation of material payoffs should affect the cooperation rate, we note that other factors may play a role as well, even if communication is not available. In particular, a key factor is the information that each player has about the other players' moves. In effect, since the EH-players cooperate conditionally, cooperation might be enhanced if players can observe whether others cooperated. As a result, the order in which players move –i.e., simultaneously or sequentially- matters.

Prediction 2: The rate of cooperation depends on the order of play.

Example: To illustrate this, consider the sequential PD game –i.e., one of the players chooses after observing her co-player's move. We will show that the equilibrium cooperation rate is higher in the sequential PD game than in the simultaneous one.

Since allocation (c, c) is the only E-allocation, it follows that both players cooperate in the unique E-path of this game (see definition 3). Consequently, the EH-norm commends the first mover to cooperate, while the second mover should cooperate if the first mover cooperated. In contrast, the EH-norm allows the second mover to choose any action if the first mover defected and hence deviated from the E-path (see definition 4). Figure 1 depicts players' payoffs if both players are EH (upper payoffs correspond to the first mover; further, E-actions are identified by an arrow).

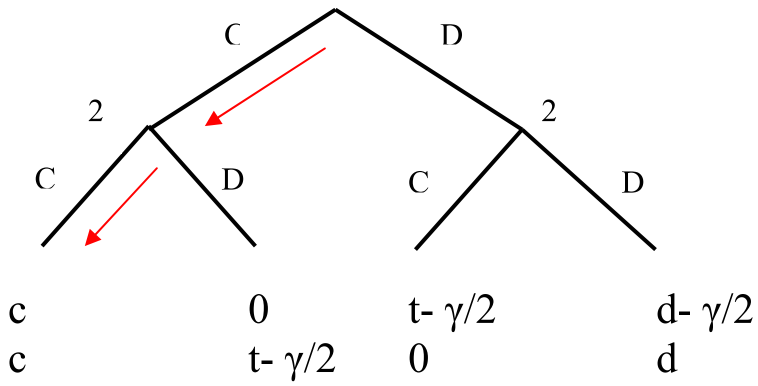


Figure 1: EH-players' payoffs in the sequential PD

Whatever the players' types, this game has a unique Perfect Bayesian Equilibrium for any parameter calibration.[9] In this equilibrium, the second mover always defects if she is a selfish or an H type. Further, and as one may confirm from Figure 1, an EH second mover defects at any information set if $\gamma < 2(t-c)$ and reciprocates the first mover's choice if $\gamma \geq 2(t-c)$ -that is, she cooperates if the first mover cooperated and defects if he defected. As a result, cooperation is profitable for the first mover if $\gamma \geq 2(t-c)$ and μ is large enough. More precisely, selfish and H first movers cooperate if $\mu \cdot c > d$, while EH first movers do the same if $\mu \cdot c > d - {}^{\gamma}\!/_{2}$, that is, if μ is larger or equal than

$$\mu^{seq} = \frac{2d-\gamma}{2c}. \tag{4}$$

Note that selfish and H first movers comply with the EH-norm in order to 'emotionally force' an EH-second mover to comply as well –see Rabin (1993, p. 1296) for a similar suggestion. As any first mover cooperates if μ is large enough, it follows that the average rate of cooperation in the sequential PD is larger than the maximum rate in the simultaneous PD. Further, comparison between expressions (3) and (4) indicates that $\mu^{sim} > \mu^{seq}$ if $\gamma \geq 2(t-c)$. This means that an EH-mover in the simultaneous PD requires a larger prior to cooperate than an EH first mover in the sequential PD, which is again in line with our suggestion that the sequential mechanism is relatively more effective in rising cooperation. Note in addition that the sequential game has a unique equilibrium, while the simultaneous one has multiple equilibria for some parameter constellations, which constitutes an additional handicap for cooperation to succeed.

Experimental evidence for prediction 2: Experimental evidence from Hayashi et al. (1999) and Clark et al. (2001) corroborates our equilibrium predictions: Second movers sometimes cooperate, but they do it conditional on the first mover's choice, while unconditional cooperation is negligible. In addition, Clark et al. (2001) report that

[9] More precisely, there exist multiple equilibria which are essentially equivalent as they only differ in the beliefs off the equilibrium path. These beliefs play no role in the games that we analyze in this paper, and for this reason we do not report them.

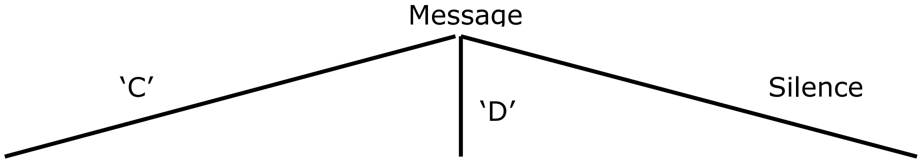
reciprocation falls as its material cost rises, something that is also consistent with our model, as reciprocation is predicted only if $\gamma \geq 2(t - c)$. Finally, Hayashi et al. (1999) and Clark et al. (2001) report that the sequential game elicits a higher rate of cooperation than the simultaneous one.▪

While predictions 1 and 2 are useful as a benchmark for later comparisons, our main focus is on how communication increases cooperation. The next prediction first considers this issue.

Prediction 3: Pre-play communication can foster cooperation. The efficacy of communication depends on the content of the message.

Example: We return to the simultaneous PD, assuming now that unilateral (*one-way*) pre-play communication is available. That is, prior to playing the PD, one of the players (the *sender*) can either send a non-binding message announcing her future move or stay silent. We call this the *communication* stage, to distinguish it from the *action* stage, where the proper PD is played. The combination of both stages forms the *entire game*.

We first elucidate what actions and messages are selected by the EH-norm and the H-norm at each information set. To start, the EH-norm selects silence or message 'C' in the communication stage. In the action stage, in turn, the EH-norm distinguishes between players: While the sender should move C if she announced 'C' or kept silent, and D if she announced 'D'; the other player (the *receiver*) should always move C. Finally, the H-norm selects any message but commends to play accordingly later, and selects any action if a player kept silent (note that this applies also to the receiver).



	C	D
C	c	0
D	$t - \gamma/2$	d

	C	D
C	$c - \gamma/2$	0
D	$t - \gamma/2$	D

	C	D
C	c	0
D	$t - \gamma/2$	d

	C	D
C	C	0
D	$t - \gamma/2$	$d - \gamma/2$

	C	D
C	$c - \gamma/2$	$- \gamma/2$
D	t	D

	C	D
C	c	0
D	t	d

Figure 2: Sender's utility payoffs for any possible strategy profile (the three upper matrices depict an EH-sender's payoffs, while the lower three ones depict an H-sender's ones)

Taking all this into account, figure 2 depicts utility payoffs for an EH and an H-sender and for any possible strategy profile. The upper tree indicates available messages, while the posterior matrices indicate the sender's payoffs for each possible message and combination of players' choices in the action stage (in the payoff matrices, the sender is assumed without loss of generality to be the row player; further, the three upper matrices correspond to an EH-sender; the three lower ones to an H-sender).

If $\gamma \geq \max\{2 \cdot (t-c), d\}$ and μ is large, the game has a PBE in which (i) any type of sender announces 'C', EH and H senders cooperate afterwards, and selfish ones defect, (ii) any type of sender defects if she made an announcement different than 'C', (iii) an EH receiver cooperates if his co-player announced 'C' and defects otherwise, and (iv) a selfish or H-receiver defects whatever the message received. Since H-senders find optimal to cooperate in this PBE, average cooperation here is larger than in the cooperative equilibrium of the simultaneous PD game with no communication (in this equilibrium, only the EH-types cooperate). Therefore, communication can raise cooperation in this game.

We now prove that the above mentioned conditions (i) to (iv) indeed characterize an equilibrium strategy profile. First, it is obvious that a selfish sender or receiver acts optimally (note that selfish senders mimic the other types' announcement to prevent signaling their type), and the same is true for an H-receiver. In addition, figure 2 indicates that EH and H-senders should also defect if they sent a message different than 'C' (among other reasons because the receiver will not cooperate then). As a result, and independently of the beliefs off the equilibrium path, defection is also optimal for an EH-receiver if the sender previously announced something different than 'C' (recall that EH-types cooperate reciprocally).

On the contrary, an EH-sender who previously announced 'C' should cooperate if $\mu \cdot c + (1-\mu) \cdot 0 \geq \mu(t-\gamma/2) + (1-\mu) \cdot d$, that is, if her prior μ is larger than (observe that this is identical to expression (3), the threshold with no communication):

$$\mu^{sim} = \frac{d}{d-t+c+\gamma/2} .$$

This line of reasoning also applies to an EH receiver who received message 'C', with the only caveat that now it is ρ + μ (and not only μ) what should be larger than threshold μ^{sim}, since both EH and H senders are expected to cooperate. In turn, inspection of figure 2 indicates that an H sender who announces 'C' would rather honor her word if

$$\mu \cdot c + (1-\mu) \cdot 0 \geq \mu(t-\gamma/2) + (1-\mu)(d-\gamma/2) \Leftrightarrow \gamma/2 \geq (1-\mu) \cdot d + \mu \cdot (t-c),$$

which holds true if $\gamma \geq \max\{2 \cdot (t-c), d\}$. To finish with the proof, note that announcement 'C' is better than any other one (in particular, it is better than silence) for an EH and an H-sender if $\mu \cdot c \geq d$.

There are several intuitions behind this equilibrium. Note first that the selfish senders go for the maximal money payoff, and hence defect independently of the message sent. In contrast, the other senders are honest and cooperate if they announced it before (provided that cooperation is not too costly). When do they announce that? Clearly, only if the receiver is expected to cooperate as well. However, the only receivers who might cooperate are the conditional cooperators –i.e., the EH-types. Consequently, the EH and H-senders commit themselves to cooperate only if the share μ of EH-types is large enough. Since cooperative announcements are likely to be truthful in this case, EH-receivers reciprocate and cooperate as well.

We stress that the previous PBE is not the only equilibrium of the game. To start, one can find slight variations of the previous equilibrium, like an equilibrium in which EH-senders keep silent and then cooperate, while H-senders announce 'C' and cooperate afterwards. More importantly, there exist additional equilibria in which all types of players defect along the equilibrium path for any parameter constellation (these equilibria with unanimous defection are the correlate of the equilibrium in which all types defect in the simultaneous PD without pre-play communication; there exist multiple such equilibria because selfish types are then indifferent between announcements). For communication to raise cooperation, therefore, we need that players coordinate on a cooperative equilibrium and not on this kind of equilibria.

Prediction 3 is a very important implication of our assumption that people have a taste for honesty. Yet we note that other alternative hypotheses could explain it. In particular, communication could foster cooperation if it acts as a coordinating device, even if players have no concern for honesty. To illustrate this point, consider the simultaneous PD again, but now assuming that both players are sufficiently inequity-averse (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). It is well known that the game has two pure strategy equilibria (mutual cooperation and defection) in this case so that coordination is a key issue (recall that something similar happens with our model). If one-way communication is available, however, one could apply the hypothesis in Farrell and Rabin (1996) that self-committing, self-signaling messages are always trusted.[10] This restriction on post-message beliefs reduces the number of equilibria of the entire game and facilitates coordination on the most efficient equilibrium. Thus, a combined model (inequity aversion + communication refinement) predicts two *refined* equilibrium paths in the entire game. In one path, the sender announces 'C' and both players cooperate afterwards; in the other path the sender keeps silent and then mutual cooperation follows –announcing 'D' is never optimal because it leads to the 'bad' equilibrium (D, D). Since communication ensures cooperation, this could explain why communication increases efficiency. We will show later, though, that this coordination hypothesis fails to replicate other predictions of our model.

Experimental evidence for prediction 3: Under certain conditions, our model predicts a rise in average cooperation if one player is allowed to communicate. This is consistent with

[10] A message is self-committing if the sender wants to honor it in case she believes that the receiver believes it -the message must be part of an equilibrium strategy profile of the action stage subgame. For instance, both 'C' and 'D' are self-committing in the simultaneous PD game if both players are sufficiently inequity averse. Further, a message is self-signaling when the sender prefers the receiver to play a best response to it if and only if the message is true –e.g., both 'C' and 'D' are self-signaling messages in the PD game if both players are sufficiently inequity averse.

the available experimental evidence. For instance, Duffy and Feltovich (2002) report that the introduction of one-way cheap talk increases the rate of cooperation from 22% to 40% (in any information condition, subjects played ten times the same simultaneous PD game against different opponents).

Further, the model indicates that cooperation fundamentally depends on the *content* of the message: Nobody cooperates after sending or receiving a 'D' message, while some types cooperate after sending or receiving a 'C' message. The intuition here is twofold: (i) A significant number of people honor their word and (ii) some people respond reciprocally to messages. Consistent with all this, Duffy and Feltovich (2002) report that receivers condition their actions on the message they receive –i.e., they cooperate significantly more when they receive message 'C' that when they receive 'D' (50.4% vs. 16.1%).[11] Moreover, senders often announce truthful messages ('C' messages are truthful half of the time, and 'D' messages 85% of the time).

To finish, we predict that cooperation requires both γ and μ to be large enough ($\mu \geq \max\{\mu^{sim}, d/c\}$). Interestingly, and as μ^{sim} and d/c depend negatively on the monetary payoff c and positively on d (μ^{sim} also depends positively on t), it follows that cooperation becomes more unlikely if, say, the difference c-d decreases (recall that the model also predicts this phenomenon if PD players cannot talk). This might explain the results in Charness (2000) from a PD experiment with unilateral communication. The payoff calibration was such that c-d was rather small, and Charness reports that, although most senders announced cooperation, most senders and receivers defected afterwards.▪

Prediction 3 indicates that communication can raise cooperation. As we will see, however, the amount of the increase depends on several factors. One of them is the order of play –i.e., sequential or simultaneous- in the action stage subgame.

Prediction 4: The net effect of communication on cooperation depends on the order of play in the action stage. In particular, messages are ineffective if the sender moves before the receiver in the action stage.

Example: Consider the sequential PD when one-way communication is available. Does pre-play talk foster cooperation, as in the simultaneous PD? Interestingly, the answer depends on *who* sends the message. To start, communication improves nothing if the sender happens to move *first* in the action stage. In effect, recall from our equilibrium analysis of the sequential PD without pre-play talk (prediction 2) that first movers only cooperate if they expect the second mover to cooperate as well, something that requires in turn that the second mover is an EH type (and that γ is large enough). In this respect, it is clear that giving the voice only to the first mover cannot increase the probability that the second mover cooperates: Selfish second movers never cooperate, and the same happens

[11] Contrary to our model, though, both senders and receivers cooperated in the same proportion.

with the H second movers if they cannot communicate and make promises. Consequently, pre-play talk should not change the incentives to cooperate of the first mover: As in the case without prior communication, she will cooperate only if she expects to be matched with an EH second mover.

The scenario is rather different when the message sender is the *second* mover. To see this, suppose that the second mover is an H type. Recall from prediction 2 that such type of player does not cooperate if she cannot talk. If she can talk, in contrast, she should announce 'C' and subsequently cooperate in equilibrium if the mass μ of EH-types is larger than d/c. The intuition is that by *credibly* promising to cooperate, an H-second mover can entice the first mover to reciprocate and cooperate as well, thus earning a higher payoff than if she announces 'D' or keeps silent.

Experimental evidence for prediction 4: To sum up, unilateral communication should increase cooperation in a sequential dilemma if the second mover can communicate, but not when the first mover is the message sender. More generally, promises from agent A *to* agent B are useless if A makes *all* choices *before* B starts moving, as these promises do not add any relevant information (to put it like this, actions 'crowd out' words) and do not affect B's incentives to cooperate. The experimental results from Charness and Dufwenberg (2006) are very much consistent with this prediction. More precisely, they study a sequential dilemma with a random shock (this is immaterial for our results), and three of their treatments are of particular interest to us: A first one in which no subject could communicate, a second one in which the first mover could communicate, and a third one in which the second mover could communicate (subjects always communicated by means of free-form messages). The authors report that the percentages of mutual cooperation that these three treatments elicited were respectively 20%, 26%, and 50%. Consistent with our model, the first two percentages are not significantly different while the third one is higher than the others. Further, mutual cooperation in the third treatment was much higher following a statement of intent or promise than otherwise (recall that subjects were free to write whatever they wanted in their messages).

We note that the coordination hypothesis (Farrell and Rabin, 1996) cannot account for the increase in cooperation in the sequential PD. The reason is that most models of other-regarding preferences (like inequity aversion models) predict a unique equilibrium in the sequential PD, and hence coordination is not an issue here. The data from Charness and Dufwenberg (2006) is also at odds with the idea that communication works because it enhances group identity (in this case, it should not matter *who* talks for pre-play talk to be effective). In contrast, it can be explained as the result of *guilt aversion* (i.e., people do not like to let down other players' expectations), provided that promises affect beliefs, as Charness and Dufwenberg (2006) hypothesize.

As additional evidence, Charness and Dufwenberg (2006) report one phenomenon that can be explained by guilt aversion but not by our model. More precisely, first movers were asked to guess the proportion of second movers who would reciprocate if the first mover cooperated, while second movers were asked to guess the average guess made by the first movers who cooperated (both were paid for accuracy). A probit regression using the data from the second and third treatments then shows that a second mover's decision to cooperate is significantly correlated with her guess, but not with a dummy for the treatment. Our theory cannot explain such correlation –observe however that the model correctly forecasts a shift in the second mover's guess if she can communicate and make promises (as in the third treatment).[12] Yet we note two things in this respect. First, guilt aversion is not the only possible explanation for this correlation, as argued in Vanberg (forthcoming). Second, we report in the next section some evidence that is at odds with guilt aversion (or at least with a simple specification of the model) but not with our model –Vanberg (forthcoming) also provides evidence that suggests that people dislike breaking promises, and that is inconsistent with guilt aversion.▪

We have just shown that the effect of communication on cooperation may depend on the structure of the action stage. In addition, the structure of the communication stage –i.e., the communication protocol- may also play a role.

Prediction 5a: The net effect of communication on cooperation depends on the communication protocol employed. In particular, cooperation *sometimes* increases with the number of message senders.

Example: To illustrate this point, we consider the simultaneous PD when bilateral (two-way), simultaneous pre-play talk is available -i.e., *both* players send messages *simultaneously* in the communication stage. In this game, the EH-norm commends both players to announce 'C' (or keep silent) and to honor in any case their prior announcements. The H-norm allows both players to send any message but they should play according to it in the action stage. If one player keeps silent, the EH-norm commends her to cooperate afterwards, whereas the H-norm selects any move.

Following a similar reasoning as the one used in prediction 3, one can easily show that there exists an equilibrium where EH and H types announce 'C' and then cooperate if μ is large enough and $\gamma \geq 2(t-c)$ (selfish types announce cooperation as well but defect afterwards). We make several remarks on this equilibrium. *First*, since any H player cooperates in this equilibrium, the corresponding cooperation rate is larger than the highest rate in the PD game with one-way communication (where the H receivers do not cooperate, recall the analysis in prediction 3). Hence, our model predicts that, under certain conditions, cooperation should increase with the number of message senders. The

[12] Charness and Dufwenberg (2006, 1593) give additional arguments against a model assuming a fixed cost of lying (recall that we do not make this assumption). For instance, they claim that people do not suffer from lying in certain contexts, as when playing poker. However, this seems easily explainable by our approach: Implicitly, the rules (norms) of poker allow some deceptive use of language -it is indeed part of the fun of poker!

intuition is clear: More H-types can commit to cooperate if the communication structure is rich. *Second*, we yet stress that there exist additional equilibria in which everybody defects along the equilibrium path. Consequently, cooperation requires that the players coordinate on the cooperative equilibrium.

Third, it is crucial that the mass μ of EH types is large. In particular, this equilibrium does not exist if μ is small but ρ is large –i.e., if there exists a large mass of H types. To clarify this point, consider the extreme case $\rho = 1$ so that both PD players are H types. One can then show that, in equilibrium, both players must announce 'D' or silence and then defect. This might appear counterintuitive. At first sight, there seems to be another equilibrium in which both H players promise cooperation and cooperate afterwards if γ is high enough. Nevertheless, a bit of reflection indicates that this behavior is not sustainable because, whatever the co-player's announcement, an H player gets more money and feels no remorse if she announces defection and defects afterwards. In short, communication and honesty alone cannot generate cooperation in the PD. *Fourth*, a model based on the joint idea that people have social preferences –as in, say, Fehr and Schmidt (1999)- and that communication acts as a coordination device (Farrell, 1987; Farrell and Rabin, 1996) seems at odds with prediction 5a. In effect, since one-way communication should be sufficient to ensure coordination on the efficient equilibrium, it follows that two-way communication should not increase cooperation.

Although cooperation in the simultaneous PD game directly depends on the number of message senders, we note that more is not always better: In some games, cooperation does not increase as more people can communicate. This point, which is very related to prediction 4, can be illustrated with the <u>sequential</u> PD. Recall from prediction 4 that allowing the first mover in the sequential PD to announce his future move is ineffectual in raising the cooperation rate. Consequently, two-way communication generates more cooperation than one-way communication only if the first mover was the message sender with one-way communication.

Experimental evidence for prediction 5a: We are not aware of any economic experiment that compares cooperation levels in the simultaneous PD game with one-way and two-way communication. However, Sally (1995) meta-analysis suggests that two-way communication has a strong positive effect on cooperation. In effect, the author finds that the presence of discussion –a form of bilateral communication- in one-shot social dilemma games is highly significant, and on average raises the cooperation rate by more than 45 percentage points. In addition, and in line with our model, Orbell et al. (1990) contend that discussion is effective because it allows subjects to make multilateral promises to cooperate, which are significantly more binding than unilateral ones. More focused experimental research would be welcome.▪

Given all previous predictions, we can rank each treatment according to the frequency of mutual cooperation in equilibrium. Table 3 summarizes our results in this respect (in the sequential PD, we assume that the message sender with one-way communication is the second mover, we also assume in each treatment that players coordinate in the cooperative equilibrium). For instance, since only the EH types are expected to cooperate in the simultaneous PD with no pre-play talk (prediction 1), it follows that the frequency of mutual cooperation is μ^2. The reader can elaborate a ranking just by directly comparing these frequencies.

		Communication Stage		
		No Pre-play Talk	One-way Talk	Two-way Talk
Action Stage	Simultaneous PD	μ^2	$\mu^2 + \rho \cdot \mu$	$\mu^2 + 2\rho \cdot \mu$
	Sequential PD	μ	$\mu + \rho \cdot \mu$	$\mu + \rho \cdot \mu$

Table 3: Frequency of mutual cooperation in each communication treatment

Table 3 clearly shows, as we have stressed throughout the paper, that the order of play in the action stage affects the cooperation rate. Our following prediction suggests that the order of play in the communication stage *might* have an effect as well.

Prediction 5b: The net effect of communication on cooperation depends on the communication protocol employed. In particular, average cooperation might vary if the order in which players communicate changes (if there are multiple message senders).

Example: To think about this point, consider first the simultaneous PD game with bilateral pre-play communication. Clearly, bilateral communication can be simultaneous or sequential (one player sends first a message, her co-player observes it and sends afterwards another message). It is then a natural question which type of bilateral communication is more effective in fostering cooperation. The answer is that both work equally well here because, whatever the players' types, both mechanisms induce the same equilibria –as one can prove with a line of reasoning similar to that of prediction 3. To put it like this, bilateral communication achieves maximal cooperation in the simultaneous PD even if no one *has the first word*.

	C1	C2	D
C1	c, c	0, 0	0, t
C2	0, 0	c, c	0, t
D	t, 0	t, 0	d, d

Table 4: A Social Dilemma with Two Equally Efficient OutcomeS

Yet we believe that this result should not be generalized to *any* social dilemma. The two-player game at Table 4 coincides with the PD game of table 1 except that each player has available two possible cooperative moves (C1, and C2). Importantly, this game has two E-paths -(C1, C1) and (C2, C2)-, as both lead to the E-allocation (c, c). If players can talk before playing the game and they talk sequentially, our model predicts cooperation in a continuum of mixed strategy equilibria if the mass of EH types is large enough (the model has also other equilibria in which players defect along the equilibrium path; the reader can prove all this by applying a similar reasoning to that of prediction 3). In these equilibria, the first message sender randomizes with some probability between announcements 'C1' and 'C2', the second sender responds with the same announcement as the first one, and both players honor their words afterwards if they are principled –of course, we need $\gamma \geq 2(t - c)$ for this to be optimal.

One can also prove that the game has multiple mixed strategy equilibria if the players talk simultaneously and not sequentially. Consequently, players have to coordinate in both cases. Our point here is that coordination *seems* much more likely to succeed in the game with sequential talk that in the other one: Giving the first word to one player may help when the action stage subgame has multiple E-paths, as in the game of table 4. One possible reason might be that a focal point (Schelling, 1960) of the type 'first come, first served' helps to coordinate players' beliefs when there is an unequivocal first mover.

Experimental evidence for prediction 5b: Some evidence from coordination games is in line with this focal point idea. Consult the survey in Camerer (2003) on matching games.▪

4. Complicating the basic setting

The model described at section 2 can be extended in many ways. We propose here a number of them, and show them to be consistent with much available experimental evidence.

4.1 Communication about past actions

In the previous section, players exchanged messages about their future moves, that is, their intentions. In general, however, players can also exchange messages about players' past moves (for expositional reasons, we take this to include messages about Nature's moves, that is, random shocks). To analyze honesty in this kind of situations, we introduce a slight change in the H and EH-norms.

For this, note first that the interpretation of a message in this setting depends on its timing. Consider a message sent at information set h_1 and such that it announces action **a**$\in M(h_2)$ at some h_2. This message means 'the mover at h_2 played **a**' if h_2 precedes h_1, and 'the mover at h_2 would choose **a** if she had to move at h_2' otherwise. Taking into account this, we say that a message is a lie about *past actions* if it announces at least one previous action that the sender knows *not* to be on any *possible* previous path of play. More precisely, we have

Definition 5: Let h' denote any information set that precedes h. A message sent at h is a lie about *past* actions if it announces action **a**'$\in A(h')$ at h' and **a**' does not point towards h.

In other words, a message is a lie about past actions if it is impossible that someone acted as announced, and the sender knows that *for sure*. Admittedly, this is a restrictive definition: In common parlance, we often consider that a message about past behavior is a lie if the sender deems *most likely* that nobody acted as announced (*even* if he is not totally sure about this issue). Thus if one says 'Company X's recorded assets and profits were not inflated' but at the same time he believes that Company X most likely manipulated its financial statements, such utterance is commonly regarded as a lie. We leave aside this kind of subtleties, though, as they complicate the model and are not essential to explain the experimental phenomena that we consider later. In any case, the H-norm of definition 2 and the EH-norm of definition 4 can be now extended as follows:

Definition 6 (The H-norm): At any h where the mover can communicate, this correspondence selects silence and any other message that is not a lie about past actions. At any other h, it selects *action* $\mathbf{a} \in M(h)$ if the actor announced **a** previously, and the whole set $M(h)$ otherwise.

Definition 7 (The EH-norm): At any h where the mover can communicate, this norm selects silence and any other message that (i) is not a lie about past actions, and (ii) announces to play an E-action at any of the *sender*'s future information sets. At any other h, the norm selects (i) action $\mathbf{a} \in M(h)$ if the mover announced **a** previously, and (ii) any E-action of h otherwise -if there is no E-action, the norm selects the whole set $M(h)$.

As the reader can confirm, the only difference between these norms and the ones considered in section 2 is that these new norms forbid lies about past actions. This is obviously immaterial for the games analyzed in section 3, where players could only communicate their intentions. Nevertheless, this point is obviously important if players can talk about past actions, as we show in what follows with a series of predictions.

Prediction 6: People may transmit more information than a standard analysis would predict.

Example: To illustrate this point, consider a sender-receiver game as in Crawford and Sobel (1982). In this class of games, one player (the *sender*) has private information about the realization of some random shock –i.e., about Nature's previous move- and *must* send a message in this regard to another player (the *receiver*), who subsequently takes an action. We stress that the sender must send a message and hence cannot keep silent (as we note below, the predictions of the model would otherwise change). The monetary payoffs for both players depend on the action chosen by the receiver *and* the state of nature, but not on the message.

In general, both the H-norm (definition 6) and the EH-norm (definition 7) commend the sender to announce the actual realization of the random shock, while the H-norm (the EH-norm) commends the receiver to choose any action (any E-action). That is, the H-norm does not restrict the receiver's choices (because she cannot send any message), while the EH-norm is more restrictive if some of the actions available are not E-actions. With this in mind, and to simplify matters, we assume the following: (i) There are only two realizations of the random shock (states of nature A and B) and both are equally likely, (ii) the receiver can choose only between two actions (a and b), (iii) the *set* of available monetary allocations coincides for any state of nature, and (iv) the available allocations are equally fair. This setting is relatively simple to study and moreover we have some experimental evidence on it.

More precisely, suppose that the sender (receiver) earns m (M) if the action and the state of nature coincide –i.e., this corresponds to the cases (a, A) and (b, B)-, while the sender (receiver) earns M (m) otherwise ($\mathrm{M} > \mathrm{m}$). Note that the monetary incentives of the sender and the receiver are totally misaligned in this game. If all players were selfish, therefore, the sender should transmit the least information possible and thus choose uniformly between messages 'A' and 'B' –for a formal proof, see Sánchez-Pagés and Vorsatz (2007). In other words: Under the standard assumption that all players are self-interested, half of the messages should be false. When there is a large population of principled types, in contrast, most messages should be truthful. To be precise, the game has the following unique perfect Bayesian equilibrium if $\gamma > 2(\mathrm{M} - \mathrm{m})$ and $\mu + \rho > ½$:

- Whatever their type, receivers trust the sender's message and play the action that coincides with the state of nature stated in the message.
- Principled senders tell the truth, and selfish ones lie.

The proof of this result goes as follows: Since allocations (M, m) and (m, M) are the only feasible ones and both maximize function (1), it follows that both a and b are E-actions, and hence selected by the EH-norm. As the H-norm also selects both actions, it follows that a principled receiver gets the same payoffs as a selfish one –i.e., both seek to maximize money payoffs. With this in mind, it is clear that choosing the action that coincides with the announced state of nature is better than the opposite pattern of behavior if (note also that moving always a or b cannot be optimal either)

$$(1 - \mu - \rho)\cdot\mathrm{m} + (\mu + \rho)\cdot\mathrm{M} > (1 - \mu - \rho)\cdot\mathrm{M} + (\mu + \rho)\cdot\mathrm{m}\ .$$

This implies that receivers should trust the sender's message and play a best response to it if $\mu + \rho > ½$. With respect to the senders, principled ones should tell the truth –i.e., announce 'A' ('B') when state of nature is A (B)- if the subsequent payoff is larger than the payoff for lying, that is, if $\mathrm{m} > \mathrm{M} - \gamma/2$. Note that principled senders decide not to lie even if they get the lowest possible monetary payoff as a result; this explains why the equilibrium is unique. On the other hand and given the receiver's behavior, selfish senders clearly maximize their monetary payoff if they lie. This finishes the proof.

This analysis can be easily extended to cover other cases, as one where the sender (receiver) earns m (M) if the action and the state of nature do *not* coincide. One can also find equilibrium predictions for other parameter constellations than the ones considered before. Thus, in case there are few principled players ($\mu + \rho < ½$) but they are 'principled enough' –that is, γ > 2(M – m)-, principled senders tell the truth and selfish ones randomize their announcements so as to leave receivers uninformed. More precisely, selfish senders tell the truth with some probability θ so that receivers are indifferent between choices a and b whatever the state of nature:

$$(1-\mu-\rho-\theta)\cdot m+(\mu+\rho+\theta)\cdot M=(1-\mu-\rho-\theta)\cdot M+(\mu+\rho+\theta)\cdot m\ .$$

The reader may compute in this case the probability with which receivers should choose each move so that selfish senders play a best response by randomizing with probability θ. Finally, there exist multiple equilibria if γ < 2(M – m) for the principled senders, as they may find sometimes profitable to lie –to get some intuition of this result, consult Sánchez-Pagés and Vorsatz (2007) for a theoretical analysis of the game when *all* players are selfish.

Finally, we note that our prior equilibrium analysis crucially hinges on the assumption that senders cannot keep silent. To illustrate this point neatly, consider the sender-receiver game previously studied, but assume now that the sender can only send a truthful message or stay silent –e.g., if state of nature is A, she can only send message 'A' or keep silent. Equilibrium predictions in this case are straightforward once one notes that both the H-norm and the EH-norm allow the sender to submit the truthful message *or* stay silent. First, and whatever her type, the receiver obviously maximizes her payoff by trusting the message (in case she receives one), while she is indifferent between any action if she receives no message. Given this, both selfish and principled senders should remain silent. Intuitively, principled types suffer a cost when they utter a false message, but not when they conceal information (both norms allow silence). As a result, they maximize their payoff by providing no information, that is, exactly as a standard analysis would predict.[13]

Experimental evidence for prediction 6: Gneezy (2005) reports lab data from two sender-receiver games which are similar to the one analyzed before (when silence is not permitted): A first one in which M = 6 and m= 5 and a second one such that M = 15 and m= 5 (all the reported amounts are in US dollars). The only difference from our prior analysis is that receivers were never informed about the values of M and m –in fact, they did not know anything about the actual payoffs, even that they were inverse, and senders could only tell them whether a or b were payoff-maximizing. To properly study the receiver's equilibrium behavior in this setting, therefore, one should make the analysis conditional on the receiver's beliefs about M and m.

[13] In other words, principled people in our model care about lies of commission, but not about lies of omission (silence). We suspect, however, that some actual senders dislike lies of omission and hence would tell the truth in this game (provided that m and M are close enough). An experimental analysis of this game could provide evidence on this respect.

To simplify the exposition, we then focus on the senders' behavior. Gneezy (2005) indicates that among 50 senders who were asked to guess the receiver's reaction to their message (they were paid for accuracy), a majority of them (82 percent) expected the receiver to trust their message –as indeed mostly happened.[14] For this group of players, our model predicts that selfish senders should tell a lie while principled types should not, provided that γ is large enough. Indeed 36 percent of the senders lied when M = 6, whereas that number rose to 52 percent when M = 15. Observe that our model can explain this significant increase: As principled senders tell the truth if $\gamma > 2(\mathrm{M} - \mathrm{m})$, they are less likely to do that as the difference M - m increases (this is particularly true if one assumes some heterogeneity in the distribution of γ).

Again consistent with our model, additional evidence from Gneezy (2005) suggests that people tell the truth because they have some kind of preference for that. In a dictator game in which dictators could choose between (dictator, dummy) allocations (M, 5) and (5, M), Gneezy reports that 66 and 90 percent of the dictators chose their payoff-maximizing allocation when M = 6 and M = 15, respectively. Since these fractions are much higher than the corresponding percentages of deception mentioned in the previous paragraph for sender-receiver games with identical payoffs, it seems reasonable to assert that "it is not only care for others that motivate behavior, but also aversion to lying" (Gneezy, 2005, p. 388).[15]

We stress that Gneezy used a between-subjects design, that is, subjects played only one of the games (i.e., either one of the sender-receiver games or one of the dictator games). Further, the available allocations (M, m) and (m, M) in the above mentioned treatments were equally fair –at least according to the E-function (1). In contrast, participants in Hurkens and Kartik (forthcoming) played both a sender-receiver game and a payoff-preserving dictator game in which only one of the two available allocations was an E-allocation -as in Gneezy (2005), receivers were uninformed about the payoff constellation. For instance, the available (sender, receiver) allocations were (4, 12) and (5, 4) in one treatment, so that (4, 12) is the only E-allocation. Interestingly, this treatment allows us to separate our three types in equilibrium (a proof of this is left to the reader). In

[14] The paper does not specify in which treatment these 50 senders participated –i.e., the value of M-, but this seems immaterial since senders knew that receivers were totally uninformed on this issue.

[15] Our model predicts that any type of dictator should choose her payoff-maximizing allocation in both dictator games and it is then consistent with the difference in behavior between the dictator games and the corresponding sender-receiver games. However, it fails to explain why some dictators chose to be generous with the co-player. We speculate that they felt obliged to follow a norm of courtesy or chivalry like: "Among two fair allocations, choose the one that favors your co-player". This could additionally explain why there was less compliance with this norm when its cost increased –i.e., when M = 15.

effect, provided that most receivers trust the message in the sender-receiver game (as it indeed happened) and that γ and $\mu + \rho$ are large enough, selfish types should choose allocation (5, 4) in the dictator game and lie in the sender-receiver game, H-types should choose allocation (5, 4) in the dictator game and tell the truth in the sender-receiver game, and EH-types should choose allocation (4, 12) and tell the truth in the sender-receiver game. In line with this, the authors report that around 33% of the senders/dictators acted as our selfish types, 43% as our H-types, and 19% as our EH-types (the remaining 5 % of the subjects chose the E-allocation and lied).[16] Consequently, this evidence is well in line with our 3-types assumption.

Additional studies also show that some people tell the truth even if that goes contrary to their material interest, especially if the cost is not high. In Sánchez-Pagés and Vorsatz (2007), participants play 50 times with re-matching a similar sender-receiver game to that of Gneezy (2005) –receivers are informed about the values of M and m, though. Although the theoretical analysis of this repeated game is complex, the fact that along the last 40 rounds both the fraction of truthful messages and that of trusting behavior are significantly above the standard prediction of 50 percent seems in line with our model, even if the authors did not include in the analysis the data from the first 10 rounds, when the rate of truth-telling was much higher.[17] In line with our model, the rate of deception increased when M increased from 2 to 9 points, while keeping m constant. See also Cai and Wang (2006) for additional evidence.▪

Prediction 7: Truth-telling may decrease after a history of deception.

Example: Consider again the sender-receiver game of prediction 6, now assuming that the players play it twice and that they change roles after playing the first round –more precisely, we assume without loss of generality that player 1 is sender in the first round and receiver in the second round. The analysis of this game will allow us to show that principled types care about previous history: They are less likely to tell the truth if they were deceived before –consult Sen (1997) on history-dependent preferences.

To simplify the analysis, we assume that the parameter constellation is the most propitious for principled senders to tell the truth (see prediction 6). Furthermore, we focus our analysis on the second round, assuming that player 1 deceived player 2 in the first round. This means that player 1 deviated from any of our two norms (see definitions 6 and 7) in the first round, and hence no principled player 2 suffers any psychological cost if she deviates as well in the second round. Consequently, the utility of any principled player in this subgame coincides always with her monetary payoff –i.e., they have the same

[16] As Hurkens and Kartik (forthcoming) note, a selfish sender would tell the truth if she (incorrectly) believed that the receiver is most likely to distrust her message. Due to this, the 43% of subjects who told the truth and chose (5, 4) might be an overestimation of the actual proportion of H-types.

[17] Paradoxically, receivers did not anticipate so much truth-telling during those first 10 rounds, that is, they did not trust the messages in a proportionate manner.

incentives as a selfish type. As a result, we can apply here a result in Sánchez-Pagés and Vorsatz (2007) which indicates that the equilibrium rate of truth-telling in this subgame should be 50 percent if all players are selfish. Since this rate is smaller than the one predicted in the one-shot game if $\mu + \rho >$ ½ (see prediction 6), our model hence predicts a decline in truth-telling after a history of deception.

This result provides further insights into players' preferences for honesty. Our model assumes that principled people are (conditionally) lie-averse, but a natural, alternative hypothesis is a fixed cost of lying (see for instance Ellingsen and Johannesson, 2004), which could be modeled by assuming that principled types get a payoff of $x_i(z) - \gamma$ when deviating from their respective norms and $x_i(z)$ otherwise. In other words, this hypothesis indicates that principled people feel badly when they lie, independently of what others do. It is clear that this hypothesis cannot predict any change in behavior depending on previous history, and hence is inconsistent with our prior result. Experimental evidence on this regard should be welcome.▪

4.2 Adding anger to the model

According to Thomas Hobbes, "covenants without the sword are nothing but words". While the experimental evidence that we have reviewed so far is somehow inconsistent with this pessimistic point of view (some people honor their word even if they expect no sanctions otherwise), introspection suggests that one can foster truth-telling by threatening to sanction liars. However, this raises a question, that is, given that using the 'sword' is often costly, why do people punish? Perhaps one reason is that they feel angry at cheaters. We can model this idea by slightly changing principled (i.e., either EH or H types) players' utility function:

$$u_i(z) = \begin{cases} x_i(z) - \gamma \cdot \mathrm{r}(\psi, \mathrm{z}) & \text{if } i \notin \mathrm{R}(\psi, \mathrm{z}),\ (0< \gamma) \\ x_i(z) - \alpha \cdot \max\limits_{j \notin R(z)} \{x_j(z)\} \cdot I(z) & \text{if } i \in \mathrm{R}(\psi, \mathrm{z}),\quad (0< \alpha \leq 1) \end{cases}$$

where $I(z)$ is an indicator function that takes value 0 if nobody deviates –i.e., if $R(z) = N$ - and 1 otherwise. That is, we assume that principled players get angry when someone deviates from the norm that they find binding, and want to punish the well-off deviator (this assumption is probably unrealistic but sufficient in two-player games, which are our focus here; more complex patterns could be easily introduced). In this regard, parameter α can be interpreted as the maximum amount of money that an angry player is willing to spend in order to reduce the earnings of the best-off deviator in one monetary unit. The available experimental data on punishment suggests that people are rarely willing to punish if the cost is larger than the harm imposed, which explains assumption $0< \alpha \leq 1$ –see López-Pérez (2008) on this. Importantly, note that principled players do not get angry if they breached the norm themselves; intuitively, a deviant player should not get annoyed at someone who misbehaved as her. The interested reader may consult López-Pérez (2008) for a more detailed discussion of these assumptions.

Prediction 8: People may punish co-players for deceiving them, even if punishment is costly. The availability of sanctions may foster truth telling.

Example: Assume that the sender-receiver game of prediction 6 has now an additional stage in which, after discovering whether the sender's message was truthful, the receiver can spend money out of her endowment to punish the sender. More precisely, each monetary unit spent by the receiver reduces the sender's payoff in $p>0$ monetary units. If $\gamma > 2(M - m)$ and $\mu + \rho > ½$ the game has a unique perfect Bayesian equilibrium (except marginal cases):

- Selfish receivers never punish, while principled ones spend their endowment to punish the sender if $1 < p\cdot \alpha$ and the sender lied before (they spend no money otherwise).
- Whatever their type, receivers trust the sender's message and play the action that coincides with the state of nature announced in the message.
- Principled senders tell the truth. Selfish senders lie if $1 > p\cdot\alpha$ or $\mu + \rho < \frac{M - m}{m\cdot p}$, and tell the truth otherwise.

The proof of this result goes as follows: To start, it is clear that selfish receivers should never spend money to reduce the other player's payment. In turn, principled receivers should feel angry at the sender only if she deviated from the H or the EH-norm. In this game, that occurs only if the sender lied before. If this is the case, they should punish the sender if $m - \alpha M < -\alpha (M - mp) \Leftrightarrow 1 < p\cdot \alpha$ (we are implicitly assuming here $M - mp > 0$, that is, the receiver cannot reduce the sender's earnings to zero; the receiver would not spend her whole endowment otherwise).

Further, an argument similar to that of proposition 2 indicates that, whatever their type, receivers maximize their payoff if they trust the sender's message, and that principled senders should tell the truth if $\gamma > 2(M - m)$. Finally, selfish agents would rather tell the truth if the probability that they are punished (which, recall, requires $1 < p\cdot\alpha$) is high enough, that is, if

$$m > (\mu + \rho)\cdot(M - mp) + (1 - \mu - \rho)\cdot M \Leftrightarrow \mu + \rho > \frac{M - m}{m\cdot p},$$

while they should lie otherwise (they are indifferent if the previous expression holds with equality; in this marginal case there are multiple equilibria). This finishes the proof.

Experimental evidence for prediction 8: The previous analysis is consistent with some of the experimental results from Sánchez-Pagés and Vorsatz (2007). *First*, there is substantial punishment and most of it is directed towards liars –interestingly, though, receivers do not punish liars if they did not trust them before, that is, if they were not deceived. *Second*, the rate of punishment positively depends on its effectiveness: The rate of punishment of a lie was equal to 42.8% if $p=9$ and equal to 25.2% if $p=2$. *Third*, punishment is primarily a response to deception, not to the distribution of material payoffs. Thus, the rate of punishment is equal to 13.4% after a sincere message that was distrusted, but rises to 42.8% after a lie that was trusted –this data refers to the treatment when $p=9$, but a similar increase occurs if $p=2$. However the (sender, receiver) allocation was (9, 1) in both situations, which indicates that the increase in punishment was the result of deception and not, say, of the inequality in players' earnings –hence, this evidence goes contrary to models of inequity aversion like Fehr and Schmidt (1999), Bolton and Ockenfels (2000), or more generally to any model that assumes that punishment should *only* depend on the available money vectors

Fourth, both truth-telling and trusting tend to increase if punishment is available. *Fifth*, truth-telling decreases if its cost increases, that is, if the difference $M - m$ increases, even if punishment is available. *Sixth*, and consistent with our hypothesis that principled types should be the only ones who punish, Sánchez-Pagés and Vorsatz (2007) indicate that those subjects who punish liars most are also most likely to tell the truth (the authors could study this issue because subjects rotated roles in their treatment). ▪

Our previous analysis considered the sanction of deception. However, the EH-types may also punish unfair behavior –i.e., deviations from the EH-path of a game. Further, games with punishment opportunities and communication introduce an interesting variable, that is, threats to punish. When are they credible? We illustrate these two points in what follows.

Prediction 9: Unfair behavior is often punished, and that tends to reduce self-interested behavior and foster players' trust. Threats to punish unfair behavior are credible if the proportion of principled types is large and the cost of punishment is sufficiently low.

Example and experimental evidence: We consider the two-player hold-up game of Ellingsen and Johannesson (2004). In this game, player 1 (the *seller*) chooses first whether to invest in some project, thus incurring in a sunk cost of 60 Swedish kronor (SEK). Both players get zero money if the seller does not invest, while they pass to play an ultimatum game with stake size 100 SEK otherwise (this money represents the revenues from the investment). That is, player 2 (the *buyer*) can now offer some money out of the 100 SEK, and player 1 can accept or reject that offer. The sharing is implemented if player 1 accepts, while no player gets any money if player 1 rejects (in addition and independently of her decision here, player 1 must always pay the sunk cost of the investment).

Ellingsen and Johannesson (2004) experimentally study three treatments. In the first treatment, there is no communication. In the second one, player 2 can send a message before the seller makes the investment decision. In the third treatment, player 1 can send a message if she chooses to invest (no restrictions were put on the content of the message). Clearly, the second treatment invited promises while the third one invited threats.

What does our model predict in each treatment? Observe first that the overall game has a unique E-allocation, which is obtained if player 1 invests, player 2 offers 80 SEK and player 1 accepts (once the sunk cost is discounted, both players end up earning 20 SEK in such a way). For this reason, the EH-norm commends to follow that path of play in the game with no communication, while one should announce to follow that path and honor always her prior announcement if she can communicate. Further, the H-norm only restricts behavior if someone can send messages, and commends to play as announced before.

Taking into account this and the induced utility payoffs, our model replicates the four main experimental results reported by Ellingsen and Johannesson (2004).[18] *First*, low offers are often rejected. According to our model, this occurs in any treatment when an EH-seller gets angry at a deviation from the (seller, buyer) 'E-sharing' (80, 20) and her anger intensity α is high enough (however and since α is never larger than one, player 1 should never reject an offer larger than 50), or when an H-seller has threatened to reject low offers in the treatment with seller communication and her parameter γ is large (hence, our analysis predicts more rejections of low offers in the seller communication treatment than in the no-communication treatment, which is in line with the reported evidence).

Second, many agents propose an equal split of the net surplus, that is, sharing (80, 20). They are the EH-buyers, who feel committed to follow the EH-norm if their parameter γ is large. Furthermore, and consistent again with the data, our model predicts that the average offer will be larger when communication is available. In the seller communication treatment, H-sellers can credibly threat to reject low offers; as a result selfish buyers raise their offers with respect to the no communication treatment. In the buyer communication treatment, in turn, promises are binding for the H-buyers and they must be generous enough to induce the seller to invest (the data indicates that buyers often made explicit promises in their messages, and they rarely violated them).

Third, communication increases investment. On one hand, H-buyers can commit to make a generous offer in the buyer communication treatment, and that should entice sellers to invest. On the other hand, sellers can ensure higher offers by credibly threatening to reject low ones in the seller communication treatment. This should raise the investment rate as well. *Fourth*, promises are more credible than threats in this game. More precisely, promises to make a generous offer in the buyer communication treatment are not violated, but threats to reject offers lower than 80 are often neglected in the seller communication treatment. As an illustration of this second point, the authors report an extreme case in which one seller accepted a (30, 70) split even if she had threatened

[18] To shorten the exposition, we do not provide a complete equilibrium analysis here, but one can get some hints about this from the detailed study of the ultimatum game without prior communication in López-Pérez (2008).

before not to accept lower offers than (80, 20). This example suggests why such threats are not credible: They are rather costly to honor and the effectiveness of punishment is not very high –this is especially true if the offer is larger than 50; in fact, and as we noted above, our model predicts that no such offer will ever be rejected if $\alpha \leq 1$. Promises, in contrast, are binding if parameter γ is large enough. Since we have not restricted the upper value of this, promises can be in principle 'more binding' than threats.

Finally, we note that our model can also explain much of the evidence from Brandts and Charness (2003). In the two-player game that the authors studied, player 1 announces her intended move and then each player chooses between a fair and an unfair move; afterwards player 2 can either punish or reward player 1. Consistent with our model, they show that deception is significantly punished (see also Ostrom et al., 1992). In addition, Ellingsen et al. (2006) show that people cease to be cooperative and honest with others who did not cooperate before. This is very well in line with two important intuitions of our model, that is, path-dependency (recall prediction 7) and the idea that people respect norms reciprocally. ▪

4.3 Other possible extensions

We may consider three additional extensions. *First*, the model assumes that principled players' bad feelings do not depend on the specific deviation that they make from their binding norm. But remorse might be higher depending on the material consequences of the deviation –e.g., cheating in a medical article might generate more remorse than cheating in a paper on ancient history!

In fact, Gneezy (2005) provides some evidence that might be consistent with this (see however Hurkens and Kartik, forthcoming). In one sender-receiver game in which, whatever the state of nature, the receiver could choose between one action leading to (sender, receiver) allocations (5, 15) and (6, 5) –as in the other treatments, though, the receiver was not informed about the available payoffs- only 17 percent of the senders lied. This contrasts with the previously mentioned 36 percent of senders who lied when $\mathrm{M} = 6$ and $\mathrm{m} = 5$, that is, truth-telling could depend not only on its price ($1 in both treatments), but also on the receiver's loss ($10 in one treatment, and $1 in the other). To explain this sort of phenomena, one could assume that parameter γ positively depends on the difference between the other's payoff *had one respected the norm* and her actual payoff.

Second, we mentioned in the introduction two possible reasons why communication fosters cooperation: Group identity and social norms. Although this paper shows that a parsimonious model based on social norms can explain much evidence, this is not to say that group identity plays no role. It seems plausible that people from group X are more likely to cooperate with someone who *declares* to be a member of group X than with another person. One could introduce this into the model by making parameter γ depend on the identity of the co-player(s). Note, however, that norms of honesty somehow predate group identity: To understand why saying 'I belong to group X' has an effect on the others, one must first be able to explain why they believe that message to be sincere.

Third, our norms are restrictive in that they forbid telling white lies –i.e., lies that, if believed by the receiver, will benefit him- and lies that induce moral behavior. For the first case, think of a doctor who tells a reassuring lie to an ill patient. For the second one, think of the French priest who, when asked by a Nazi official whether he had hidden some fugitive Jews in his church, lied by saying that he had not. Our norms would forbid that lie, but most of us would agree that it was commendable: Had the priest told the truth, the official would have behaved morally wrongly. One can model all these ideas by introducing a norm whose prescriptions depend on the actor's beliefs about other players' future actions. An example is a norm that tolerates a lie when one expects that another player will respect the EH-norm if he is deceived.

5. Conclusion

Much experimental evidence confirms that communication fosters cooperation. This is difficult to explain using standard game-theoretical hypotheses, which largely imply that "concepts like lying, credibility, and credulity –all essential features of strategic communication- do not have fully satisfactory operational meanings".[19] In this paper we argue that communication works *mainly* because it allows promises and because many people care about social norms of cooperation and honesty in a reciprocal manner.[20]

Our approach throws light on the closely related issues of truth-telling, deception, and credibility. When is someone expected to lie? There are two crucial requisites here. *First*, the person should not care (much) about lying. According to our model, people are more likely to lie if the expected monetary gain increases -e.g., when the stake size is large, the relationship non-repeated (think of house negotiations), and the likelihood of being detected by an angry and revengeful co-player low- or if sufficiently many others are expected to lie or behave unfairly.[21] *Second*, she should expect others to trust her, as her lie would be ineffective otherwise. Heterogeneity is crucial in this regard: Since players' types are private information and a significant part of the population is expected to be honest (if truth-telling is not too costly), dishonest guys find easier to cheat others.

The model can be used to study how pre-play communication affects phenomena as diverse as bargaining, collusion between firms, conflict, charity giving, revolutions, team

[19] Crawford and Sobel (1982, p. 1450). Among other extensions of their model, they suggest "allowing lying to have costs for [sender] S, uncertain to [receiver] R, in addition to those inherent in its effect on R's choice of action" (*ibid.*). Our paper follows this line.

[20] Emotions like shame and guilt seem crucial here, and our model implicitly recognizes this. Indeed, many techniques for detecting lies rely on the idea that lying generates some emotional anxiety. For an article on lie detection, consult *The Economist*, July 8th 2004.

[21] In relation with this, anecdotal evidence indicates that some professional groups like politicians and lawyers are expected to lie more frequently than others like doctors or professors; even if this image was false, it might become a self-fulfilling prophecy if many people come to believe it.

behavior, and voting –e.g., why do voters care about promises by politicians? As an illustration, we make two remarks on revolutions and demonstrations, which are typical examples of social dilemmas (participants in these events usually demand the provision of public goods, like income redistribution or changes in the political system).

First, our model indicates that people are more likely to join the mass –i.e., to cooperate- if pre-play communication is available: By announcing their intention to cooperate or by announcing that they are already cooperating, people can encourage others to cooperate as well. For this reason, cooperation positively depends on the number of message receivers and senders, which implies in turn that groups have an interest in controlling mass media. This partly explains, for instance, why rebels in revolutions or military takeovers often try to control broadcasting stations from the first moments of the rebellion, and why their opponents strive to keep them isolated.[22] The following excerpt from the September 29th 2007 issue of *The Economist* about the popular revolts against the military dictatorship in Myanmar (former Burma) is enlightening in this respect:

> "One genuine difference [with previous pro-democracy protests in Myanmar] is that, in the age of the internet and digital cameras, images of the spectacular protests in Yangon, the main city, have spread at lightning speed across Myanmar itself, *encouraging* people in other towns to stage demonstrations of their own; and around the world […]" (our emphasis; the military junta largely cut internet connections and mobile-phone lines a few days later).

Second, those groups for which multilateral communication is very costly or unfeasible –at least among large numbers of people- should engage in less collective action. For instance, K. Marx and F. Engels noted in *The Communist Manifesto* that revolutions were less likely among peasants than among proletarians, one likely reason being that geographical distance tended to hinder communication and, as a result, cooperation (indeed, Marx and Engels believed that capitalism was 'digging its own grave' because of its tendency to concentrate workers in large-scale industries).

To finish, we would like to note that communication does not only give an opportunity for making promises; but also one for teaching, thus raising productivity, and for discussing moral issues with others –some social researchers have speculated that dialogue might have a positive effect in avoiding conflict. To understand all this, however, one must understand first why people believe (or not) what others say. This article offers some insights in this respect.

Bibliography

- Aumann, R. (1990). “Nash Equilibria are not Self-Enforcing”, in Gabszewicz, J. J., J. F. Richard and L. A. Wolsey, eds., *Economic Decision-making: Games, Econometrics and Optimisation*, Elsevier, pp. 201-6.
- Aumann, R., and S. Hart (2003). “Long Cheap Talk”, Econometrica 71, 1619-1660.
- Becker, G. (1996). *Accounting for Tastes*, Harvard University Press.
- Bicchieri, C. (2002). “Covenants Without Swords: Group Identity, Norms, and Communication in Social Dilemmas”, Rationality and Society, 14(2), 192-228.
- Bolton, G. E., and A. Ockenfels (2000). “ERC: A Theory of Equity, Reciprocity, and Competition”, American Economic Review, 90(1), pp. 166-93.
- Brandts, J., and A. Schram (2001). “Cooperation and Noise in Public Good Experiments: Applying the Contribution Function Approach”, Journal of Public Economics 79, 399-427.
- Brandts, J., and G. Charness (2003). “Truth or Consequence: An Experiment”, Management Science 49, 116-130.
- Cai, H., and J. Wang (2006). “Overcommunication in Strategic Information Transmission Games”, Games and Economic Behavior, 95, 384-394.
- Camerer, C. (2003). *Behavioral Game Theory-Experiments in Strategic Interaction*, Princeton University Press.
- Charness, G., (2000). “Self-Serving Cheap Talk: A Test of Aumann’s Conjecture”, Games and Economic Behavior, 33, 177-194.
- Charness, G., and M. Rabin (2002). “Understanding Social Preferences with Simple Tests”, Quarterly Journal of Economics, 117, 817-869.
- Charness, G., and M. Dufwenberg (2006). “Promises and Partnerships”, Econometrica, 74(6), 1579-1601.
- Clark, K., S. Kay, and M. Sefton (2001). “When are Nash Equilibria Self-Enforcing? An Experimental Analysis”, International Journal of Game Theory 29, 495-515.
- Cox, J., D. Friedman, and S. Gjerstad (2007). "A Tractable Model of Reciprocity and Fairness", Games and Economic Behavior, 59, 17-45.
- Crawford, V., and J. Sobel (1982). “Strategic Information Transmission”, Econometrica 50, 1431-1452.
- Crawford, V. (1998). “A Survey of Experiments on Communication via Cheap Talk”, Journal of Economic Theory 78, 286-298.
- Croson, R. T. A. (2000). "Thinking like a Game Theorist: Factors affecting the Frequency of Equilibrium Play", Journal of Economic Behavior and Organization, 41, 299-314.

[22] Of course, control of the media is not only important for cooperation but also for coordination. We leave this issue for further research.

- Demichelis, S. and J. W. Weibull (2008). "Language, Meaning, and Games: A Model of Communication, Coordination, and Evolution", American Economic Review 98(4), 1292-1311.
- Duffy, J., and N. Feltovich (2002). "Do Actions Speak Louder Than Words? Observation vs. Cheap Talk as Coordination Devices", Games and Economic Behavior, 39, 1-27.
- Dufwenberg, M., and G. Kirchsteiger (2004). "A Theory of Sequential Reciprocity", Games and Economic Behavior, 47, 268-98.
- Ellingsen, T., and M. Johannesson (2004). "Promises, Threats, and Fairness", Economic Journal, 114, 397-420.
- Ellingsen, T., M. Johannesson, J. Lilja, and H. Zetterqvist (2006). "Trust and Truth", mimeo.
- Elster, J. (1989). "Social Norms and Economic Theory", Journal of Economic Perspectives, 3(4), 99-117.
- Elster, J. (1999). *Alchemies of the Mind: Rationality and the Emotions*. Cambridge University Press.
- Falk, A., and U. Fischbacher (2006). "A Theory of Reciprocity", Games and Economic Behavior 54, 293-315.
- Farrell, J. (1987). "Cheap Talk, Coordination, and Entry", RAND Journal of Economics, 18(1), 34-39.
- Farrell, J., and M. Rabin (1996). "Cheap Talk", Journal of Economic Perspectives, 10, 103-118.
- Fehr, E. and K. Schmidt (1999). "A Theory of Fairness, Competition and Cooperation", Quarterly Journal of Economics, 114(3), 817-68.
- Fehr, E. and S. Gächter (2000). "Fairness and Retaliation: The Economics of Reciprocity", Journal of Economic Perspectives, 14, 159-81.
- Fehr, E., and K. Schmidt (2006). "The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories", in S. C. Kolm, and J. M. Ythier (Eds.), *Handbook of the Economics of Giving, Altruism and Reciprocity*, Volume 1, Elsevier B. V.
- Fischbacher, U., S. Gächter, and E. Fehr (2001). "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment", Economics Letters 71, 397-404.
- Geanakoplos, J., D. Pearce, and E. Stacchetti (1989). "Psychological Games and Sequential Rationality", Games and Economic Behavior 1, 60-79.
- Gneezy, U. (2005). "Deception: The Role of Consequences", American Economic Review 95(1), 384-394.
- Harsany, J. C., and R. Selten (1988). *A General Theory of Equilibrium Selection in Games*, MIT Press.

- Hayashi, N., E. Ostrom, J. Walker, and T. Yamagishi (1999). "Reciprocity, Trust and the Sense of Control: A Cross-Societal Study", Rationality and Society 11, 27-46.
- Hurkens, S., and N. Kartik. "Would I Lie to You? On Social Preferences and Lying Aversion", forthcoming in Experimental Economics.
- Kartik, N. (2008). "Strategic Communication With Lying Costs", mimeo.
- Ledyard, J. (1995). "Public Goods: A Survey of Experimental Research", in J. Kagel and A. E. Roth (Eds.), *Handbook of Experimental Economics*, Princeton Univ. Press.
- Levine, D. K. (1998). "Modeling Altruism and Spitefulness in Experiments", Review of Economic Dynamics, 1, 593-622.
- López-Pérez, R. (2005). "Shame and Guilt: Their Effect on Preferences," mimeo.
- López-Pérez, R. (2008). "Aversion to Norm-Breaking: A Model", Games and Economic Behavior, 64, 237-267.
- Miettinen, T. (2005). "Promises and Lies: A Theory of Pre-Play Negotiations", Mimeo.
- Orbell, J., R. Dawes, and A. van de Kragt (1990). "The Limits of Multilateral Promising", Ethics 100, 616-627.
- Ostrom, E., J. Walker, and R. Gardner (1992). "Covenants with and without a Sword: Self-Governance is Possible", American Political Science Review, 86(2), 404-417.
- Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics", American Economic Review 83, 1281-1302.
- Rabin, M. (1994). "A Model of Pre-game Communication", Journal of Economic Theory 63, 370-391.
- Rapoport, A. and A. M. Chammah (1965). *Prisoner's Dilemma: A Study in Conflict and Cooperation*. Ann Arbor, MI: University of Michigan Press.
- Sally, D. (1995). "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992", Rationality and Society 7:1, 58-92.
- Sanchez-Pagés, S. and M. Vorsatz (2007). "An Experimental Study of Truth-Telling in a Sender-Receiver Game", Games and Economic Behavior 61, 86-112.
- Schelling, T. (1960). *The Strategy of Conflict*. Cambridge: Harvard University Press.
- Sen, A. (1997). "Maximization and the Act of Choice", Econometrica 65, 745-779.
- Vanberg, C. "Why do people keep their promises? An experimental test of two explanations", forthcoming in Econometrica.