Edinburgh School of Economics

Discussion Paper Series

Number 128

An Experimental Study of Truth-Telling in a Sender-Receiver Game

Santiago Sánchez-Pagés (University of Edinburgh)
Marc Vorsatz (Universitat Autonoma de Barcelona)

Date
November 2004

Published by
School of Economics
University of Edinburgh
30 -31 Buccleuch Place
Edinburgh EH8 9JT
+44 (0)131 650 8361
http://www.ed.ac.uk/schools-departments/economics

THE UNIVERSITY *of* EDINBURGH

An Experimental Study of Truth-Telling in a Sender-Receiver Game*

Santiago Sánchez-Pagés† and Marc Vorsatz‡

November 10, 2004

Abstract

A recent experimental study of Cai and Wang [5] on strategic information transmission games reveals that subjects tend to transmit more information than predicted by the standard equilibrium analysis. To evidence that this overcommunication phenomenon can be explained in some situations in terms of a tension between normative social behavior and incentives for lying, we show that in a simple sender-receiver game subjects incurring in costs to punish liars tell the truth more often than predicted by the equilibrium analysis whereas subjects that do not punish liars after receiving a deceptive message play equilibrium strategies. Thus, we can partition the subject pool into two groups, one group of subjects with preferences for truth-telling and another group taking into account only economic incentives.

Keywords: Experiment, Sender-Receiver Game, Strategic Information Transmission, Truth-Telling.

JEL-Number: C72, C73, D83.

*Both authors thank Jordi Brandts for many helpful discussions and Raúl López, Rosemarie Nagel, Jordi Massó, Fernanda Rivas and Larry Samuelson for their useful comments. We also thank Marco Faravelli for helping us to conduct the experiment. This research has been possible thanks to the financial support of the Development Research Trust Fund of the University of Edinburgh. Vorsatz acknowledges financial support from the fellowship 2001FI 00451 of the Generalitat de Catalunya and the research grant BEC2002-02130 of the Ministerio de Ciencia y Tecnología de España.

†Address: Edinburgh School of Economics, University of Edinburgh, 50 George Square EH8 9JY, Edinburgh, U.K., E-Mail: ssanchez@staffmail.ed.ac.uk.

‡Corresponding author. Address: Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, Edifici B, 08193 Cerdanyola del Vallès, Spain, E-Mail: mvorsatz@idea.uab.es.

1 Introduction

In several strategic environments individuals have incentives to lie about their private information.[1] But by behaving strategically they disrespect one of the oldest ethical principles, a social norm telling us *not to lie*. This tension between incentives and normative social behavior makes it difficult to predict the outcome of this type of interactions. It is our objective to show, with the help of an experiment, that in situations that can be modelled as a particularly simple *sender-receiver game*, a considerable number of subjects have preferences for truth-telling, whereas the rest of the subjects follow only the material incentives underlying the interaction.

"Strategic information games", introduced by Crawford and Sobel [8], are an obvious way of modelling the tension described above. In this class of games, a player called the "sender" has private information about the true state of the world. Then, she transmits a message about the actual state to the receiver who takes a subsequent action that is payoff-relevant for both participants. The main insight of Crawford and Sobel [8] is that less information about the true state is transmitted as the preferences of the sender and the receiver become less aligned.

In the first experimental study on strategic information games, Dickhaut et. al [9] corroborated this theoretical prediction. But more recently, Cai and Wang [5] offered clear experimental evidence of an *overcommunication phenomenon*: senders truthfully reveal their private information more often than predicted by the most informative equilibrium of the standard model of preference maximization. Although the authors explain the observed ab-

[1] Examples include oligopolistic competition (Galor [15]), financial advice (Morgan and Stocken [19]) and electoral competition (Heidhues and Lagerlof [16]).

normalities successfully by means of a behavioral type analysis (see among others Bosch-Domènech et. al [3], Costa-Gomes et. al [6] and Crawford [7]) or the Logit Quantal Response equilibrium concept (McKelvey and Palfrey [17] and [18]), they leave as an open question whether the overcommunication phenomenon is caused by social preferences such as trust or honesty.

Our aim is to show that the tension between incentives and normative social behavior is the driving force underlying the overcommunication result. To this end, we study the experimental behavior of a group of subjects in two very similar constant-sum sender-receiver games. The *Benchmark Game* proceeds as follows: In the beginning of the game, one out of two payoff tables is randomly picked. The selected table determines players' (strictly positive) payoffs as a function of the action to be taken by the receiver later on. Then, the sender, who is the only player informed about Natures choice, submits a message about the actual payoff table; hence, she implicitly decides to tell the truth or to lie. After observing this message, the receiver either trusts or distrusts the sender. Finally, both participants are paid according to the selected payoff table.

Since the payoff tables are constructed in such a way that the preferences of the sender and the receiver are conflictive (the best case scenario for the receiver is the worst for the sender and vice versa), the sender does not have any incentive to transmit information, or, to say it differently, the sender plays a strategy such that the posterior beliefs of the receiver are equal to the prior beliefs. Given our model specification, all strategies in which the sender lies with probability one-half generate these beliefs consistently and can thus be supported in equilibrium. In the first step of our analysis, we recover the overcommunication phenomenon of Cai and Wang [5]: Subjects playing the Benchmark Game in the role of the sender lie in 44.93% of

our observations, a percentage significantly smaller than the equilibrium prediction of 50% (Hypothesis 1).

To show that this result is caused by a considerable number of subjects with preferences for truth-telling we extend our original set-up by introducing punishments. In the *Punishment Game*, the receiver is informed about the actual payoff table once he has taken an action. Then, he chooses between accepting the payoff distribution induced by the Benchmark Game and reducing the payoffs of both participants to zero.

According to the standard model of preference maximization, individuals care only about their own payoffs, and therefore, the receiver should never punish the sender. But the limitations of the purely rational model are already well-documented. For example, the distributional models of Fehr and Schmidt [13] and Bolton and Ockenfels [2] incorporate experimental evidence showing that some individuals are inequality averse. So they are willing to pay money in order to reduce income disparities. Recently Brandts and Charness [4] have found evidence of even more complex preferences. Individuals not only take into account the whole payoff distribution, because the notion of *procedural justice*[2] - the utility attached to a payoff distribution depends on how this distribution has been reached - plays a crucial role in economic interactions. As a matter of fact, Brandts and Charness [4] study in the laboratory a game in which the sender transmits a message regarding her/his intended play in a 2×2 simultaneous move game and show that the receiver's willingness to punish the sender after revealing the result from the simultaneous move game depends on whether or not the sender played according to the reported message.

[2]The concept of procedural justice was introduced in decision theory by Sen [20] as an extension of the standard model of preference maximization over material outcomes.

We derive our predictions with respect to this game by looking at the punishment rates after different histories. Given that the game is symmetric, histories can be summarized by whether or not a message is truthful and whether or not the receiver trusts that message. The punishment rates are equal to 0% after history (truth, trust), 1.6% after (lie, distrust), 5.4% after (truth, distrust), and 25.2% after (lie, trust). Since the payoff distributions after (truth, distrust) is equal to the one after (lie, trust), we confirm the importance of procedural justice in social interactions (Hypothesis 2).

Finally, we show that the overcommunication phenomenon can be explained in terms of social preferences. First, we identify a number of subjects with concerns for procedural justice. We find that a group of 15 out of 66 individuals punish liars frequently after the history (lie, trust) and that they account for 90% of all punishments. Then, we analyze how these subjects behave as senders and find that they tell the truth in over 70% of the cases whereas the rest of subjects do so in only the 52% (the overall punishment rate is 57%). This result supports our intuition that individuals with a strong sense for procedural justice should, consistently, be responsible for the overcommunication phenomenon (Hypothesis 3).

The remainder of the paper is organized as follows: In the next Section, we formally introduce the games and our experimental hypotheses. In Section 3, we explain the experimental procedures. In the following Section, we present our results. We conclude in Section 5.

2 Theoretic Analysis and Experimental Predictions

In this Section, we introduce the Benchmark and the Punishment games and derive several null hypotheses from their sequential equilibria. Then,

we present alternative hypotheses derived from the overcommunication phenomenon and the incorporation of procedural justice.

The Benchmark Game

Let $N = \{S, R\}$ be the set of players. We call S the "sender" and R the "receiver". At the beginning of the game, Nature selects between payoff tables A and B with equal probability, e.g. $p(A) = p(B) = 0.5$. Only the sender is informed about the payoff table actually chosen. Selecting table $x \in \{A, B\}$ means that the final payoffs are realized according to it and depend only on the action U or D to be taken by the receiver later on.

Table A	Sender	Receiver
Action U	2	1
Action D	1	2

Table B	Sender	Receiver
Action U	1	2
Action D	2	1

Table 1: Payoff Tables

After the sender has been informed, she chooses a mixed strategy with support on the message space $M = \{A, B\}$. Formally, if Nature selects table A, the sender communicates with probability p_A that this table represents the actual payoff scheme. Thus, she lies in this case with probability $1 - p_A$. Similarly, if Nature selects table B, then the sender communicates with probability $1 - p_B$ that this table represents the actual payoff scheme and lies with probability p_B.

Next, we describe the receiver's belief system. If $m = \{A\}$ (the sender transmits message A), then the receiver believes that with probability $\mu(A|A)$ the actual payoff scheme is represented by table A whereas he thinks that

with probability $\mu(B|A) = 1 - \mu(A|A)$ table B is the one determining payoffs. If $m = \{B\}$, then the receiver believes that with probability $\eta(A|B)$ table A determines payoffs and that with probability $\eta(B|B) = 1 - \eta(A|B)$ table B is the one doing so. Taking into account these beliefs, the receiver chooses a mixed strategy with support on the action set $\mathcal{A} = \{U, D\}$. Formally, if $m = \{A\}$, then the receiver takes action U with probability q_A and action D with probability $1 - q_A$. Similarly, if $m = \{B\}$, then the receiver takes action U with probability q_B and action D with probability $1 - q_B$. Finally, both individuals receive their payments. □

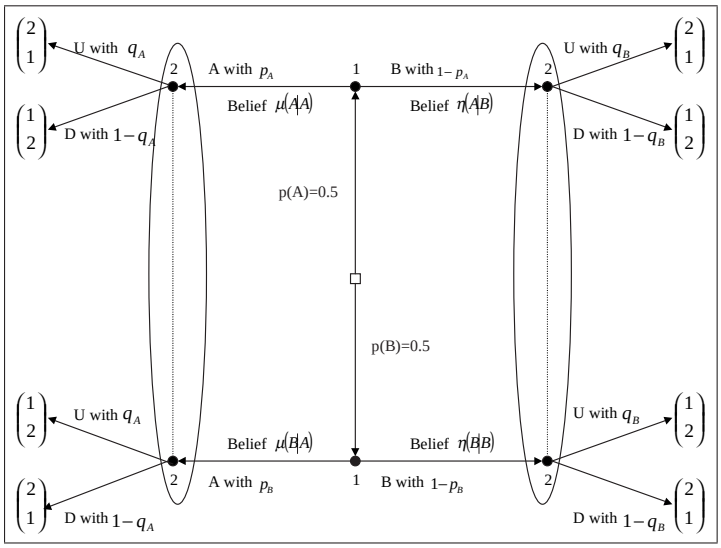


Figure 1: The Benchmark Game

The Benchmark Game is well suited to analyze the tension between social preferences and incentives for two reasons. First, in order to minimize the possibility of mistakes, the Benchmark Game has a simple payoff structure and a very intuitive set of equilibria. Second, truth-telling is a dichotomous choice in the sense that (a) there are only two state variables

and (b) the sender's strategy set boils down to the messages *truth* and *lie*. This is important, because messages can otherwise be interpreted in many different ways. To see this suppose that the state and message spaces are both equal to $\{1, 2, 3, 4, 5\}$. If the sender transmits message $\{3\}$ and the receiver learns afterwards that the true state is $\{4\}$, then some subjects may find this message really close to the true state; after all the sender could have sent message $\{1\}$ instead. But for others the message may be simply a lie worth to be punished. Hence, richer state and message spaces give room to a wide variety of behaviors and to a complexity that lies out of the scope of the paper.[3]

Proposition 1 *The set of sequential equilibria of the Benchmark Game is given by the set of strategies* $(p_A^*, p_B^*, q_A^*, q_B^*) = (p, p, q, q)$*, where* $p, q \in [0, 1]$*, and the supporting belief system* $(\mu^*(A|A), \eta^*(B|B)) = \left(\frac{1}{2}, \frac{1}{2}\right)$*.*

Proof: See Appendix A. ■

The intuition of Proposition 1 is as follows: Since the preferences of the sender and the receiver are not aligned, the sender plays a strategy that leaves the receivers prior beliefs unchanged. The strategies generating these posterior beliefs in a consistent manner are all those in which the sender submits message A with a constant probability, e.g. $p_A = p_B = p \in [0, 1]$. To see this note that if the sender plays for example the strategy "always transmit message A" (this strategy is equal to $p_A = p_B = 1$), then the receiver does not get any additional information from the message. Then, since the sender's message does not convey any information, the receiver can

[3]The importance of the size of the message space is reported by Blume et. al [1]. The authors show that in a sender-receiver game with multiple equilibria it depends on the size of message space whether subjects converge to play a separating or a pooling equilibrium.

as well ignore it. This game becomes thus equivalent to the following one: Nature selects the tables A and B with equal probability and the receiver chooses q (the probability to play U) to maximize his expected payoff that is equal to $p(A)(q+2(1-q))+p(B)(2q+1-q) = 1.5$ and thus independent of q. Therefore, any constant strategy $q_A = q_B = q \in [0,1]$ is optimal.

The set of pooling equilibria is quite large, yet there is an easy way to identify them. Given p_A and p_B, the probability that the sender lies in the Benchmark Game is equal to $l_1(p_A, p_B) = p(A)(1-p_A) + p(B)p_B$. The notion of trust, on the other hand, refers to the beliefs of the receiver. Given p_A and p_B, we denote by $p(m = x)$ the probability that the sender transmits message $x \in \{A, B\}$. Then, the probability that the receiver trusts the sender in the Benchmark Game is equal to $t_1(p_A, p_B) = p(m = A)\mu(A|A) + p(m = B)\eta(B|B)$. According to Corollary 1 the sender lies with probability one-half in any sequential equilibrium of the Benchmark Game, a strategy foreseen correctly by the receiver in terms of trust.

Corollary 1 *Let (p_A^*, p_B^*) be an equilibrium strategy for the sender in the Benchmark Game. Then, $l_1(p_A^*, p_B^*) = t_1(p_A^*, p_B^*) = \frac{1}{2}$.*

Our first null hypothesis is given by Corollary 1. The first part of the corresponding alternative hypothesis is derived from the overcommunication phenomenon of Cai and Wang [5] and states that the sender lies less than predicted by the standard model. If true, the best response function of the receiver dictates that he should always take action D after message A and action U after message B, or, in other words, that the receiver should always take the action *trust*.[4] This leads us to hypothesize additionally that

[4]Following Farrell and Rabin [12], we interpret the revealed decision of the receiver as the result from a maximization process involving subjective beliefs about the truthfulness of the message.

receivers take the action *trust* in more than fifty percent of all occasions.

HYPOTHESIS 1: *In the Benchmark Game, the senders lie in less than fifty percent of the cases and the receivers trust the senders in more than fifty percent of the cases.*

Next, we introduce punishments into the Benchmark Game and find that the source of the overcommunication phenomenon is the preference for truth-telling shared by a considerable number of subjects.

The Punishment Game

The Punishment Game extends the Benchmark Game. Let H be the set of all histories of the Benchmark Game. Given $h \in H$, the receiver reduces with probability $r_2(h) \in [0,1]$ the payoffs of both participants to 0 and accepts with probability $1 - r_2(h)$ the distribution of payoffs induced by the Benchmark Game. Afterwards, both individuals receive their payments. □

Given the strategy (p_A, p_B) of the sender in the Punishment Game, $l_2(p_A, p_B)$ and $t_2(p_A, p_B)$ denote the probabilities that the sender lies and that the receiver trusts the sender's message, respectively. It is easy to calculate the set of sequential equilibria of the Punishment Game because, from a purely materialistic point of view, it is never optimal for the receiver to reduce payoffs. This observation allows us to draw the following conclusion.

Corollary 2 *In all sequential equilibria of the Punishment Game, (a) for all* $h \in H, r_2^*(h) = 0$ *and (b)* $l_2(p_A^*, p_B^*) = t_2(p_A^*, p_B^*) = \frac{1}{2}$.

The second null hypothesis is given by Corollary 2. To derive the corresponding alternative hypothesis first note that the set H can be summarized

by the histories h_1 = (truth, trust), h_2 = (truth, distrust), h_3 = (lie, trust), and h_4 = (lie, distrust).[5] Different punishment rates after these histories will reveal the presence of social preferences.

The distributional models of Fehr and Schmidt [13] and Bolton and Ockenfels [2] take into account that some individuals are inequality averse and care not only about their own payoff but rather about the whole payoff distribution. But according to Brandts and Charness [4] preferences are even more complex, because the utility attached to a particular payoff distribution also depends on how this distribution has been reached. In particular, the authors show that receivers punish the senders more often if a payoff distribution results from a deceptive message. Thus the receiver should punish the sender more frequently after history h_3 = (lie, trust) than after history h_2 = (truth, distrust) although both payoff distributions are identical. Still, we should expect the punishment rate after history h_2 = (truth, distrust) to be strictly positive, because some individuals may be inequality averse and prefer the payoff distribution (0,0) over (2,1). We do not expect any punishments after the histories h_1 = (truth, trust) and h_4 = (lie, distrust), because the receiver interpreted the message correctly and the resulting payoff distribution (1,2) is favorable to him. If we reject the null hypothesis in favor of the alternative one, then a sender who lies is in more danger of being punished than a truth-teller. As a consequence, we hypothesize that truth-telling is enhanced in the Punishment Game with respect to the Benchmark Game and that the receivers trust more in the former than in the latter.

[5] In our experimental sessions we do not ask subjects to elicit their mixed strategies, rather we derive them from the repeated observation of pure strategies. Then, the fact that the payoff tables A and B are symmetric and the probabilities $p(A)$ and $p(B)$ are identical allows us to write the set H in the way described above.

HYPOTHESIS 2: *In the Punishment Game, the receivers punish the senders only after the histories* $h_2 = (truth, distrust)$ *and* $h_3 = (lie, trust)$ *with the punishment rate being higher after* h_3*. Moreover, the senders lie less and the receivers trust more in the Punishment than in the Benchmark Game.*

According to our main hypothesis the overcommunication phenomenon can be explained because a considerable group of individuals has preferences for truth-telling. To check this we perform a final consistency test on the Punishment Game. After observing the experimental results, we divide our subject pool into two different groups, one group of subjects punishing liars frequently after history h_3 = (lie,trust) and another group containing the rest of the subjects. Given this division, the third null hypothesis states that there is no difference in the level of truth-telling among the two groups when subjects play the Punishment Game as senders.[6] On the other hand, the corresponding alternative hypothesis states that the group of subjects with a high sense of procedural justice accounts for most of the overcommunication phenomenon; that is, these subjects tell the truth very often whereas the rest of the subjects lie in about fifty percent of the occasions.

HYPOTHESIS 3: *In the Punishment Game, the group of subjects punishing liars very often after history* h_3 = (*lie,trust*) *accounts for most of the overcommunication phenomenon.*

[6]We use a role rotation mechanism in our experimental sessions so that every subject plays the Punishment Game half of the time in each role. For more on this see Section 3.

3 Experimental Design and Procedures

We conducted our experimental sessions at the University of Edinburgh in May 2004. Since all economic students at this university have an E-mail account associated to their matriculation number, we promoted the experiment mainly via electronic newsletters. Students from other academic disciplines were recruited through flyers distributed on the campus and further announcements made on information boards. As a result, 132 undergraduate students from nearly all faculties participated in one of our experimental sessions. We organized a total of ten sessions, five on the Benchmark Game and five on the Punishment Game. Twelve subjects participated in the first four sessions and eighteen subjects in the fifth and last session of each treatment. No subject took part in more than one session.

To perform this experiment we employed the computer software Z-Tree developed by Fischbacher [14]. At the beginning of a session subjects met in a computer room and sat down in front of one of the computers. The computers were placed in such a way all subjects could only see her/his own screen. We placed next to each computer a closed envelope containing instructions, a questionnaire, and a payment receipt. After subjects had filled out the questionnaire we read the instructions aloud (see Appendix B for the instructions corresponding to the Punishment Game).

The computer randomly divided the subjects participating in the same session into groups of six without revealing the matching. We informed every subject that s/he would only play against subjects belonging to the same group. Therefore, the fact that the number of subjects differed across sessions should not matter. Hence, we implicitly divided our subject pool into a total of twenty-two groups of six subjects, eleven groups playing each

treatment. In each of the fifty rounds of an experimental session the computer matched the subjects belonging to the same group into three new pairs and assigned different roles (sender or receiver) within pairs. The matchings were constructed in such a way that after fifty rounds every subject played the game exactly ten times against each of her/his five opponents. Moreover, every subject met every opponent five times in each role.

In every round, after pairs had been formed and roles had been assigned, the sender was informed of whether table A or B represented the actual payoff scheme. Afterwards, the sender transmitted a message from the message space $M = \{A, B\}$ telling the receiver which table corresponds to the actual payoff scheme. Then the receiver chose an action from the action set $\mathcal{A} = \{U, D\}$. This constituted the end of the round in the Benchmark Game. In the sessions corresponding to the Punishment Game, the receiver was further informed about the induced payoffs of her/his action. Finally, s/he had to decide between accepting these payoffs or reducing the payoff of both participants to zero.

At the end of a session, we called subjects one by one to step forward to the control desk for payment. In addition to the five pounds show up fee, subjects received ten pence per point obtained. As a result, the average payment in the one hour session corresponding to the Benchmark and the Punishment Game was equal to 12.5 pounds and 11.74 pounds, respectively.

4 Results

4.1 Overcommunication in the Benchmark Game

According to our first null hypothesis, the senders should lie in the Benchmark Game with probability one-half. In the histogram in the left part of

Figure 2, we represent the frequencies of truthful messages in the sessions corresponding to the Benchmark Game. Since a subject was exactly 25 times in the role of the sender, in equilibrium s/he should tell the truth 12.5 times. The data seem to be slightly shifted to the right of the theoretical mean, although there is no clear evidence of a statistically significant difference. In the right panel of Figure 2, it is possible to observe that the percentage of subjects telling the truth is extraordinary high in the first rounds and declines over time in such a way that it stays on average just above of the 50% line predicted by the standard theory. We eliminate this learning effect by excluding in our statistical analysis the data from the first ten rounds .

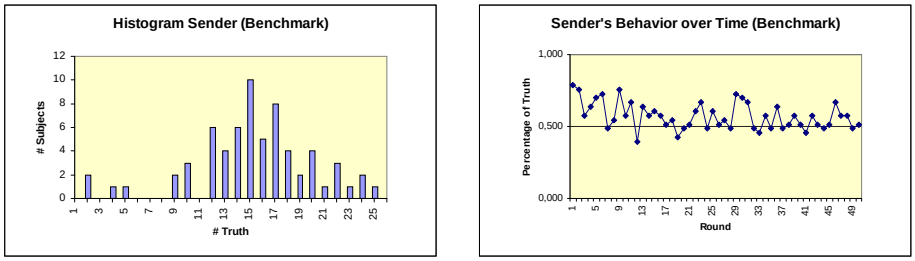


Figure 2: Sender's Behavior in the Benchmark Game

Since subjects belonging to the same group play the Benchmark Game more than once against the other group members, actions within a given group are likely to be correlated over time. This is the reason why we calculate for every group the percentage of truthful messages over the last forty rounds. This procedure allows us to obtain eleven independent observations, one observation for each group. The overall percentage of truth-telling in the last forty rounds is equal to 0.5507, a percentage significantly greater than 0.5 (p-value of the one-tailed Wilcoxon rank-sum test = 0.0615; p-value of the one-tailed t-test = 0.0459).

Next, we provide evidence in favor of the second part of Hypothesis 1, namely that the receivers adjust their beliefs in the correct direction and trust the senders in more than fifty percent of all occasions. To this end, we interpret the action of the receiver as the result of a maximization process involving subjective beliefs about the truthfulness of the message. For example, if a subject observes message A and takes action D afterwards, then this action reveals in our understanding that the subject trusts the sender's message. In the histogram in the left panel of Figure 3, we can clearly see that a lot of receivers trust more often than the theoretical prediction of 12.5 times. If we analyze the evolution of this percentage over time, one can observe that it is particularly low in the first rounds before it stabilizes well above the fifty percent line.

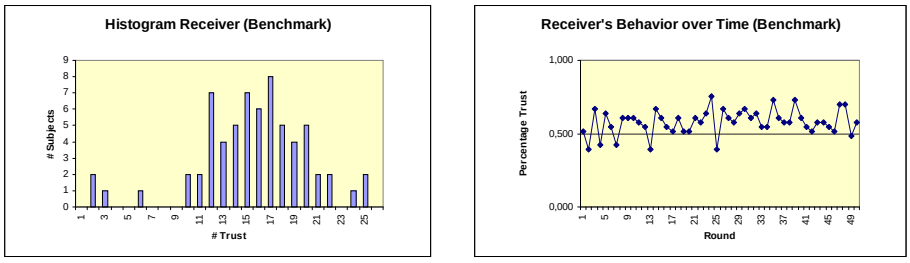


Figure 3: Receiver's Behavior in the Benchmark Game

In the last forty rounds of the experiment the receivers trust the senders' messages in the 58.7% of all observations. This value is significantly greater than the theoretical prediction of 50% (p-value of the one-tailed Wilcoxon rank-sum test < 0.0001; p-value of the one-tailed t-test < 0.0001). Hence, we reject the prediction of the standard model in favor of Hypothesis 1.

4.2 Procedural Justice in the Punishment Game

So far we have shown the existence of an overcommunication phenomenon in the Benchmark Game. We are now ready to analyze the origin of this result. In Figure 4 below we present the punishment behavior of the receivers. For consistency reasons we only consider punishments in the last forty rounds, and therefore, we have a total of 1320 observations (11 groups of 40 rounds and 3 observations per round). The senders told 740 times the truth and lied in 580 occasions. The receivers trusted the message 520 times when the sender had told the truth and 396 times when the sender had lied.

	Truth	Lie
Trust	0	0,25253
Distrust	0,05455	0,0163

Table 2: Punishment Behavior

The punishment rate is highest, more than 25%, after history h_3 = (lie,trust). We use the normal approximation of the binomial distribution in order to establish that this proportion is significantly greater than zero (p-value of the one-tailed Z-test < 0.0001). We also find, as expected, that the punishment rate after history h_2 = (truth, distrust) is significantly greater than zero. We attribute the positive punishment rate after history h_4 = (lie, distrust) to mistakes made by some subjects. Yet, our main prediction is confirmed: The willingness to punish the sender depends on whether or not a payoff distribution has been reached by means of a deceptive message, because the punishment rate after history h_3 = (lie, trust) is greater than the

one after the history h_2 = (truth, distrust). The test of equal proportions sustains this hypothesis (p-value of the one-tailed Z-test < 0.0001).

We analyze next whether subjects behave consistently across the two treatments. The histogram in left panel of Figure 4 looks quite similar to the one corresponding to sender's behavior in the Benchmark Game although it seems that the shift to the right from the theoretical mean has increased. In the right panel of Figure 4, we observe that the percentage of subjects telling the truth is quite high in the first rounds and declines over time, a behavior we have encountered before. Nevertheless, in the latter rounds there are less values below the fifty percent line.

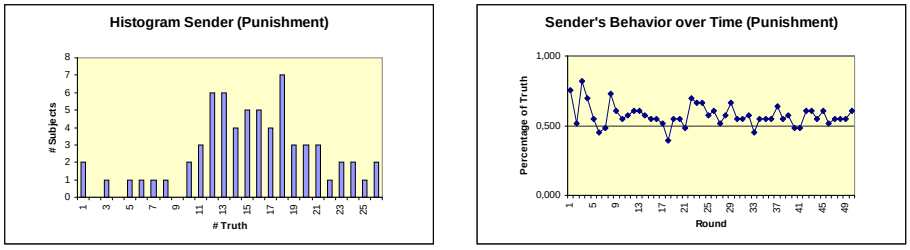


Figure 4: Sender's Behavior in the Punishment Game

The percentage of subjects telling the truth in the last forty rounds is equal to 56.29%. This percentage is significantly greater than 50% (p-value of the one-tailed Wilcoxon rank-sum test = 0.0499; p-value of the one-tailed t-test = 0.0343), but it is not significantly greater than the corresponding percentage in the Benchmark Game of 55.07% (p-value of the one-tailed Wilcoxon rank-sum test = 0.40; p-value of the one-tailed t-test = 0.385).

The picture looks quite differently if we compare the receivers' behavior across the two treatments. The histogram in the left panel of Figure 5 indicates that the receivers trust more in the Punishment than in the

Benchmark Game. This intuition is confirmed in the right panel of Figure 5, because the percentage of receivers trusting the sender seems to increase over time and stays well above the fifty percent line. On the aggregate, in the last forty rounds the percentage of trustful receivers is equal to 69.3%. This percentage is significantly greater than the one corresponding to the Benchmark Game of 58.7% (p-value of the one-tailed Wilcoxon rank-sum test < 0.0001; p-value of the one-tailed t-test < 0.0001).

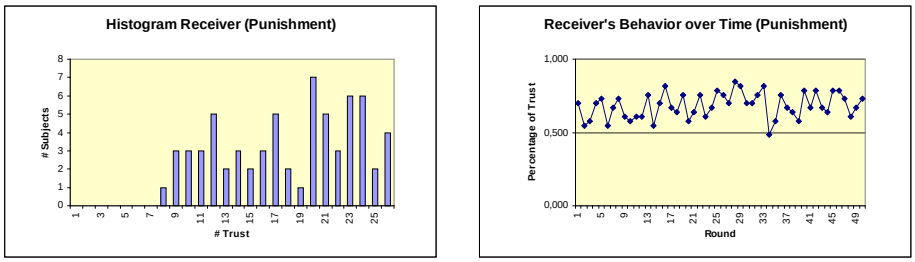


Figure 5: Receiver's Behavior in the Punishment Game

The introduction of punishments seems to induce receivers to believe that senders will often tell the truth in order to avoid a possible moral outrage caused by deceptive messages. But on the contrary, when subjects play as senders they seem to consider the punishment as an incredible threat and they barely change their behavior with respect to the Benchmark Game.

4.3 The Separability Hypothesis

In the previous two parts of this Section we have prepared the ground for our main contribution, namely that that the tension between normative social behavior and incentives is driving the overcommunication phenomenon. After observing the experimental results, we first divide our subject pool into two groups, one group of subjects containing those who punish liars

frequently and another group containing the rest of subjects. We obtain this division in the following way: In the last forty rounds of an experimental session corresponding of the Punishment Game every subject is twenty times in the role of the receiver. Since the sender lies with probability 0.437 and the receiver trusts the message with probability 0.694, every subject plays, in expected terms, the history h_3 = (lie, trust) 6.06 times in the role of the receiver. The punishment rate after h_3 is equal to 0.2525, and therefore, every subject is expected to punish the sender 1.53 times. Hence, we consider that all subjects punishing the sender in at least three occasions after h_3 have serious concerns for procedural justice. This condition is met by fifteen out of sixty-six subjects. Not surprisingly, this group of subjects accounts for 90% of all punishments.

Given this classification, the role rotation mechanism allows us to study how these fifteen subjects behave in the Punishment Game. On the aggregate, they tell the truth in 70.66% of all observations. This probability is significantly greater than 56.29%, the percentage of truth-telling corresponding to the whole subject pool (p-value of the one-tailed Z-test < 0.0001). The rest of the subjects, on the other hand, tell the truth in only 52.05% of the cases, a percentage not significantly greater than the theoretical prediction of 50% (p-value of the one-tailed Z-test ≥ 0.0945). Therefore, we reject the third null hypothesis -the percentage is the same for both groups of subjects -in favor of Hypothesis 3. This result suggests not only that individuals with a strong notion of procedural justice behave consistently across roles, rather they also show preferences for truth-telling and account for most of the information transmitted by the senders. This interpretation is further strengthened if we analyze how the beliefs of these two groups vary. The subjects with a serious notion for procedural justice trust in 86%

of the occasions, whereas the rest of subjects do so only in 64,51%.

5 Conclusion

Communication is the most natural way to exchange information. Experimental studies such as the ones of Duffy and Feltovich [10] and [11] have shown that individuals are able to achieve Pareto improving allocations by means of cheap talk. In particular, the authors show that if subjects announce to cooperate in the Prisoner's Dilemma, then this message often reflects the truth. Moreover, receivers reciprocate and cooperate as well so the Pareto-efficient outcome is frequently implemented. On the other hand, Crawford [7] shows that in some sender-receiver games a rational individual can feint a boundedly rational one by sending a deceptive message .

These two results raise some questions. When can the receiver trust the senders' messages? And why do the senders transmit truthful messages? Our aim was to show that the overcommunication phenomenon is not necessarily due to a lack of sophistication or rationality but to the fact that individuals follow normative social behaviors such as truth-telling. To this end, we studied the behavior of a group of subjects in a simple sender-receiver game offering no material incentives to do so.

In the first step, we recovered the overcommunication phenomenon of Cai and Wang [5] (e.g. on the aggregate the senders transmit information more than predicted by the standard model of preference maximization). Then, we introduced punishments and showed that, in accordance with the results of Brandts and Charness [4], the willingness to punish the sender is higher after a deceptive message. Finally, we sustained our main hypothesis by showing that if we subtract from our subject pool the group of subjects

who often punish liars after a deceptive message, then that very same group tells the truth very often whereas the rest of the subjects behave roughly according to the standard equilibrium prediction. Thus, if moral subjects are excluded, then the overcommunication phenomenon vanishes.

The existence of moral agents who reject material incentives to misbehave opens some fascinating questions: What are the implications on mechanism design? Or on the organization of the firm? And on the elaboration of policy prescriptions?

References

[1] A. Blume, D. DeJong, Y. Kim, and G. Sprinkle, *Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games*, American Economic Review **88** (1998), 1323–1339.

[2] G. Bolton and A. Ockenfels, *ERC: A Theory of Equity, Reciprocity, and Competition*, American Economic Review **90** (2000), 166–193.

[3] A. Bosch-Domènech, R. Nagel, J. García-Montalvo, and A. Satorra, *One, Two, (Three), ..., Infinity: Newspaper and Lab Beauty-Contest Experiments*, American Economic Review **92** (2002), 1687–1701.

[4] J. Brandts and G. Charness, *Truth or Consequence: An Experiment*, Management Science **49** (2003), 116–130.

[5] H. Cai and J. Wang, *Overcommunication and Bounded Rationality in Strategic Information Transmission Games: An Experimental Investigation*, Mimeo (2003).

[6] M. Costa-Gomes, V. Crawford, and B. Broseta, *Cognition and Behavior in Normal Form Games: An Experimental Study*, Econometrica **69** (2001), 1193–1235.

[7] V. Crawford, *Lying for Strategic Advantages: Rational and Boundedly Rational Misrepresentations of Intentions*, American Economic Review **93** (2003), 133–149.

[8] V. Crawford and J. Sobel, *Strategic Information Transmission*, Econometrica **50** (1982), 1431–1451.

[9] J. Dickhaut, K. McCabe, and A. Mukherji, *An Experimental Study of Strategic Information Transmission*, Economic Theory **6** (1995), 389–403.

[10] J. Duffy and N. Feltovich, *Do Actions Speak Louder Than Words? Observation vs. Cheap Talk as Coordination Devices*, Games and Economic Behavior **39** (2002), 1–27.

[11] _____, *Words, Deeds and Lies: Strategic Behavior in Games with Multiple Signals*, Mimeo (2003).

[12] J. Farrell and M. Rabin, *Cheap Talk*, Journal of Economic Perspectives **10** (1996), 103–118.

[13] E. Fehr and K. Schmidt, *A Theory of Fairness, Competition, and Cooperation*, Quaterly Journal of Economics **114** (1999), 817–864.

[14] U. Fischbacher, *Z-Tree Tutorial Version 2.1*, Zurich University (2002).

[15] E. Galor, *Information Sharing in Oligopoly*, Econometrica **53** (1985), 329–343.

[16] P. Heidhues and J. Lagerlof, *Hiding Information in Electoral Competition*, Games and Economic Behavior **42** (2003), 48–74.

[17] R. McKelvey and T. Palfrey, *Quantal Response Equilibria in Extensive Form Games*, Experimental Economics **1** (1995), 9–41.

[18] ———, *Quantal Response Equilibria in Normal Form Games*, Games and Economic Behavior **10** (1998), 6–38.

[19] J. Morgan and P.C. Stocken, *An Analysis of Stock Recommendations*, RAND Journal of Economics **34** (2003), 183–203.

[20] A. Sen, *Maximization and the Act of Choice*, Econometrica **65** (1997), 745–779.

Appendix A: Proof of Proposition 1

Remember that p_A denotes the probability that the sender submits message A when the signal is table A and that p_B is equal to the probability that the sender submits message A when the signal is table B. We have to to divide our analysis into three different cases.

Case 1: Suppose that $0 < p_A^* + p_B^* < 2$. We derive the best response correspondence for the receiver who takes the strategies p_A and p_B as given. Suppose that the sender transmits message A. By sequential rationality the receiver updates his beliefs according to Bayes' rule, and therefore, he thinks that the probability $\mu(\text{payoff } A|\text{ message } A)$ (e.g. the true payoff scheme is represented by table A conditional on message A) is equal to

$$\begin{aligned}\mu(\text{payoff } A|\text{ message } A) &= \frac{p(\text{message } A|\text{ payoff} A)p(\text{payoff } A)}{p(\text{message } A)} \\ &= \frac{0.5p_A}{0.5(p_A+p_B)} = \frac{p_A}{p_A+p_B} \equiv \mu.\end{aligned}$$

Similarly, let $\mu(\text{payoff } B|\text{ message } A) = 1-\mu$ be the belief that table B represents the actual payoffs when the sender submitted message A before. Given μ, the receiver chooses q_A (the probability to take action U conditional on message A) in order to

$$\max_{q_A} \quad (\mu(q_A + 2(1-q_A)) + (1-\mu)(2q_A + 1 - q_A)).$$

This maximization problem is equivalent to

$$\max_{q_A} \quad (1 + \mu + q_A(1-2\mu)),$$

and therefore, the best response correspondence for the receiver is

$$q_A^*(\mu) = \begin{cases} 1 & \text{if } \mu < \frac{1}{2} \\ [0,1] & \text{if } \mu = \frac{1}{2} \\ 0 & \text{if } \mu > \frac{1}{2}, \end{cases} \quad \text{, or,} \quad q_A^*(p_A, p_B) = \begin{cases} 1 & \text{if } p_A < p_B \\ [0,1] & \text{if } p_A = p_B \\ 0 & \text{if } p_A > p_B. \end{cases}$$

If, on the other hand, the sender submits message B, then the belief that the actual payoff scheme is represented by table B is equal to

$$\begin{aligned} \eta(\text{payoff } B|\text{ message } B) &= \frac{p(\text{message } B|\text{ payoff } B)p(\text{payoff } B)}{p(\text{message } B)} \\ &= \frac{0.5(1-p_B)}{0.5(1-p_A)+0.5(1-p_B)} = \frac{1-p_B}{2-p_A-p_B} \equiv \eta. \end{aligned}$$

Similarly, let $\eta(\text{payoff } A|\text{ message } B) = 1-\eta$ be the belief that table A represents the payoff scheme when the sender submitted message B before. Given η, the receiver chooses q_B (the probability to take action U conditional on message B) in order to

$$\max_{q_B} \quad ((1-\eta)(q_B + 2(1-q_B)) + \eta(2q_B + 1 - q_B)).$$

This maximization problem is equivalent to

$$\max_{q_B} \quad (2 - \eta + q_B(2\eta - 1)),$$

and therefore, the best response correspondence of the receiver is

$$q_B^*(\eta) = \begin{cases} 1 & \text{if } \eta > \frac{1}{2} \\ [0,1] & \text{if } \eta = \frac{1}{2} \\ 0 & \text{if } \eta < \frac{1}{2}, \end{cases} \quad \text{, or,} \quad q_B^*(p_A, p_B) = \begin{cases} 1 & \text{if } p_B < p_A \\ [0,1] & \text{if } p_A = p_B \\ 0 & \text{if } p_B > p_A. \end{cases}$$

Next, we calculate the receivers optimal mixed strategy p_A^* (the probability that the sender submits message A if the payoff scheme is represented by table A) and p_B^* (the probability that the sender submits message A if the payoff scheme is represented by table B). To do so we consider three different cases:

Case A: Suppose that $p_A^* < p_B^*$. Then, it follows from the optimal behavior of the receiver that $q_A^*(p_A^*, p_B^*) = 1$ and $q_B^*(p_A^*, p_B^*) = 0$, and therefore, the optimal strategies p_A^* and p_B^* must be the solution to the following maximization problem: Choose p_A and p_B in order to

$$\max_{p_A, p_B} \quad 0.5\left(2p_A + p_B + 1 - p_A + 2\left(1 - p_B\right)\right).$$

This maximization problem is equivalent to

$$\max_{p_A, p_B} \quad 0.5\left(3 + p_A - p_B\right).$$

But the solution to this problem is such that $p_A^* = 1$ and $p_B^* = 0$, and therefore, we have reached a contradiction. We conclude that there does not exist any equilibrium in which $p_A^* < p_B^*$.

Case B: Suppose that $p_A^* > p_B^*$. Then, it follows from the optimal behavior of the receiver that $q_A^*(p_A^*, p_B^*) = 0$ and $q_B^*(p_A^*, p_B^*) = 1$, and therefore, the optimal strategies p_A^* and p_B^* must be the solution of the following maximization problem: Choose p_A and p_B in order to

$$\max_{p_A, p_B} \quad 0.5\left(p_A + 2\left(1 - p_A\right) + 2p_B + 1 - p_B\right).$$

This maximization problem is equivalent to

$$\max_{p_A,p_B} \quad 0.5\left(3-p_A+p_B\right).$$

But the solution to this problem is such that $p_A^*=0$ and $p_B^*=1$, and therefore, we have reached a contradiction. We conclude that there does not exist any equilibrium in which $p_A^*>p_B^*$.

Case C: Suppose that $p_A^*=p_B^*$. Then, it follows from the best response correspondences of the receiver that $q_A^*\in[0,1]$ and $q_B^*\in[0,1]$. Thus, the sender faces the problem

$$\begin{aligned}\max_{p_A,p_B} \quad & 0.5p_A\left(2q_A+1-q_A\right)+0.5\left(1-p_A\right)\left(2q_B+1-q_B\right)+\\ & 0.5p_B\left(q_A+2\left(1-q_A\right)\right)+0.5\left(1-p_B\right)\left(q_B+2\left(1-q_B\right)\right),\end{aligned}$$

that is equivalent to

$$\max_{p_A,p_B} \quad 0.5\left(3+p_A\left(q_A-q_B\right)+p_B\left(q_B-q_A\right)\right).$$

Hence, the best response correspondences for the sender are

$$p_A^*\left(q_A,q_B\right)=\begin{cases}1 & \text{if } q_B<q_A\\ [0,1] & \text{if } q_A=q_B\\ 0 & \text{if } q_B>q_A\end{cases} \quad \text{and} \quad p_B^*\left(q_A,q_B\right)=\begin{cases}1 & \text{if } q_B>q_A\\ [0,1] & \text{if } q_A=q_B\\ 0 & \text{if } q_B<q_A.\end{cases}$$

From inspection we see that the set of mixed strategies $(p_A^*,p_B^*;q_A^*,q_B^*)=(p,p;q,q)$, where $p\in(0,1)$ and $q\in[0,1]$, can be sustained as equilibrium strategies. Finally, plug the optimal strategies into the definitions of the beliefs in order to obtain that $\mu^*(A|A)=\eta^*(B|B)=\frac{1}{2}$.

Case 2: Suppose that $p_A^*=p_B^*=0$. Observe from the best correspondence of the sender in case 1.C that $p_A^*=p_B^*=0$ can only be sustained as an equilibrium strategy if $q_A^*=q_B^*$. Moreover, $p_A^*=p_B^*=0$ implies

that $\eta^*(B|B) = \frac{1}{2}$ and $q_B^*(\eta^*(B|B)) \in [0,1]$. Since the sequential game we study consists of two-periods and the cardinality of the action space of both players is equal to two, any belief $\mu \in [0,1]$ is consistent. Yet, we yield from the best response correspondence $q_A^*(\mu)$ in case 1 that $q_A^* = q_B^*$ if and only if $\mu = \frac{1}{2}$. Therefore, we conclude that the set of mixed strategies $(p_A^*, p_B^*; q_A^*, q_B^*) = (0, 0; q, q)$, where $q \in [0,1]$, together with the belief system $\mu^*(A|A) = \eta^*(B|B) = \frac{1}{2}$ constitutes a set of sequential equilibria.

Case 3: Suppose that $p_A^* = p_B^* = 1$. Observe from the best correspondence of the sender in case 1.C that $p_A^* = p_B^* = 1$ can only be sustained as an equilibrium strategy if $q_A^* = q_B^*$. Moreover, $p_A^* = p_B^* = 1$ implies that $\mu^*(A|A) = \frac{1}{2}$ and $q_A^*(\mu^*(A|A)) \in [0,1]$. Although any belief $\eta \in [0,1]$ is consistent, we yield from the best response correspondence $q_B^*(\eta)$ in case 1 that $q_A^* = q_B^*$ if and only if $\eta = \frac{1}{2}$. Therefore, we conclude that the set of mixed strategies $(p_A^*, p_B^*; q_A^*, q_B^*) = (1, 1; q, q)$, where $q \in [0,1]$, together with the belief system $\mu^*(A|A) = \eta^*(B|B) = \frac{1}{2}$ constitutes a set of sequential equilibria. ■

Appendix B: Instructions of the Punishment Game

Welcome

Thank you for coming. The purpose of this session is to study how people make decisions in a particular situation. If you have any questions, feel free to raise your hand and your question will be answered so everyone can hear. From now until the end of the session unauthorized communication of any nature with any other participant is prohibited. The experiment will be conducted through computers and all interactions between you will take place through them.

During the session you will play a game that gives you the opportunity to make money. What you earn depends partly on your decisions and partly on the decisions of others. At the end of the session, the amount you earned will be paid to you privately in cash.

We start with a brief instruction period. During the instruction period you will be given a description of the experiment. We are about to begin.

General Instructions

In your envelope you will find a questionnaire and an official receipt. Fill in the questionnaire and write down your name and matriculation number in the receipt. You will need both forms to receive your payment at the end of the session. Your personal data will be kept confidential and will be used for statistical purposes only.

In this session, you will play a game which is repeated for 50 rounds. Before the first round, the computer will randomly divide the participants into groups of six. This division will last for the entire session. Participants within each group will play only among themselves. The assignment process is random and anonymous so you will not know who is in your group.

Next, we will go over a brief tutorial. Please interrupt at any time if you have a question.

At the beginning of each round, you will be randomly joined with another participant from your group to form a pair. In each pair, one participant is randomly chosen to be the **Sender**, and one to be the **Receiver**. Remember that this process is random and the assignment changes every round.

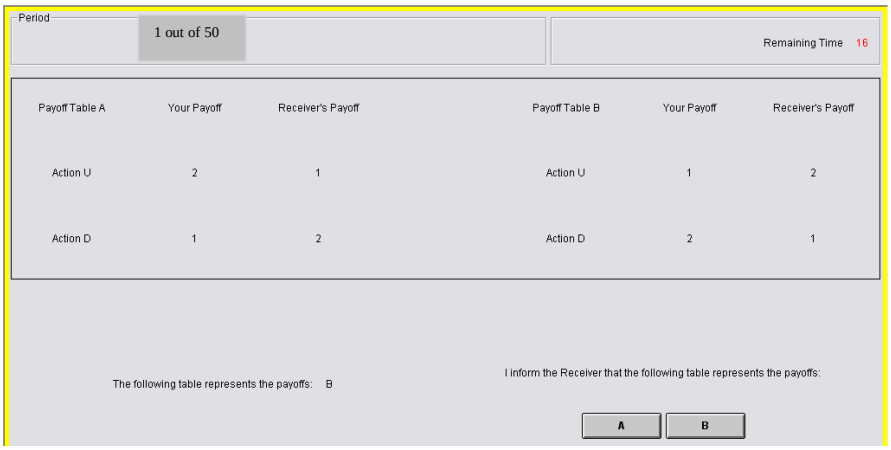
Each round, after pairs have been formed and roles have been assigned, the computer selects one of the following two payoff tables. Final payoffs for both participants will be determined according to the selected table and the action **U** or **D** taken by the Receiver later on.

Table A	Sender	Receiver
Action U	2 Points	1 Point
Action D	1 Point	2 Points

Table B	Sender	Receiver
Action U	1 Point	2 Points
Action D	2 Points	1 Point

Sender's Instructions

At the beginning of the round **only** the Sender will be informed about the actual payoff table chosen by the computer. The Sender is the first one to take a decision in the game. S/He must communicate to the Receiver whether the payoff table chosen by the computer is either table **A** or table **B**. **Please, take into account that the Sender is free to tell the truth or to lie**. The computer screen for the Sender is as follows:

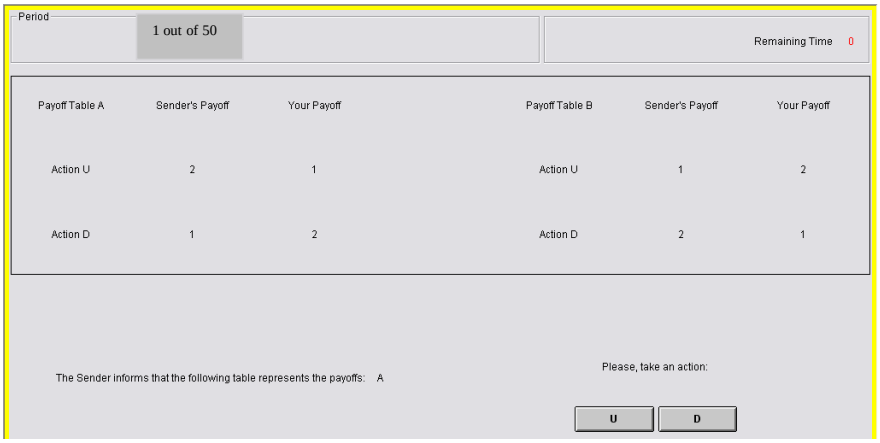


The two tables at the top of the screen represent payoffs according to tables **A** and **B**. Below you find the information whether table **A** or table **B** was chosen by the computer (in our example it is table **B**). On the inferior right corner there are two buttons labelled **A** and **B**. By clicking on the buttons **A** or **B** you inform the Receiver that you have observed the corresponding table. The Sender has 20 seconds to take this decision. **This is the only decision the Sender takes.**

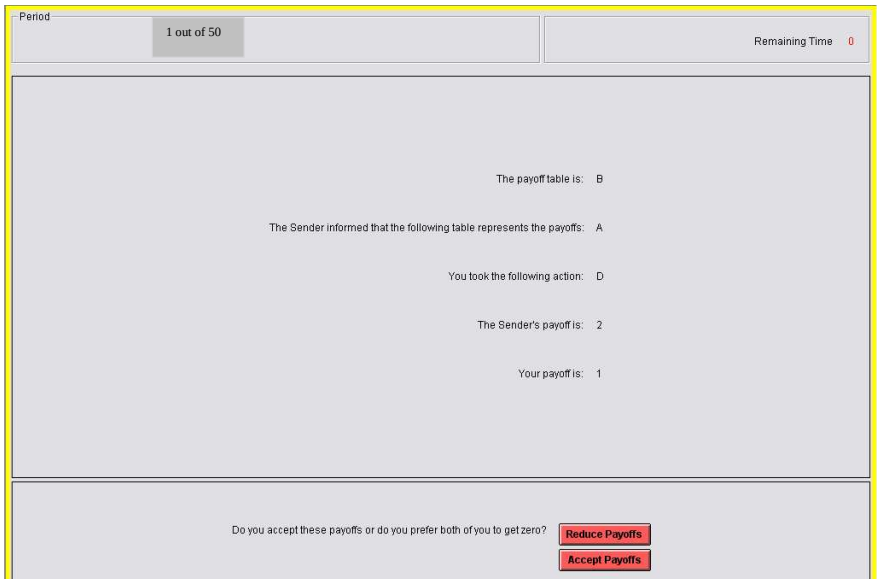
Receiver's Instructions

The Receiver takes two decisions. First, once the Receiver got the Sender's message, s/he has to decide between actions **U** and **D**. The computer screen for the Receiver is as follows:

The two tables at the top of the screen represent payoffs according to tables **A** and **B**. Below you find the message from the Sender regarding the table



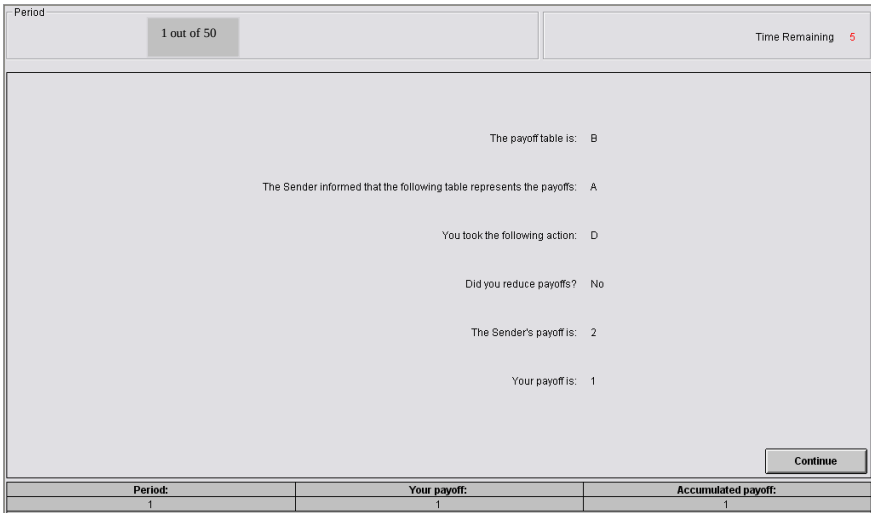
s/he observed (in our example the Sender has informed the Receiver that s/he observed table **A**). On the inferior right corner there are two buttons labelled **U** and **D**. By clicking on the buttons **U** or **D** you take the corresponding action. The Receiver has 20 seconds to take this decision. Once this action is taken, a new screen appears summarizing the outcome of the round so far.



Now the Receiver is asked to take the second decision: S/He must either accept the current payoff distribution or reduce the payoff of both participants to zero. By clicking on the button **Reduce Payoffs** or **Accept Payoffs**, the Receiver takes the corresponding action. The Receiver has 15 seconds to take this decision.

Summary of the Round

The final screen is a summary of the round: It indicates the actual payoff table, the message chosen by the Sender, the actions taken by the Receiver, and the earnings of both participants in this round. Additionally, you are also informed about your accumulated payoff.



The screen above is the Receiver's summary. It indicates that the Sender chose message **A** whereas the Receiver took action **D** and accepted the payoffs. Therefore, the Sender gets 2 Points and the Receiver 1 Point. At the end of a round, click on **Continue**. The experiment will nevertheless proceed automatically to the next round in 10 seconds.

Payment

The Points you accumulate during the course of the session will determine your payment in addition to the £**5** show-up fee. The exchange rate Points/£ is **10p per Point**. At the end of the experiment, take your questionnaire and receipt to the counter for payment. They will be matched to our computer printout. Once you are paid, you may leave.