

IFN Working Paper No. 883, 2011

Credible Communication and Cooperation: Experimental Evidence from Multi-stage Games

Ola Andersson and Erik Wengström

Credible Communication and Cooperation: Experimental Evidence from Multi-stage Games *

Ola Andersson[†] and Erik Wengström[‡]

September 19, 2011

Abstract. It is well known that communication often serves as a facilitator for cooperation in static games. Yet, communication can serve entirely different purposes in dynamic settings as communication during the game may work as a means for renegotiation, potentially undermining the credibility of cooperative strategies. To explore this issue, this paper experimentally investigates cooperation and non-binding communication in a two-stage game. More specifically, two treatments are considered: one with only pre-play communication and one where subjects can also communicate intra-play between the stages of the game. The results highlight a nontrivial difference concerning the effects of pre-play communication between the two treatments. Pre-play communication only has a significant impact on cooperation when no intra-play communication is possible. The results suggest that the credibility of pre-play messages may depend crucially on future communication opportunities.

JEL: C92; C72

Keywords: Communication, Cooperation, Renegotiation, Experiments

*We are thankful for helpful comments made by Antonio Cabrales, Hans Carlsson, Tore Ellingsen, Hakan J. Holm, Dan Houser, Magnus Johannesson, Rosemarie Nagel, Randolph Sloof, participants at the ESA European Meeting in Nottingham 2006, the ESA North-American Meeting 2006 in Tucson, the First Nordic Workshop in Behavioral and Experimental Economics in Oslo 2006 and seminar participants at Lund University, Stockholm School of Economics, University of Padua and Universitat Pompeu Fabra. We are also thankful to Kerstin Enflo and Fredrik Hansen for help in conducting the experiment. Financial support from the Swedish Competition Authority, Riksbankens Jubileumsfond and the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

[†]Research Institute of Industrial Economics (IFN), P.O. Box 55665, SE-102 15 Stockholm Sweden. Tel: +46 (0)8-665 45 21, Fax: +46 (0)8 665 45 99, email: ola.andersson@ifn.se

[‡]Department of Economics, Lund University and Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK-1353, Copenhagen K, Denmark; email: erik.wengstrom@econ.ku.dk

1 Introduction

Communication is often regarded as an important means for achieving cooperation. Yet, typically, it does not directly affect the payoffs of the game. From a theoretical perspective, this raises the question of when such cheap talk messages affect outcomes. Farrell and Rabin (1996) argue that a necessary condition for a message to be credible is that it is self committing, in the sense that a player must have an incentive to comply with the message if it is believed.¹ That is, the proposed actions have to be part of a Nash equilibrium profile. When the underlying game is a multi-stage game, a natural extension is to require that the proposed actions have to be part of a subgame perfect equilibrium (SPE) strategy.

In a multi-stage game, communication can take place not only before the game (pre-play) but also between the stages of the game (intra-play). Hence, when evaluating the credibility of a given message, the players have to take into account the fact that intentions may be revised in subsequent communication rounds. To give an illustration, note that players in multi-stage games can propose cooperation by threatening to punish non-cooperative behavior in the future. Such messages may be deemed credible if they are part of an SPE. Yet, if players can communicate intra-play, it is not obvious that threats of punishment will be carried out after a defection, as punishments usually hurt both players. By communicating again, they may be able to renegotiate and coordinate on an outcome that is preferred by both, instead of carrying out the threats. Hence, the credibility of a given message can be eroded by future communication opportunities, and intra-play communication may thereby impede rather than facilitate cooperation. This reasoning has been acknowledged in the literature on renegotiation proofness.² The following quote from Bernheim, Peleg, and Whinston (1987) captures the core of these theoretical studies:

”When players have unlimited ability to communicate and reach non-binding agreements regarding their strategy choices, a meaningful agreement requires more than the Nash best-response property. This is true because coalitions of players can typically ar-

¹This line of argument has not been undisputed theoretically (see Aumann (1990)).

²See for example Bernheim, Peleg, and Whinston (1987), Blume (1994), Farrell and Maskin (1989), Bernheim and Ray (1989) and van Damme (1989).

range mutually beneficial agreements to deviate from a Nash equilibrium". Bernheim, Peleg, and Whinston (1987)

Hence, in the presence of intra-play communication opportunities, a message should propose actions that are immune to such renegotiation opportunities in order to be credible.

The experimental evidence of communication in multi-stage games comes mostly from repeated static games. These studies point to the fact that the cooperation-enhancing effect of communication depends on the underlying static game. For instance in posted offer markets the effect of communication is weak and loses its impact with repetition (Holt and Davis, 1990; Cason, 1995).³ On the other hand, for other types of games, such as prisoners' dilemma and coordination games, numerous experimental studies have demonstrated cooperation and coordination enhancing effects of communication.⁴ Moreover, Bochet, Page, and Putterman (2006) show, in a public goods game, that the level of cooperation might depend on the type of communication allowed. However, none of these studies explicitly analyze the credibility issues that arise in multi-stage games.⁵

The aim of this paper is to test the effect of introducing intra-play communication on the credibility of pre-play communication and cooperation. We use a simple set-up, which is yet rich enough to generate clear predictions based on renegotiation proofness. In short, subjects first play a prisoners' dilemma game and then a coordination game with two pareto-ranked equilibria. The payoffs of the game are such that players can sustain cooperation in the prisoners' dilemma by threatening to play the inferior equilibrium in

³See also Isaac and Plott (1981), Isaac, Ramey, and Williams (1984), and Isaac and Walker (1985) for an early series of experiments that investigate the impact of face-to-face communication in repeated double auction, posted offer and sealed bid offer games. In general, they find that subjects are unable to sustain collusive outcomes. Bochet and Putterman (2009) reports no effect of structured communication in repeated public good games.

⁴See for example Ledyard (1995) and Sally (1995) for comprehensive surveys of the extensive literature on communication in social dilemma situations and Cooper, Dejong, Forsythe, and Ross (1992), Crawford (1998) and Blume and Ortmann (2007) for studies of coordination games. Similarly, evidence from recent auction experiments reviewed in Kagel and Levin (2008) suggest that communication typically lead to lower prices in many repeated auction settings; one reason is that communication facilitates bid rotation among the buyers.

⁵The only exception is a recent paper by Cooper and Kuhn (2010). We will return to a more detailed discussion of this paper and other related literature in Section 5.

the coordination game. Pre-play messages signalling such intentions are likely to be considered credible if intra-play communication is not allowed. In contrast, if players can communicate intra-play, the threat of punishment may be non-credible since players can renegotiate away from costly punishments. We run two treatments; one where only pre-play communication is allowed and one with both pre-play communication and intra-play communication (i.e. subjects get a second chance of communicating in-between the prisoners' dilemma and the coordination game). Pre-play communication has a significant and positive impact on cooperation when no intra-play communication is possible. In particular, we find that sending or receiving any message has a positive impact on cooperation, but the effect is stronger if the message contains intentions of cooperation. In contrast, no such effects are found when intra-play communication is also allowed. This result is in line with the predictions that intra-play communication opportunities may impede the credibility of pre-play messages trying to establish cooperation.

The rest of the paper is organized as follows: Section 2 presents the theoretical model that forms the basis for our experimental setup, Section 3 describes the experimental setup, Section 4 shows the results, Section 5 discusses related literature and Section 6 concludes the paper.

2 Theoretical results

Consider a two-player two-stage game $\Gamma = [\{1, 2\}, \{g^i(s)\}_{i=1}^2, \{S^i\}_{i=1}^2]$ where S^i is player i 's strategy space, and $g^i : S^1 \times S^2 \rightarrow \mathbb{R}$ give i 's utility for each profile s of strategies. In what follows we will theoretically analyze the game in Figure 1, denoted $\hat{\Gamma}$, which will form the foundation of the experiment described in Section 3. The first stage is a prisoners' dilemma game and the second stage is a coordination game. It is straightforward to see that the only Nash equilibrium of stage 1 in $\hat{\Gamma}$ is the action pair (D_1, D_2) . Indeed, it is a dominant strategy equilibrium. The second stage is a coordination game with two Nash-equilibria that entails playing either action pair (H_1, H_2) or (L_1, L_2) . The two Nash-equilibria are pareto-ranked so that both players strictly prefer the outcome of (H_1, H_2) to that of (L_1, L_2) .

It is well known that in two-stage games there may exist subgame perfect equilibrium (SPE) strategies that do not prescribe playing a Nash equilibrium action profile in each stage game. In the literature, these kinds of

Stage 1		
	D	C
D	80,80	180,70
C	70,180	140,140
⇓		
Stage 2		
	H	L
H	150,150	60,90
L	90,60	100,100

Figure 1: **The two-stage game $\hat{\Gamma}$.**

strategies are often referred to as cooperative or collusive strategies. We follow the former strand of literature and, in order to separate cooperative equilibria from non-cooperative equilibria, we say that a strategy profile is *Cooperative* if it does *not* entail playing actions (D_1, D_2) in stage 1. Clearly, a strategy profile that prescribes playing (D_1, D_2) in stage 1 and any of the two Nash-equilibria in stage 2 is subgame perfect. However, the following Proposition shows that $\hat{\Gamma}$ has Cooperative equilibria as well.

Proposition 1 *Any cooperative action profile of stage 1 can be sustained as a subgame-perfect equilibrium outcome.*

Proof. To validate this claim we start out by analyzing stage 2. Since this is the last stage of the game, it is obvious that the only possible equilibrium actions must entail playing either (H_1, H_2) or (L_1, L_2) . Now, looking at stage 1 it is straightforward to see that any combination of stage 1 actions can be sustained as a SPE by using the stage 2 action pairs (L_1, L_2) as threats if necessary. In particular, any of the action profiles (C_1, C_2) , (C_1, D_2) or (D_1, C_2) can be sustained by the threat of playing (L_1, L_2) in case of defection and (H_1, H_2) otherwise. ■

From the vantage point of standard game theory, the viability of co-operation does not rely on the presence of communication opportunities. Moreover, any equilibrium survives the presence of communication since it can always be ignored. However, given the multiplicity of equilibria in $\hat{\Gamma}$, communication might serve as a coordination device to select an equilibrium. In addition communication may be particularly important for those equilibria that require coordinated punishments.

There exists ample experimental evidence showing that communication has coordination enhancing effects in many static games (see for example Crawford (1998)). With these results in mind, one might first expect that communication helps coordination on efficient SPE in multi-stage games as well. However, as discussed in the introduction, allowing for communication in multi-stage games introduces new complexities.⁶ As a consequence, we need to distinguish between communication that occurs pre-play and communication that occurs intra-play. When only pre-play communication is allowed and given that players are egoistic and rational, it is natural to assume that, a message has to prescribe actions that are part of an SPE strategy profile. As argued in the introduction, when intra-play communication is allowed, it is reasonable to require in addition that the proposed actions are immune to renegotiation opportunities. In what follows we give the definition of renegotiation proofness as given in Bernheim, Peleg, and Whinston (1987) with the restriction to two-player games.⁷

Definition 1 *A strategy profile $s^* \in S_1 \times S_2$ in Γ is **Renegotiation-proof**:*

- i) if the restriction of s^* to any proper subgame of Γ is a Nash equilibrium of that subgame.*
- ii) And if for any proper subgame of Γ , there does not exist another **Renegotiation-proof** strategy s' such that $g^i(s') \geq g^i(s^*)$ for $i = 1, 2$, with a strict inequality for at least one player and the restriction of s^* and s' to that subgame.*

As we will see in the subsequent proposition, there is a stark difference in the equilibrium prediction between subgame perfection and renegotiation

⁶Once again, we note that if players can commit to ignoring any intra-play communication, then punishments are credible.

⁷Note that Bernheim, Peleg, and Whinston (1987) refers to this concept as perfectly coalition-proof Nash equilibrium.

proofness.

Proposition 2 *The game $\hat{\Gamma}$ has a unique renegotiation-proof strategy profile $s^* = \{(D_1, D_2), (H_1|h, H_2|h)\}$ where h denotes any history of stage 1 behavior. Clearly, the strategy profile is non-Cooperative.*

Proof. To validate this, we start out by analyzing stage 2. First, we have to check that there are no incentives to renegotiate in any subgame starting at stage 2. In addition, any strategy must be a Nash equilibrium of that subgame. It is easy to conclude that the action pair (L_1, L_2) cannot be part of any renegotiation-proof strategy, since players would then have a collective interest in renegotiating to (H_1, H_2) . To see the force of this requirement we now look at stage 1. First note that a renegotiation-proof strategy cannot entail anything other than (H_1, H_2) in stage 2; thus no threats of punishments are possible. The implication of this is that no action pair other than (D_1, D_2) can be sustained as a renegotiation-proof equilibrium in stage 1. Thus, the game $\hat{\Gamma}$ has a unique renegotiation-proof strategy profile that is non-cooperative. ■

Based on these two propositions we now state our first hypothesis.

Hypothesis 1 *The frequency of C actions will be higher in $\hat{\Gamma}$ when players can only communicate pre-play compared to the case when they can communicate both pre-play and intra-play.*

Even though communication opportunities lie at the heart of the theory of renegotiation proofness, communication does not enter into Definition 1 above. Nor can we make any predictions regarding whether players will communicate and what type of communication is to be expected in equilibrium. Therefore, the following paragraph is somewhat more explorative.

As mentioned earlier, we know from the experimental literature on static games that communication of intentions serves as an important device to achieve coordination on Pareto efficient equilibria. A naïve extrapolation of these results to $\hat{\Gamma}$ suggests a strong cooperative effect from sending or receiving a message containing an intention of playing a cooperative strategy. However, since intra-play communication provides subjects with means for renegotiation, we conjecture that it is relevant to restrict attention to renegotiation proof strategies when intra-play communication is available.

Therefore, as a direct consequence of Proposition 2, we expect pre-play messages expressing cooperative intentions not to be credible in situations with opportunities of intra-play communication. Consequently, we also expect the general level of cooperation to be higher in the absence of intra-play communication opportunities, as players are able to more successfully coordinate on a cooperative equilibrium using the pre-play messages.

Based on these arguments, we now end this section by stating our second hypothesis.

Hypothesis 2 *Sending or receiving a message that suggests playing a cooperative strategy has a stronger effect on the probability of playing the C action in $\hat{\Gamma}$ when players can only communicate pre-play compared to the case when they can communicate both pre-play and intra-play.*

We proceed by describing the experimental design that sets out to test these hypotheses.

3 The Experiment

The basis of the experiment is the two-stage game depicted in Figure 1, which was analyzed in the preceding section. To test our hypotheses we varied the structure of communication between treatments in the following way:

- **Treatment P – Pre-play Communication:** Subjects played game $\hat{\Gamma}$, with communication only allowed before the first stage (pre-play communication).
- **Treatment PI – Pre-play and Intra-play Communication:** Subjects played $\hat{\Gamma}$, with communication allowed both before the game and between the two stages of the game (both pre-play and intra-play).

As mentioned above, the theory on renegotiation proofness is silent regarding the mode of communication. However, since renegotiation assumes the ability to coordinate, two-way communication seemed like the natural choice. Moreover, previous experimental findings emphasize that two-way communication is superior in facilitating coordination in symmetric games (see for example Crawford, 1998). When allowed, communication took the

following form: Both subjects in a pair were given the opportunity to send a cheap-talk message of their intended play in the two-stage game.⁸ We designed the pre-play communication stage so that the intended action in the second stage could be made contingent on the opponent's choice in stage 1.⁹

Regarding the choice between free-form communication and structured communication, we conjectured that free-form communication, would lead subjects to form more of a group identity, thus increasing the importance of social preferences. In addition, if communication is free-form, we might invalidate the renegotiation arguments since subjects can deliver verbal punishments and it is not clear that both subjects have an incentive to renegotiate away from those.¹⁰ More precisely, subjects indicate the action, C or D , for stage 1. For stage 2 subjects indicate the action, H or L , they would choose if the other player chooses C in stage 1, and which action, H or L , they would choose if the other player instead chooses D in stage 1. This pre-play message structure was identical between the two treatments. In the PI treatment, subjects had an additional opportunity to communicate before the second stage of the game. At this stage, subjects wanting to communicate could indicate whether they would play H or L in stage 2. For both pre-play and intra-play communication, if both wanted to send a message, a sequential design was implemented where it was randomly determined, with equal probability, which of the players got to send the first message. The second player was then informed about the message of the first player before sending his/her suggestion. We chose the sequential structure since we found it most suitable for studying renegotiation. In particular, we believed that this structure would minimize the risk of subjects failing to coordinate verbally before the game. See the appendix for a detailed description of the communication protocol.

⁸Note that contrary to most previous studies the choice to communicate was endogenous. Exceptions are Andersson and Wengström (2007) and Andersson and Holm (2010).

⁹This allows subjects to communicate threats central to sustaining cooperation in multi-stage games.

¹⁰Cf. the discussion of Cooper and Kuhn (2010) in section 5. Therefore, we implemented structured communication, where only possible contents were the intended action choices in the game.¹¹

The experiment was conducted at Lund University, Sweden in 2006.¹² A total of 98 subjects participated in eight sessions. The subjects were recruited by posters and targeted emails to students of introductory economics classes. A between-subjects design was implemented in which subjects played one of the treatments. After each stage, they learned the outcome of that stage. On completing the second stage, they were anonymously re-matched with a new subject and, in order to avoid reputation-building, they never met the same subject again. In total, each subject played the two-stage game eight times. At the end of the experiment, the accumulated payoffs during the eight rounds of play were converted into Swedish kronor according to an exchange rate of 1 experimental currency unit = 0.15 kronor.¹³ Subjects also received a show-up fee of 20 kronor and the average earnings were 206 kronor. The experiment took approximately 1 hour and 15 minutes to conduct.¹⁴

4 Results

We start addressing our two research hypothesis stated in Section 2 by looking at subjects' behavior in stage 1. Thereafter, we analyze stage 2 behavior using a more explorative approach.

4.1 Stage 1

Table 1 reports the rate of C actions in stage 1 for each period of the experiment. We note that the cooperation rate is higher in the P treatment in all but two periods. Acknowledging the fact that individual choices cannot be considered independent observations (even in the first period due to the opportunity of sending messages), we calculate the fraction of times when at least one individual in the group choose a C action. This measure is indeed independent across pairs in the first period. Using Fisher's exact test we found no difference between the two treatments in the first period

¹²The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

¹³At the time of experiment 1 kronor \approx EUR 0.11.

¹⁴A transcript of the instructions and the communication stage procedure can be found in the appendix.

(two-sided $p=0.466$).^{15,16} Taken together, although we observe higher level of C actions in the P treatment, we do not find significant statistical support for Hypothesis 1.

Table 1: **Rate of C Actions per Period**

Before proceeding to test Hypothesis 2—that cooperative messages were more closely associated with cooperative play in the P treatment—we categorize the messages into a few broad categories. Taking the perspective of player i we divide the messages into the following categories, where the first argument denotes intended play in stage 1, and the second argument denotes intended play in stage 2 given the opponent's (j 's) stage 1 action:

- Punish/Reward message: $PR = (C_i, (H_i|C_j, L_i|D_j))$
- Reward/Reward message: $RR = (C_i, (H_i|C_j, H_i|D_j))$
- Defect and Coordinate message : $DC = (D_i, (H_i|C_j, H_i|D_j))$
- Other Message: $OM =$ a message that is not PR, RR or DC
- No Message: $NM =$ no message.

The PR type of message corresponds to the predicted equilibrium strategies in the P treatment, whereas a DC message indicates the renegotiation-proof strategy in the PI treatment. Table 2 reports the rates of different messages. Since the theoretical model is silent in regards to what messages players should send, there is no clear-cut reference chart to compare with. However, with previous experimental results in mind, we expect the bulk of messages to be used for signalling cooperative intentions. Overall, the most common type of message is the RR message followed by the PR message. This might seem surprising given the equilibrium strategies. One possible

¹⁵The test produced insignificant results in all periods except period 4. However, one needs to be careful here since observations cannot be considered independent beyond period 1.

¹⁶Similar result were also obtained from various probit and linear probability model regressions using C action as the dependent variable, clustering at the session level. The sign of the treatment variable indicated lower levels of C actions in the PI treatment, but this effect was not statically significant, neither in specifications with only the treatment dummy nor in specifications including the same control variables as the basic specification of table 8 in B

explanation for the high number of RR messages is that some subjects are driven by a focus on maximizing the social efficiency of each stage game, ignoring the strategic element of the two-stage game.¹⁷ We will take more explicit account of preferences for efficiency in the simple theoretical model that we sketch in A. Finally, we note that in both treatments it is common that subjects decided not to send a message.

Table 2: **Rate of Messages**

To investigate the importance of communication for cooperation in stage 1, we calculated the fraction of times each subject played C given a certain type of message. The pattern of messages in each pair is divided into the following categories:¹⁸

- PR_{11} (RR_{11}) : both players sent a PR (RR) message.
- PR_{10} (RR_{10}) : player i sent a PR (RR) message but player i 's opponent did not.
- PR_{01} (RR_{01}) : player i did not send a PR (RR) message but player i 's opponent sent a PR (RR) message.
- NM_{11} : neither of the players sent a message.

Table 3 shows the averages of the individual fractions of C choices, given the specified messages structure (the decision-maker as player i). Two things are worth pointing out. First, the frequencies of C actions are higher after one of the pair sent a message, compared to the case when neither of the subjects sent a message. Second, for all groups with communication, the fractions of C actions are higher in the P treatments than in the PI treatments. To test whether these differences are significant, we start out by using average level of C actions at the session level as observations. Using the Mann-Whitney test, we find that the level of C actions are higher in the

¹⁷Yet another potential explanation of the high number of RR messages is offered in a recent experimental study by Houser, Xiao, McCabe, and Smith (2008). They find a similar pattern in their data and argue that threats create a cognitive shift that crowds-out norm-based motivations for cooperation. Even though their experiment is different from ours, we conjecture that this might provide a partial explanation of observed behavior. Indeed, if subjects anticipate this effect, they will refrain from using such messages. Similar results are obtained for trust games (Fehr and Rockenbach, 2003).

¹⁸ DC messages are excluded from Table 3 due to their low frequency.

P treatment when both sent PR messages, PR_{11} (two sided p-value: 0.028) and for RR_{10} (two sided p-value: 0.043). The first result is in fully in line with our hypothesis that subjects intending to implement punish reward strategies are more likely to carry out such a strategy in the P treatment. The latter result is more surprising but underlies that messages are treated differently in the two treatments. As this approach is very conservative, making use of only the 8 session averages as observations, we move on to exploring the role of communication using regression analysis.

Table 3: Fraction of C actions given a certain message in Stage 1

We estimate a linear probability model (LPM) using stage 1 action as the dependent variable (1 indicating C and 0 indicating D) with individual random effects and clustered standard errors on the session level.¹⁹ As independent variables we use the different types of messages, the actions of the subjects' previous opponents and a linear time trend. The results of these regressions are summarized in Table 4.

Table 4: Linear probability model, individual random effects, C Actions in Stage 1

The first specification (1) of table 4 confirms that there is no significant overall treatment effect on the propensity of playing C in the first stage. More interestingly, the other specifications in the table confirm the pattern from table 3; namely that messages have different impact on actions in the two treatments. First, in specifications (2) and (3) we observe that sending messages of types PR and RR has a significant impact on the propensity of playing C . In addition receiving PR and RR also has some effects. However, from specifications (4) and (5) it becomes evident that these effects are mainly driven by the effect messages has in the P treatment. The interaction terms between message-types and the PI treatment shows that sending PR and, to a somewhat lesser extent, RR have significantly lower effect on

¹⁹We also estimated an LPM without individual random effects as well as an LPM with individual fixed effects (still clustering at the session level). Both provided nearly identical results. See Table 8 and Table 9 in B for details. In addition, we estimated a probit model and it also yields similar results but at the cost of very lengthy analysis since we cannot take coefficients or their significance level at face value (Ai and Norton, 2003).

playing C . The same is true for receiving PR and RR . Taken together, we have confirmed that sending and receiving messages, especially of the PR type, has a significantly higher effect on the propensity of playing C in the P treatment, which is in line with Hypothesis 2.

Taken together these findings give strong support to Hypothesis 2 and highlights an interesting and nontrivial difference concerning the effects of pre-play communication between the two treatments. The strong cooperative effect of pre-play communication vanishes when further communication possibilities are introduced. The next step is to explore if the results from stage 1 can be explained by subjects' behavior in stage 2. In particular we set out to test the history independence assumption that underlies the renegotiation argument.

4.2 Stage 2

In this section we turn to investigating behavior in the second stage of the game. The theoretical framework laid out in Section 2 suggests that subjects may be able to coordinate on the Pareto-dominant stage-game equilibrium (H_1, H_2) in the second stage in both treatments. In the P treatment, we expect subjects to be able to coordinate on a cooperative SPE entailing (H_1, H_2) along the equilibrium path in stage 2. In the PI treatment, the argument is that subjects would use intra-play communication to coordinate on (H_1, H_2) , regardless of the behavior in the first stage. Underlying this argument is an assumption that subjects let bygones be bygones, leading to equilibrium strategies that are history independent.

The data from stage 1 presented in the previous section reveals that subjects' behavior is not entirely captured by the stylized behavior along the equilibrium paths described in the previous paragraph. In particular, the treatment effect on the overall propensity to play C is less strong than expected. Investigating the data from stage 2 is hence of special interest as it may provide explanations for the observed heterogeneity in stage 1 behavior.

Our analysis of the stage 2 actions starts out by giving the following labels to the four possible stage 1 outcomes for player i :

Loser:	(C_i, D_j)
Winner:	(D_i, C_j)
Mutual defection:	(D_i, D_j)
Mutual cooperation:	(C_i, C_j)

Table 5 depicts the frequencies of H actions at stage 2 conditional on the stage 1 outcome. The first column shows frequencies unconditional on what type of pre-play message was sent (if any). Whereas column two and three show the corresponding frequencies when at least one in the pair sent a PR or RR message. Following mutual cooperation almost all subjects coordinated on (H_1, H_2) in stage 2. At the other end of the scale, the losers display the lowest levels of H actions for each category. We also observe that, conditional on mutual defection, there is a slightly higher fraction of H actions in the PI treatment lending support to the renegotiation hypothesis. Yet this difference is not significant. Calculating the fraction of times that the subjects played H conditional on the outcome in stage 1 allows us to test the (bygones be bygones) assumption that players will always coordinate on the high equilibrium when intra-play communication is possible. To operationalize this, we calculate for each session the average rate of playing H conditional on mutual cooperation or not. Using the Wilcoxon matched-pairs test we find that there is a significant difference at the 10 percent level (two sided p-value= 0.0679).

In order for the PR message to be credible the L action must be chosen after defection by the opponent, in the P treatment. The fraction of H actions after a PR message has been sent and the opponent chose the D action is 0.6 in the P treatment. So in a majority of times the threat was not carried out. Yet there are only 10 observations where a PR message was followed by a D action in the P treatment so we are not keen on drawing any far-fetched conclusions from this observation. We note that the corresponding fraction in the PI treatment is 0.72, which indicates that the comparative statics goes in the right direction, even though the difference is not statistically significant.²⁰

Table 5: **Fractions of H actions in Stage 2**

²⁰The corresponding number of observations in the PI treatment was 32.

Another interesting issue regarding behavior in stage 2 is whether subjects in the PI treatment used the opportunity to communicate intra-play. Table 6 below summarizes the frequency of the different types of messages broken down by outcome in stage 1. Notably, lowest frequency of messages indicating H is observed for the losers, again suggesting that not all subjects seem to let bygones be bygones. Interestingly, the highest rate of H messages is observed after mutual defection in Stage 1, indicating that subjects that play along the equilibrium path outlined by the renegotiation-proof equilibrium do indeed use messages to coordinate on the Pareto dominant stage-game equilibrium (H_1, H_2) .²¹

Table 6: **Intra-play messages in the PI treatment**

%begincenter

In conclusion, stage 2 behavior seems not to be in line with the standard theoretical predictions put forward in Section 2.²² Subjects appear to have behaved differently both in terms of messages and actions depending on the outcome of the first stage, thus rejecting the history independence hypothesis. Hence, although the analysis of stage 1 data gives support to our hypotheses, the stage 2 data suggests that not everyone let bygones be bygones. In A, we try to rationalize our results by extending the theoretical model in Section 2 to incorporate distinct behavioral types. More precisely, in addition to the type of sophisticated players considered previously, we add players that are driven by either reciprocal preferences or efficiency concerns. These two types of motives have previously been found to be important in contexts related to ours. For example, when classifying subjects according to a range of norms, Lopez-Perez and Vorsatz (2009) report that reciprocity and efficiency are the two most prevalent norms. Extending the model with such types, we can find an equilibrium that explains some of the core deviations from the theoretical framework outlined in Section 2. Firstly, the introduction of the efficiency driven types implies that we can observe RR messages in equilibrium. Secondly and less trivial, in the equilibrium we characterize, there is a higher correlation between sending or receiving PR

²¹Yet in absolute terms only 44.9 percent of the times such a message was sent.

²²To have more thorough look at behavior in stage 2 of the PI treatment, we also ran regressions on the propensity to send an H message as well as the propensity to play H . These regressions are displayed in Tables 10 and 11 together with a short discussion in B.

or RR messages and playing C in the P treatment than in the PI treatment. Thirdly, since the composition of types will be different for a given stage 1 outcome in the two treatments, the fraction of L actions in stage 2 is expected to be higher in the PI treatment.

5 Comparison to Previous Literature

The experimental literature concerning renegotiation proofness is scarce. A notable exception is Davis and Holt (1999) wherein the occurrence of renegotiation in two-stage games is investigated.²³ Contrary to the renegotiation prediction, they find some evidence that subjects use punishments to sustain cooperative strategies. However, since they do not allow for communication, subjects are left without any obvious means of coordinating on the mutually efficient deviation, which is central to the renegotiation argument.

Based on a renegotiation argument, Andersson and Wengström (2007) experimentally investigate the effect of (possibly) costly intra-play communication in a repeated Bertrand duopoly. They find that costless intra-play communication leads to unstable collusive coalitions, whereas restricting renegotiation by making communication costly results in a high frequency of the collusive outcome.

In a recent paper, Cooper and Kuhn (2010) investigate the effect of intra-play communication on cooperation. Interestingly, they do not find the same negative association between cooperation and intra-play communication as we do. They report that allowing for intra-play communication in addition to pre-play communication raises cooperation levels. Their approach has many similarities with the approach in this paper; however, it differs at least in one important way and as we will argue this might explain the differences in our results. Just as we do in this paper, they set up a two-stage game where a prisoners' dilemma is followed by a coordination game with Pareto-ranked equilibria.

The major difference compared to our approach is that they use written free-form communication. The use of free-form communication allows subjects to deliver verbal punishments in case of defection, which may reduce the scope for renegotiation. Renegotiation builds on the idea that defecting

²³See also Croson, Gomes, McGinn, and Nöth (2004) for related paper that experimentally studies coalition formation in a merger and acquisition context

from a cooperative path may be profitable since there is a mutual interest in renegotiating away from costly punishments. If free-form messages can be used to hurt the defector, while making the sender of the message better off, defecting from the cooperative path may no longer be viewed as profitable. Put differently, even though renegotiation is profitable in terms of actions, if the perceived cost of a verbal punishment outweighs the gain of defecting, a player would prefer to cooperate in the first stage.²⁴ Consistent with this line of reasoning, Houser and Xiao (2005) report that the threat of verbal punishments increases offers in ultimatum games.

In addition, recent research shows that free-form communication dramatically increases altruistic behavior, which may further remove the incentive for defecting in the first period (see for example Mohlin and Johannesson (2008) and Andreoni and Rao (2010)). Although we are not aware of any study that directly compares free-form communication with the type of structured communication we implement, related evidence by Lundquist, Ellingsen, Gribbe, and Johannesson (2007) suggests that the altruism-inducing effects are likely to be stronger with free-form communication than structured communication. Taken together, there exist several reasons to believe that free-form communication will weaken the outlook for renegotiation-proofness to have any relevance, which may in turn explain the difference between our results and those of Cooper and Kuhn (2010).

Our results also relate to the findings in Ellingsen, Johannesson, Lilja, and Zetterqvist (2009). In their setup, subjects play a prisoners' dilemma unaware that they will meet the same subject again in a subsequent bargaining game. As in our setup, they either allow or disallow intra-play communication, whereas pre-play communication is always allowed. In their conclusion, they conjecture that knowing they will meet the same subject again in a subsequent game will induce subjects to be less opportunistic in the prisoners' dilemma. Although not directly related, our results indicate that the degree of opportunism may in addition depend on the presence of intra-play communication.

The notion of guilt aversion (Charness and Dufwenberg, 2006) and the similar concept of consistency (Ellingsen and Johannesson, 2004) can also be mentioned. These models assume that people are acting to avoid letting

²⁴Indeed, there is support for the renegotiation hypothesis in their data. The authors report that, given that a defection has occurred, a higher degree of renegotiation actions in the intra-play treatment.

others down. Ellingsen and Johannesson (2004) argue that communication creates commitment, and people face a personal cost of being inconsistent. Charness and Dufwenberg (2006) claim that communication may shift players' perceptions of what others expect them to do.²⁵ Guilt can be a potential explanation of why we do not observe as large treatment differences as we expect. Indeed, if guilt is a very strong driver of behavior, then we should not observe any treatment difference in our experiment. Yet, behaviorally it is not evident to us how this commitment device depends on the possibility of future communication as our results suggest. Moreover, since we chose a narrow communication channel, only allowing subjects to send pre defined messages, we conjecture that the guilt aversion effect is rather weak in our setting.

Taken together, the results of our paper point out that the timing and frequency of communication are important and should be taken into account when analyzing data. A recent paper on voluntary contribution games with face-to-face, chat-room and numerical communication (Bochet, Page, and Putterman, 2006), finds that it is only with face-to-face communication that subjects are able to sustain high levels of contributions to the public good.²⁶ However, in addition to changing the mode of communication between treatments, the authors also alter the timing and frequency of communication.²⁷ The authors attribute their results to the stronger commitment effect in face-to-face and chat-room communication. In light of our results, this effect could also be driven by the change in frequency and timing of communication in their experiment. In particular they find that there is a drop in contributions in rounds that are succeeded by a communication round, possibly indicating that subjects hope to be able to renegotiate in the communication stage.²⁸ Future research should try to disentangle these

²⁵In a recent experimental paper Vanberg (2008) is able, by a clever design, to distinguish between these two potential explanations. His findings support the commitment effect of Ellingsen and Johannesson (2004). Ellingsen, Johannesson, Tjötta, and Torsvik (2010) find a similar result.

²⁶See Brosig, Weimann, and Yang (2004) and Charness and Dufwenberg (2010) for similar results.

²⁷More specifically: In face-to-face communication treatment subjects only communicate pre-play; in the chat-room communication treatment subjects communicate pre-play as well as before period 4 and 7 (out of 8 periods); in the numerical communication treatment subjects communicate pre-play as well as before every period.

²⁸We thank an anonymous referee for pointing this out.

two potential explanations.

6 Conclusions

In this paper, we find that more communication possibilities do not necessarily lead to more cooperation. We report that sending or receiving pre-play messages has a positive and significant effect on cooperation if there is no possibility of intra-play communication. However, no similar effect is found when intra-play communication is allowed. This finding points out that the credibility of pre-play messages, signalling intentions to cooperate, depend on the opportunities of communication at later stages of the game.

Together with our earlier work presented in Andersson and Wengström (2007), our findings point to a robust effect of communication in multi-stage games. More communication opportunities do not imply more cooperation. In the light of the existing results from static games that point out the cooperative-enhancing effect of communication, our findings highlight that the fact these results do not directly carry over to multistage games.

With our parameters, signalling H after being defected upon and then playing L in the hope that the defector played H is not that costly for the proposer, but delivers a harsh punishment to the defector. Expecting that some subjects might use this opportunity may have reduced our treatment differences. Therefore, one interesting extension to our work would be to explore the effects of making coordination in the second stage harder, for example by increasing the number of players or making the Pareto dominant equilibrium more risky.

A A Simple Theoretical Extension

The theoretical model in Section 2 implicitly assumes that players are rational and selfish. However, as mentioned earlier, there is extensive experimental evidence from a wide range of situations that reciprocity is an important driving force behind cooperative behavior. A large fraction of experimental subjects are willing to cooperate as long as others also cooperate (Fehr and Fischbacher, 2003). Furthermore, driven by negative reciprocity, people have been found to engage in costly punishments of people who do not cooperate (Fehr and Gächter, 2000). Hence, in our setting, it is not far-fetched

to believe that some subjects will not renegotiate after being cheated, but instead choose to punish the cheater by playing the L action in the second stage.

Efficiency concerns have been singled out as another important driver of cooperative behavior in economic experiments (Engelmann and Strobel, 2004).²⁹ We note that existence of such preferences in our subject pool might shed light on why we observe a high fraction of RR messages. In a recent experimental paper, Lopez-Perez and Vorsatz (2009) classify subjects according to a range of norms and report that many subjects have either preferences for efficiency or reciprocity. This gives us further support to focus on these two types of players when trying to understand behavior in our experiment.

In what follows, we will present a very simplistic extension of our previous model by adding players who comply with these two norms (reciprocity or efficiency).³⁰ We assume that these players are honest in the sense of sending messages that signal their true intentions in the game.³¹ In addition, we assume that, since these players are complying with a norm, they do not update their behavior upon sending or receiving any message. This last assumption might, of course, be questioned but since the game with communication has up to six stages, where players can update their beliefs about what type of opponent they are facing, we restrict attention to such player types for sake of tractability. We denote these types (E)fficient types and (R)eciprocal types respectively. In addition to E and R types we assume that there exist (S)ophisticated types, who are rational and selfish. Assume that the type S 's prior distribution of beliefs over player types is uniform.³²

Here is a summary of the main assumptions behind the three types:

²⁹Engelmann and Strobel (2004) define efficiency as maximizing the sum of payoffs.

³⁰In particular, we will not make any attempt at formulating utility functions that rationalize the behavior of these two types.

³¹See Ellingsen and Östling (2010) and Demichelis and Weibull (2008) for recent papers that also introduce a preference for honesty.

³²This plausibility of such a distribution of types is given support by Lopez-Perez and Vorsatz (2009), who report that that reciprocal types and efficiency types are about equally likely. Moreover, in a meta study covering 129 studies of the dictator game, Engel (2010) finds that 36% of the dictators give nothing to the recipient.

- E* type:** Plays strategy: $(C_i, H_i | C_j, H_i | D_j)$
Sends truthful messages (i.e. RR and H)
Does not update beliefs
- R* type:** Plays strategy $(C_i, H_i | C_j, L_i | D_j)$
Sends truthful messages (i.e. PR and H or L)
Does not update beliefs
- S* type:** Chooses messages and actions strategically
Believes each type (E , R and S) to be equally likely (ex ante)
Updates beliefs based on messages and stage 1 actions

In Table 7 we present a particular Perfect Bayesian Nash equilibrium (PBE) profile that we believe contributes to the understanding of the main observations in the experiment. Looking at the left part of the table (P-treatment), the first column displays the player types that are matched and the order in which they send messages. For example in the fourth row, E, R represents a situation in which one E type and one R type are matched and the E type sends the first message. Row five, R, E , represents a match with the same types but where the R type sends the first message. The second and third columns represent the messages sent by the two players; M1 indicates the first message and M2 the second message in the sequence.

The fourth and fifth columns display the action-pairs in stage 1 and stage 2 respectively. The first player type's action is the first entry in each row in columns S1 and S2. The right part of the table (PI-treatment) is identical to the left part, but with the addition of the intra-play communication stage columns, M3 and M4. Note also that in the intra-play message sequence we preserve the order from the player type column so that the first player type sends M3 and the second M4.

Table 7: PBE outcomes in a mixed population of E, R and S types.

Since only S types update their behavior in response to a particular message, we can focus on situations where S is involved. First note that, if an S type knows that she is meeting an E type, she will choose D , since that player will play the H action in stage 2 independent of history. In contrast, if an S type knows she is meeting an R type, she will always choose C . When facing an identical type, an S type chooses C in the P-treatment and D in the PI-treatment. The picture is complicated by the fact that in Table

7 there is no complete separation of types before the stage 1 actions are taken. In particular, the S type sends the same message as an R type. For instance, the second row describes the type profile described in Section 2 where two S -type players are matched. In the P-treatment, since both S and R types send identical messages, the S receiver can only deduce that she is not facing an E type. Responding with a PR is thus optimal since it confirms to an S type that she will cooperate.³³ Both players then stick to the agreement and play C, C and H, H in equilibrium.³⁴ In the PI-treatment the S type tries to mimic the behavior of an R type and send a PR message in an attempt to maximize profits. As a result, all that a receiving S type can deduce is that she is not meeting an E type. The difference from the former case is that there are renegotiation opportunities if both players are S types. Since the distribution of types is uniform, it is optimal for the S type to choose D in stage 1.³⁵ After stage 1, uncertainty is resolved and they both send H messages and play the corresponding H action pair. The situations in rows six to nine, where an S type meets an E or R type, are more simple to solve since one of the players in the pair does not update her strategy. We note that in cases where the S type sends the first message she always sends PR in order to mimic the behavior of a R type, and always plays the D action in stage 1. We leave it up to the reader to confirm that this is indeed an equilibrium strategy.

We are now in a position to compare this theoretical extension with the observations in the experiment. Firstly, the mere presence of E types explains why we would observe the RR messages in the experiment. Of course this follows trivially by assuming that E types exist, but it is worth pointing out. In addition to this observation we can make two less obvious

³³Assume that upon receiving anything other than PR , the S type always plays D in stage 1.

³⁴Of course this can be questioned since we can always argue that S types ignore any message and play the D strategy. However, in line with the discussion in Section 2 we assume that S types play the action that corresponds to the message in stage 1. Also, it is worth pointing out that the message profile PR is self-committing in the sense of Farrell and Rabin (1996).

³⁵Choosing the D action is optimal if $180\alpha + 80(1 - \alpha) + 100\alpha + 150(1 - \alpha) \geq 140\alpha + 70(1 - \alpha) + 150$ where α is the conditional probability that the opponent is of type R given that given that the opponent is known to be of S or R type. Straightforward calculations reveal that the conditional probability of meeting a R type has to be below $6/7$, a condition which is clearly met by our assumptions.

remarks:

Remark 1 *Although we will observe some D actions in stage 1 in the P -treatment, there is a higher fraction of such actions in the PI -treatment. Moreover, keeping the message structure in mind, we conclude that the correlation between sending or receiving a RR or PR message and playing the C action in stage 1 will be higher in the P -treatment.*

This first remark is consistent with the results presented from stage 1 in Section 4 where we observed that RR and PR messages were more often followed by C actions in the P treatment than in the PI treatment.

Remark 2 *There will be a higher fraction of L actions in stage 2 of the PI -treatment.*

The second remark is in line with the observations from stage 2 in Section 4 and illustrates the selection issue discussed there. In particular, the remark shows that we might observe less H actions in the PI treatment in the asymmetric cases where one of the subjects played D in the first stage (which is consistent with the data presented in the two top rows of Table 5).

B Additional tables

Tables 8 and 9 present regression estimates similar to those found in Table 4 of the main text. Table 8 presents result from a pooled regression that does not take the panel structure into account and Table 9 displays results from a panel regression using the fixed effects estimator. Standard errors are clustered at the session level in all regressions.

We also ran regressions on the propensity to send an H message as well as the propensity to play H . These regressions are displayed in Tables 10 and 11. Concerning the propensity to choose H presented in Table 10 we confirm that stage 2 behavior is related to stage 1 outcome, with the lowest frequency of H actions occurring after being in the Loser category in stage 1. Moreover, we observe that the sending or receiving an PR message has a significant positive effect on playing H . Sending H and L messages also affects stage 2 behavior. Perhaps more surprisingly, when there is mutual defection, receiving a PR message in the first stage lowers the propensity to play H in the second stage, which suggests that when both deviated the threat contained in the PR message was not carried out. In Table 11 the

probability of sending H is also shown to depend on the outcome of stage 1. In addition, we observe that receiving a DC message has a positive impact on sending H . The effect of PR messages is sensitive to the stage 1 outcome, with the Loser and Mutual Cooperation outcomes giving rise to higher levels of H messages in presence of receiving a PR message than the Winner and Mutual Defection.

Table 8: **Linear probability model, C Actions in Stage 1**

Table 9: **Linear probability model with individual fixed effects, C Actions in Stage 1**

Table 10: **Linear probability model with individual random effects, H Actions in Stage 2**

Table 11: **Linear probability model with individual random effects, H Messages in Stage 2**

C A translation of the instructions

The text in italics is only shown to participants in the PI treatment.

General information

Welcome to this experiment on economic decision making. Read the instructions thoroughly. Do not talk during the experiment. If you have any questions, please raise your hand and one of us will approach you and you may quietly ask your question.

In the experiment you will have the possibility of earning money that will be paid out to you by Löneenheten at Lund University. Whatever happens in the experiment you are guaranteed a show-up fee of 20 kronor. In addition you can earn much more. How much more will depend on your and the other participants' choices. In the experiment you earn experimental

thalers which will be converted into kronor at the end of the experiment at the rate 1 thaler = 0.15 kronor.

The experiment consists of 8 rounds with the exact same structure. Before each round you will be anonymously matched with another participant. Please note that you will be paired with a new participant in each round.

The experiment

Each round consists of two periods in which you meet the same participant. In the first period you and your co-participant are asked to choose between two options T and B. How much you earn depends on your choice, but also on your co-participant's choice. In the first figure below your payoffs for the four different outcomes are shown in bold type. Your co-participant's payoffs for the different outcomes are shown in normal type.

Period 1

		Co-participant's choice	
		T	B
Your choice	T	80 , 80	180 , 70
	B	70 , 180	140 , 140

After you have made your choices in the first period, you will be informed about each other's choices in period 1. Period 2 then follows, where you and your co-participant once again are asked to choose between two options, this time it is L and R. In the figure below your payoffs for the four different outcomes are shown in boldface. Your co-participant's payoffs are shown in normal style.

Period 2

		Co-participant's choice	
		L	R
Your choice	L	150 , 150	60 , 90
	R	90 , 60	100 , 100

After you have made your choices in the second period, you will be informed about each other's choices in period 2. Thereafter you will be re-matched with a new participant and restart in period 1. In total you will thus carry out 8 rounds consisting of 2 periods each.

Message

Before you make your choice in the first period you have the opportunity of sending a message to your co-participant where you state what you intend to do in the two periods. Your co-participant has the same opportunity to send a message to you about his/her intentions. If the both of you choose to send a message, chance will determine which one of you sends the first message. The participant who sends the first message will not see the other participant's message, whereas the one that sends last will see the co-participant's message before he/she sends his/her message.

Before you make your choice in the second period you will once again have the opportunity to send a message to your co-participant. This message will be similar to the first except that you can only state your intentions for the second period.

The messages are non-binding, that is, you do not have to choose what you state in your messages.

Before we start the actual experiment you will be asked to perform a simple test. The questions in the test are constructed to check that everybody has understood the structure of the experiment. Everybody has to answer the questions correctly before the experiment can start, so take the opportunity to read the instructions again.

D A description of the communication protocol

Communication sequence (identical for both the pre-play and intra-play communication stage)

Subjects first decided whether they wanted to send a message or not.

If only one of the two subjects indicated interest in sending a message, he sent his message and was thereafter informed that his opponent did not choose to send a message. The subject not sending a message could observe his opponent's message.

If both decided to send a message, we implemented a sequential structure and a random draw with equal probabilities deciding which of the two had to send the first message. The first sender had no information about the opponent's choice of message when sending his own message. The second sender could see what the first player sent before sending his own message. Finally, before they moved on, the first sender was informed about the second sender's message.

Communication content

In the experiment, subjects could only indicate their intended actions of the game. They indicated their intentions by clicking radio buttons.

In the Pre-play communication phase, they could indicate their intentions for the first stage of the game as well as their intentions for the second stage of the game. Note that they could condition their intentions for the second stage on their opponent's play in the first stage. That is, they could specify one intention in case the opponent played C in the first stage and another intention if the opponent played D. Note that the actions referred to as C, D H and L in the paper were labeled T, B, L and R respectively in the experiment. See Figure 2 below for a screen shot of the subject interface for the pre-play communication stage.

In the PI treatment, a second intra-play communication stage was included between the two stages of the game in addition to the pre-play communication stage described above. In the intra-play communication stage, players had the opportunity to state their intention regarding the last stage of the game. Again, they indicated their intention by clicking the appropriate radio button. See Figure 3 for a screen shot of the intra-play communication stage.

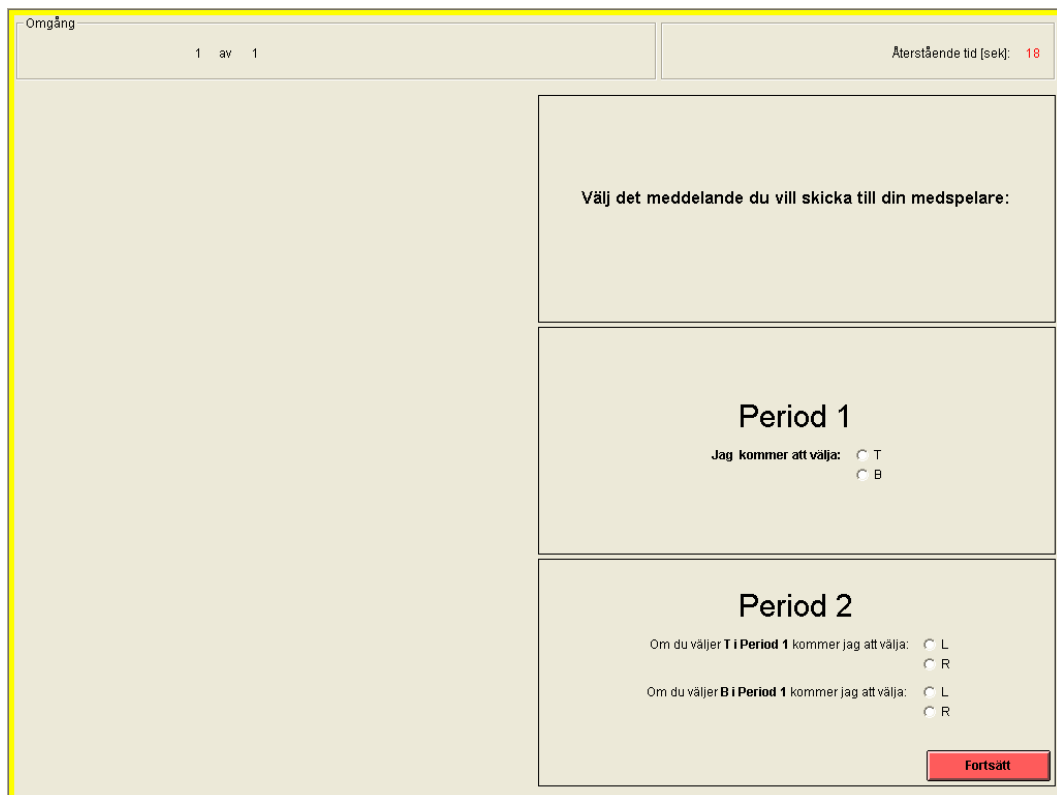


Figure 2: Screenshot of the Pre-play communication stage

Translation of pre-play communication stage:

- Top box: Choose the message that you would like to send to your opponent.
- Middle box: Period 1: I will choose:
- Bottom box: Period 2: If you choose T/B in Period 1 I will choose:

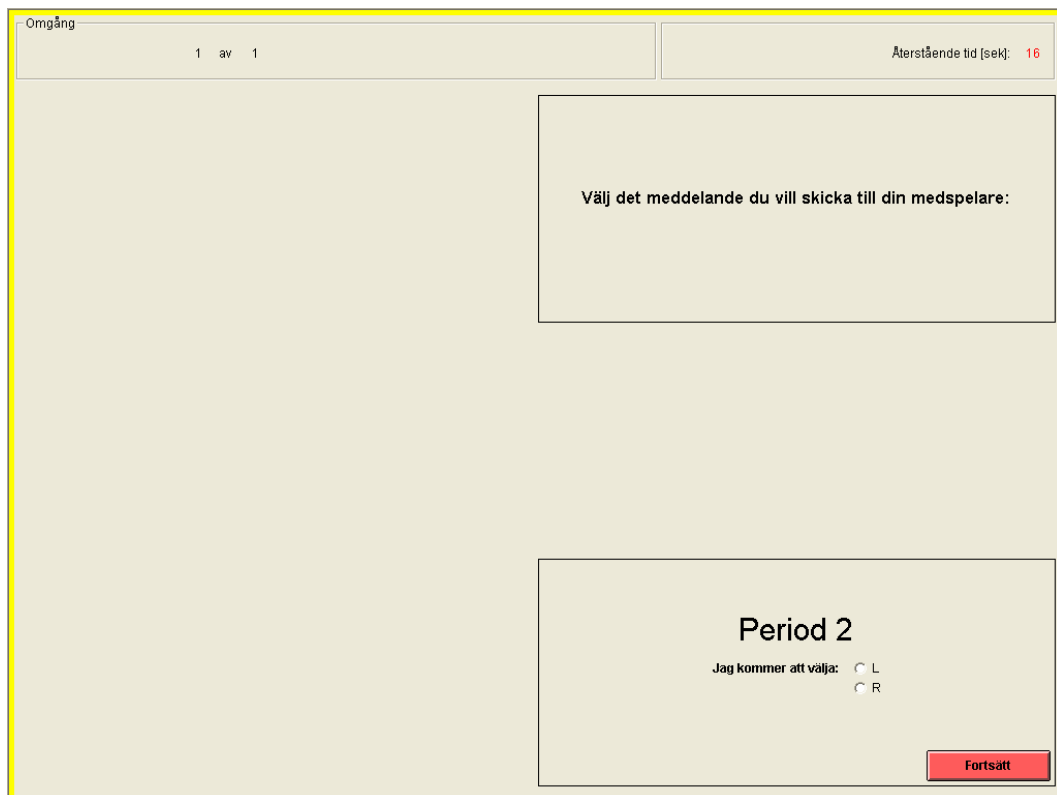


Figure 3: Screenshot of the Intra-play communication stage.

Translation of intra-play communication stage:

- Top box: Choose the message that you would like to send to your opponent.
- Bottom box: Period 2: I will choose:

References

- AI, C., AND E. C. NORTON (2003): “Interaction terms in logit and probit models,” *Economics Letters*, 80(1), 123–129.
- ANDERSSON, O., AND H. J. HOLM (2010): “Endogenous communication and tacit coordination in market entry games: An explorative experimental study,” *International Journal of Industrial Organization*, 28(5), 477 – 495.
- ANDERSSON, O., AND E. WENGSTRÖM (2007): “Do Antitrust Laws Facilitate Collusion? Experimental Evidence on Costly Communication in Duopolies,” *Scandinavian Journal of Economics*, 109(2), 321–339.
- ANDREONI, J., AND J. M. RAO (2010): “The Power of Asking: How Communication Affects Selfishness, Empathy, and Altruism,” Working Paper 16373, National Bureau of Economic Research.
- AUMANN, R. (1990): *Nash Equilibrium are not Self-Enforcing* in Gabszewicz, J., Richard, J and Wolsey L. (eds) *Economic Decision Making: Games, Econometrics, and Optimisation: Essays in Honor of Jacques Dreze*. Elsevier, Amsterdam.
- BERNHEIM, D., B. PELEG, AND M. WHINSTON (1987): “Coalition-Proof Nash Equilibria I. Concepts,” *Journal of Economic Theory*, 42, 1–12.
- BERNHEIM, D., AND D. RAY (1989): “Collective Dynamic Consistency in repeated games,” *Games and Economic Behavior*, 1, 295–326.
- BLUME, A. (1994): “Intraplay Communication in Repeated Games,” *Games and Economic Behavior*, 6, 181–211.
- BLUME, A., AND A. ORTMANN (2007): “The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria,” *Journal of Economic Theory*, 132(1), 274–290.
- BOCHET, O., T. PAGE, AND L. PUTTERMAN (2006): “Communication and punishment in voluntary contribution experiments,” *Journal of Economic Behavior & Organization*, 60, 11–26.

- BOCHET, O., AND L. PUTTERMAN (2009): “Not just babble: Opening the black box of communication in a voluntary contribution experiment,” *European Economic Review*, 53(3), 309–326.
- BROSIG, J., J. WEIMANN, AND C. YANG (2004): “Communication, Reputation, and Punishment in Sequential Bargaining Experiments,” *Journal of Institutional and Theoretical Economics*, 160, 576–606(31).
- CASON, T. (1995): “Cheap Talk Price Signaling in Laboratory Markets,” *Information economics and policy*, 7, 183–204.
- CHARNESS, G., AND M. DUFWENBERG (2006): “Promises and Partnership,” *Econometrica*, 74, 1579–1601.
- CHARNESS, G., AND M. DUFWENBERG (2010): “Bare promises: An experiment,” *Economics Letters*, 107(2), 281 – 283.
- COOPER, D. J., AND K.-U. KUHN (2010): “Communication, Renegotiation, and the Scope for Collusion,” *mimeo*.
- COOPER, R., D. DEJONG, R. FORSYTHE, AND T. ROSS (1992): “Communication in Coordination Games,” *Quarterly Journal of Economics*, 107(2), 739–771.
- CRAWFORD, V. (1998): “A Survey of Experiments on Communication via Cheap Talk,” *Journal of Economic Theory*, 78, 286–298.
- CROSON, R. T. A., A. GOMES, K. L. MCGINN, AND M. NÖTH (2004): “Mergers and Acquisitions: An Experimental Analysis of Synergies, Externalities and Dynamics,” *Review of Finance*, 8(4), 481–514.
- DAVIS, D., AND C. HOLT (1999): “Equilibrium Cooperation in Two-stage Games: Experimental Evidence,” *International Journal of Game Theory*, 28, 89–109.
- DEMICHELI, S., AND J. W. WEIBULL (2008): “Language, Meaning, and Games: A Model of Communication, Coordination, and Evolution,” *American Economic Review*, 98(4), 1292–1311.
- ELLINGSEN, T., AND M. JOHANNESSON (2004): “Promises, Threats and Fairness,” *Economic Journal*, 114, 397–420.

- ELLINGSEN, T., M. JOHANNESSON, J. LILJA, AND H. ZETTERQVIST (2009): “Trust and Truth,” *Economic Journal*, 119, 252–276.
- ELLINGSEN, T., M. JOHANNESSON, S. TJÖTTA, AND G. TORSVIK (2010): “Testing guilt aversion,” *Games and Economic Behavior*, 68(1), 95 – 107.
- ELLINGSEN, T., AND R. ÖSTLING (2010): “When Does Communication Improve Coordination?,” *American Economic Review*, 100(4), 1695–1724.
- ENGEL, C. (2010): “Dictator Games: A Meta Study,” Working Paper Series of the Max Planck Institute for Research on Collective Goods 2010-07, Max Planck Institute for Research on Collective Goods.
- ENGELMANN, D., AND M. STROBEL (2004): “Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments,” *The American Economic Review*, 94, 857–869(13).
- FARRELL, J., AND E. MASKIN (1989): “Renegotiation in Repeated Games,” *Games and Economic Behavior*, 1, 327–360.
- FARRELL, J., AND M. RABIN (1996): “Cheap Talk,” *Journal of Economic Perspectives*, 10, 03–118.
- FEHR, E., AND U. FISCHBACHER (2003): “The nature of human altruism,” *Nature*, 425, 785–791.
- FEHR, E., AND S. GÄCHTER (2000): “Cooperation and Punishment in Public Goods Experiments,” *American Economic Review*, 90(4), 980–994.
- FEHR, E., AND B. ROCKENBACH (2003): “Detrimental effects of sanctions on human altruism,” *Nature*, 422, 137–140.
- FISCHBACHER, U. (2007): “z-Tree: Zurich toolbox for ready-made economic experiments,” *Experimental Economics*, 10(2), 171–178.
- HOLT, C., AND D. DAVIS (1990): “The effects of non-binding price announcements on posted-offer markets,” *Economics letters*, 34, 307–310.
- HOUSER, D., AND E. XIAO (2005): “Emotion expression in human punishment behavior,” *Proceedings of the National Academy of Sciences*, 102, 7398–7401.

- HOUSER, D., E. XIAO, K. MCCABE, AND V. SMITH (2008): “When punishment fails: Research on sanctions, intentions and non-cooperation,” *Games and Economic Behavior*, 62(2), 509–532.
- ISAAC, R., AND C. PLOTT (1981): “The opportunity for conspiracy in restraint of trade: An experimental study,” *Journal of Economic Behavior and Organization*, 2, 1–30.
- ISAAC, R., V. RAMEY, AND A. WILLIAMS (1984): “The effects of market organization on conspiracies in restraint of trade,” *Journal of Economic Behavior and Organization*, 5, 191–222.
- ISAAC, R., AND J. WALKER (1985): “Information and conspiracy in sealed-bid auctions,” *Journal of Economic Behavior and Organization*, 6, 139–159.
- KAGEL, J., AND D. LEVIN (2008): “Auctions: A Survey of Experimental Research 1995-2007,” *Prepared for the Handbook of Experimental Economics, vol 2*.
- LEDYARD, J. (1995): *Public goods: A survey of experimental research* pp. 94–111, In *The handbook of experimental economics*, ed. J. Kagel and A. Roth. Princeton University Press, Princeton, NJ.
- LOPEZ-PEREZ, R., AND M. VORSATZ (2009): “An exploration of the content of social norms using simple games,” Mimeo.
- LUNDQUIST, T., T. ELLINGSEN, E. GRIBBE, AND M. JOHANNESSON (2007): “The cost of lying,” Working Paper Series in Economics and Finance 666, Stockholm School of Economics.
- MOHLIN, E., AND M. JOHANNESSON (2008): “Communication: Content or relationship?,” *Journal of Economic Behavior & Organization*, 65(3-4), 409–419.
- SALLY, D. (1995): “Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992,” *Rationality and Society*, 7, 58–92.
- VAN DAMME, E. (1989): “Renegotiation-proof Equilibria in Repeated Prisoners’ Dilemma,” *Journal of Economic Theory*, 47, 206–217.

VANBERG, C. (2008): “Why Do People Keep Their Promises? An Experimental Test of Two Explanations,” *Econometrica*, 76, 1467–1480.

Tables

Table 1: **Rate of C Actions per Period**

Treatment	Period								Mean
	1	2	3	4	5	6	7	8	
P	0.65	0.61	0.54	0.59	0.52	0.41	0.48	0.35	0.52
PI	0.58	0.62	0.35	0.29	0.37	0.44	0.33	0.33	0.41

Table 2: **Rate of Messages**

Treatment	<i>PR</i>	<i>RR</i>	<i>DC</i>	<i>OM</i>	<i>NM</i>
<i>P</i>	0.15	0.38	0.04	0.04	0.39
<i>PI</i>	0.27	0.25	0.06	0.03	0.38

Table 3: **Fraction of C actions given a certain message in Stage 1**

Treatment	PR_{11}	PR_{10}	PR_{01}	RR_{11}	RR_{10}	RR_{01}	NM_{11}
P	1.00	0.68	0.63	0.74	0.70	0.48	0.14
PI	0.42	0.56	0.45	0.43	0.43	0.40	0.32

Note: The numbers of the two first rows are based on the individual fractions of times the player i chose C , given the specified pattern of messages.

Table 4: **Linear probability model, individual random effects, C Actions in Stage 1**

VARIABLES	(1)	(2)	(3)	(4)	(5)
PI Treatment	-0.108 [0.127]	-0.112 [0.0923]	-0.0929 [0.0704]	0.147* [0.0781]	0.164** [0.0798]
Sent PR		0.234*** [0.0606]	0.247*** [0.0759]	0.375*** [0.0426]	0.431*** [0.0559]
Sent RR		0.208*** [0.0762]	0.175* [0.0907]	0.350*** [0.0686]	0.300*** [0.110]
Sent DC		-0.0155 [0.0738]	0.0480 [0.0803]	-0.0234 [0.120]	-0.00576 [0.0948]
Sent OM		-0.0437 [0.120]	-0.0364 [0.0978]	0.0150 [0.203]	-0.0400 [0.193]
Received PR		0.0981 [0.0598]	0.115* [0.0600]	0.252*** [0.0582]	0.279*** [0.0600]
Received RR		0.0721* [0.0437]	0.0513 [0.0417]	0.159*** [0.0297]	0.138*** [0.0384]
Received DC		-0.116 [0.0855]	-0.0916 [0.0996]	-0.117 [0.138]	-0.122 [0.121]
Received OM		-0.109 [0.0976]	-0.0855 [0.0723]	0.0152 [0.0855]	-0.0451** [0.0215]
Sent PR×PI				-0.244*** [0.0598]	-0.314*** [0.0638]
Sent RR×PI				-0.289*** [0.0886]	-0.262* [0.136]
Sent DC×PI				-0.0173 [0.141]	0.0597 [0.101]
Sent OM×PI				-0.103 [0.216]	0.0259 [0.207]
Received PR×PI				-0.252*** [0.0728]	-0.275*** [0.0729]
Received RR×PI				-0.173*** [0.0505]	-0.178*** [0.0460]
Received DC×PI				-0.0154 [0.163]	0.0471 [0.163]
Received OM×PI				-0.225 [0.144]	-0.0603 [0.142]
Period	-0.0360*** [0.00628]		-0.0238*** [0.00581]		-0.0270*** [0.00583]
History 1			0.318*** [0.118]		0.293** [0.130]
History 2			-0.0443 [0.193]		-0.00858 [0.207]
Constant	0.681*** [0.0859]	0.374*** [0.0962]	0.305*** [0.0446]	0.236*** [0.0529]	0.189*** [0.0578]
Observations	784	784	686	784	686
Number of id	98	98	98	98	98

Notes: Robust standard errors clustered on session level in brackets. Period indicates period 2, 3, ..., 8. (Period one data excluded due to the history variables.). History 1 describes the frequency of C actions of the opponents up to the current period. History 2 describes the frequency

Table 5: Fractions of H actions in Stage 2

Outcome in stage 1	All		PR		RR	
	P	PI	P	PI	P	PI
Loser						
<i>fraction of H</i>	0.79	0.57	0.50	0.40	0.89	0.68
<i>total # obs</i>	63	89	16	43	37	41
Winner						
<i>fraction of H</i>	0.87	0.78	0.63	0.65	0.95	0.85
<i>total # obs</i>	63	89	16	43	37	41
Mutual defection						
<i>fraction of H</i>	0.82	0.85	0.75	0.82	0.85	0.88
<i>total # obs</i>	114	156	8	60	34	58
Mutual cooperation						
<i>fraction of H</i>	0.99	0.99	1.00	1.00	1.00	1.00
<i>total # obs</i>	128	82	56	46	98	42

Note: PR (RR) means at least one sent PR (RR) message before stage 1. Behavior from interactions where one subject sent a PR message and the other an RR message, hence enters both columns.

Table 6: Intra-play messages in the PI treatment

Stage 1 outcome	<i>No Message</i>		<i>H Message</i>		<i>L Message</i>	
	#	%	#	%	#	%
Looser	57	64.0	27	30.3	5	5.6
Winner	52	58.4	34	38.2	3	3.4
Mutual Defection	78	50.0	70	44.9	8	5.1
Mutual Cooperation	46	56.1	36	43.9	0	0
Total	233	56.0	167	40.1	16	3.9

A7: PBE outcomes in a mixed population of E, R and S types.

Type	P-treatment				PI-treatment					
	M1	M2	S1	S2	M1	M2	S1	M3	M4	S2
E, E	RR	RR	C, C	H, H	RR	RR	C, C	H	H	H, H
S, S	PR	PR	C, C	H, H	PR	PR	D, D	H	H	H, H
R, R	PR	PR	C, C	H, H	PR	PR	C, C	H	H	H, H
E, R	RR	PR	C, C	H, H	RR	PR	C, C	H	H	H, H
R, E	PR	RR	C, C	H, H	PR	RR	C, C	H	H	H, H
E, S	RR	PR	C, D	H, H	RR	PR	C, D	H	H	H, H
S, E	PR	RR	D, C	H, H	PR	RR	D, C	H	H	H, H
S, R	PR	PR	C, C	H, H	PR	PR	D, C	H	L	L, L
R, S	PR	PR	C, C	H, H	PR	PR	C, D	L	H	H, H

Note: The first column displays which types are matched and the order in which they send messages. Columns marked M1 and M2 describe the pre-play messages of the first player and second player, respectively. Columns marked M3 and M4 state the intra-play messages of the first and second player, respectively. S1 and S2 display the actions of the two players in stage 1 and stage 2 (in the order in which they send messages).

B 8: Linear probability model, C Actions in Stage 1

VARIABLES	(1)	(2)	(3)	(4)	(5)
PI treatment	-0.108 [0.127]	-0.110 [0.0821]	-0.0850 [0.0735]	0.155* [0.0746]	0.170* [0.0775]
Sent PR		0.329*** [0.0670]	0.312*** [0.0806]	0.501*** [0.0283]	0.506*** [0.0584]
Sent RR		0.315*** [0.0733]	0.270** [0.0878]	0.422*** [0.0753]	0.366** [0.109]
Sent DC		-0.0683 [0.0905]	0.0102 [0.0942]	-0.104 [0.129]	-0.0384 [0.108]
Sent OM		-0.0278 [0.151]	-0.0169 [0.116]	0.126 [0.219]	0.0940 [0.212]
Received PR		0.124* [0.0639]	0.129* [0.0600]	0.285*** [0.0630]	0.291*** [0.0551]
Received RR		0.102* [0.0477]	0.0778 [0.0492]	0.171*** [0.0467]	0.137* [0.0612]
Received DC		-0.0619 [0.0809]	-0.0490 [0.0959]	-0.0753 [0.140]	-0.0732 [0.139]
Received OM		-0.145* [0.0728]	-0.145* [0.0715]	-0.120 [0.0741]	-0.178** [0.0716]
Sent PR×PI				-0.312*** [0.0465]	-0.348*** [0.0668]
Sent RR×PI				-0.244** [0.0964]	-0.219 [0.121]
Sent DC×PI				0.0192 [0.144]	0.0510 [0.118]
Sent OM×PI				-0.296 [0.245]	-0.190 [0.238]
Received PR×PI				-0.289*** [0.0746]	-0.303*** [0.0638]
Received RR×PI				-0.170** [0.0685]	-0.153* [0.0756]
Received DC×PI				-0.0319 [0.166]	-0.00359 [0.185]
Received OM×PI				-0.0582 [0.112]	0.0441 [0.145]
Period			-0.0237*** [0.00665]		-0.0275*** [0.00708]
History 1			0.404** [0.164]		0.380* [0.176]
History 2			-0.396 [0.316]		-0.400 [0.332]
Constant	0.519*** [0.103]	0.305*** [0.0863]	0.321*** [0.0637]	0.182** [0.0537]	0.238*** [0.0639]
Observations	784	784	686	784	686
R-squared	0.012	0.154	0.206	0.186	0.244

Notes: Robust standard errors clustered on session level in brackets. Period indicates period 2, 3, ..., 8. (Period one data excluded due to the history variables.). History 1 describes the frequency of C actions of the opponents up to the current period. History 2 describes the frequency of H actions of the opponents up to the current period. *** indicates significance at the 1% level

B 9: Linear probability model with individual fixed effects, C Actions in Stage 1

VARIABLES	(1)	(2)	(3)	(4)
Sent PR	0.228*** [0.0615]	0.241*** [0.0761]	0.296*** [0.0553]	0.384*** [0.0662]
Sent RR	0.211*** [0.0780]	0.176* [0.0922]	0.305*** [0.0698]	0.252* [0.113]
Sent DC	-0.0171 [0.0693]	0.0476 [0.0740]	0.00414 [0.107]	0.00171 [0.0934]
Sent OM	-0.0422 [0.117]	-0.0377 [0.0959]	-0.0391 [0.186]	-0.104 [0.165]
Received PR	0.0945 [0.0603]	0.112* [0.0600]	0.233*** [0.0573]	0.280*** [0.0627]
Received RR	0.0744* [0.0427]	0.0534 [0.0416]	0.146*** [0.0261]	0.141*** [0.0367]
Received DC	-0.119 [0.0823]	-0.0936 [0.0959]	-0.141 [0.141]	-0.143 [0.115]
Received OM	-0.107 [0.0964]	-0.0853 [0.0730]	0.0673 [0.0927]	0.00458 [0.0208]
Sent PR×PI			-0.193** [0.0733]	-0.290*** [0.0806]
Sent RR×PI			-0.301** [0.0921]	-0.279 [0.150]
Sent DC×PI			-0.0349 [0.139]	0.0562 [0.100]
Sent OM×PI			-0.0245 [0.199]	0.0989 [0.179]
Received PR×PI			-0.233** [0.0721]	-0.268** [0.0768]
Received RR×PI			-0.165** [0.0501]	-0.189*** [0.0440]
Received DC×PI			0.000325 [0.171]	0.0800 [0.167]
Received OM×PI			-0.286 [0.165]	-0.0848 [0.154]
Period		-0.0233*** [0.00579]		-0.0266*** [0.00550]
History 1		0.340*** [0.116]		0.167 [0.138]
History 2		-0.0674 [0.190]		0.385 [0.258]
Constant	0.315*** [0.0566]	0.251*** [0.0455]	0.343*** [0.0258]	0.230** [0.0702]
Observations	784	686	784	686
R-squared	0.055	0.111	0.087	0.163
Number of id	98	98	98	98

Notes: Robust standard errors clustered on session level in brackets. Period indicates period 2, 3, ..., 8. (Period one data excluded due to the history variables.). History 1 describes the frequency of C actions of the opponents up to the current period. History 2 describes the frequency

**B 10: Linear probability model with individual random effects, H
Actions in Stage 2**

VARIABLES	(1)	(2)	(3)	(4)	(5)
Winner	0.246*** [0.0682]	0.240*** [0.0643]	0.233*** [0.0632]	0.236*** [0.0612]	0.235*** [0.0837]
Mutual Cooperation	0.381*** [0.0404]	0.384*** [0.0564]	0.348*** [0.0529]	0.351*** [0.0524]	0.311*** [0.0645]
Mut Defection	0.259*** [0.0420]	0.240*** [0.0418]	0.243*** [0.0603]	0.218*** [0.0698]	0.218*** [0.0660]
Sent PR		-0.129** [0.0523]	-0.0799** [0.0353]	-0.0816** [0.0355]	-0.0782** [0.0386]
Sent RR		0.0184 [0.0291]	0.0339 [0.0487]	0.0333 [0.0481]	0.0388 [0.0450]
Received PR		-0.0846** [0.0426]	-0.0561 [0.0396]	-0.0799* [0.0449]	-0.107*** [0.0241]
Received RR		-0.0243 [0.0492]	-0.0260 [0.0385]	-0.0429 [0.0436]	-0.0230 [0.0378]
Sent L			-0.496*** [0.172]	-0.493*** [0.172]	-0.486*** [0.174]
Received L			-0.238* [0.139]	-0.229 [0.140]	-0.222 [0.142]
Sent H			0.106*** [0.0352]	0.109*** [0.0376]	0.102*** [0.0361]
Received H			0.0279 [0.0269]	0.0228 [0.0265]	0.0210 [0.0301]
Received PR×Loser					-0.0196 [0.147]
Received PR×Mutual Defection				0.0651*** [0.0207]	0.0870* [0.0487]
Received PR×Mutual Cooperation					0.115 [0.0836]
Received RR×Mutual Defection				0.0436 [0.0446]	
Constant	0.578*** [0.0353]	0.646*** [0.0436]	0.603*** [0.0530]	0.614*** [0.0557]	0.620*** [0.0683]
Observations	416	416	416	416	416
Number of id	52	52	52	52	52

Notes: Robust standard errors clustered on session level in brackets. *** indicates significance at the 1 % level. ** indicates significance at the 5 % level.

**B 11: Linear probability model with individual random effects, H
Messages in Stage 2**

VARIABLES	(1)	(2)	(3)	(4)
Winner	0.0859 [0.0750]	0.0724 [0.0597]	0.179*** [0.0271]	0.0620 [0.0954]
Mutual Cooperation	0.114*** [0.0399]	0.109** [0.0459]	0.0744 [0.0605]	0.132*** [0.0331]
Mutual Defection	0.0773* [0.0424]	0.0600* [0.0318]	0.105 [0.0692]	0.0739** [0.0352]
Sent PR		-0.0714 [0.112]	-0.0662 [0.104]	-0.0697 [0.111]
Sent RR		-0.00338 [0.134]	-0.000489 [0.130]	0.00467 [0.134]
Sent DC		-0.0557 [0.119]	-0.0384 [0.104]	-0.0507 [0.120]
Received PR		0.0332 [0.0537]	-0.178*** [0.0321]	0.0341 [0.0541]
Received RR		-0.0156 [0.0454]	-0.0171 [0.0417]	0.0413 [0.0721]
Received DC		0.173*** [0.0332]	0.163*** [0.0225]	0.171*** [0.0307]
Received PR_Sucker			0.338*** [0.0715]	
Received PR×Mutual Defection			0.148 [0.157]	
Received PR×Mutual Cooperation			0.414*** [0.0905]	
Received RR×Loser				-0.0262 [0.133]
Received RR×Mutual Defection				-0.0815 [0.0802]
Received RR×Mutual Cooperation				-0.105 [0.0938]
Constant	0.332*** [0.0579]	0.350*** [0.0935]	0.317*** [0.0714]	0.340*** [0.0991]
Observations	416	416	416	416
Number of id	52	52	52	52

Notes: Robust standard errors clustered on session level in brackets. *** indicates significance at the 1 % level. ** indicates significance at the 5 % level.