Whom are you talking with? An experiment on credibility and communication structure

Gilles Grandjean[‡]   Marco Mantovani[*†]   Ana Mauleon [‡§]
Vincent Vannetelbosch [‡§]

September 2014

Abstract

The paper analyzes the role of the structure of communication - i.e. who is talking with whom - on the choice of messages, on their credibility and on actual play. We run an experiment in a three-player coordination game with Pareto ranked equilibria, where a pair of agents has a profitable joint deviation from the Pareto-dominant equilibrium. According to our analysis of credibility, the subjects should communicate and play the Pareto optimal equilibrium only when communication is public. When pairs of agents exchange messages privately, the players should play the Pareto dominated equilibrium and disregard communication. The experimental data conform to our predictions: the agents reach the Pareto-dominant equilibrium only when announcing to play it is credible. When private communication is allowed, lying is prevalent, and players converge to the Pareto-dominated equilibrium. Nevertheless, at the individual level, players' beliefs and choices tend to react to messages even when these are non-credible.

Keywords: cheap talk, coordination, coalitions, experiment.
JEL codes: C72, C91, D03, D83.

[*]Corresponding author, e-mail address: `marco.mantovani@unimib.it`
[†]Department of economics, University of Milan Bicocca.
[‡]CEREC, Université Saint-Louis - Bruxelles.
[§]CORE, Université catholique de Louvain.

1   Introduction

This paper investigates in a laboratory experiment whether pre-play communication helps achieving the efficient equilibrium in a coordination game among three players. The game under study (the coalitional prisoner's dilemma, see Figure 1) has two Pareto-ranked pure strategy Nash equilibria, and each pair of players has a profitable joint deviation from the Pareto-dominant equilibrium. In particular, we study how the structure of communication, i.e. who is talking with whom, influences the choice of messages, the beliefs and the choice of the players in this game.

Communication may help the players to coordinate on a Nash equilibrium. Suppose players meet before the game and engage in pre-play communication. If an agreement is reached in the meeting, it has to be a Nash equilibrium. Furthermore, if the game features Pareto-ranked Nash equilibria, the common perception is that communication helps achieving the Pareto-dominant one. Experiments in 2 by 2 stag hunt games have shown that communication indeed leads to higher rate of coordination on the efficient equilibrium [e.g. Charness, 2000; Cooper et al., 1992]. In this paper, we analyze whether the structure of communication affects coordination and efficiency.

In the experiment, the subjects play eight rounds of the coalitional prisoner's dilemma, under *strangers* matching. We implement a *between-subjects* design, with four treatments. The treatments differ in the structure of pre-play communication. Communication takes the form of structured messages, by which agents reveal their intended action.

In a baseline treatment, players do not communicate (*NoCom*). Each of the other treatments implements one of the possible symmetric communication structures. These are: i) *Public*: each player sends a public message to both of the other group members; ii) *Private*: each player sends a private message to each of the other group members; iii) *Both*: each player sends both the private and the public messages to the other group members.

Theoretical models on pre-play communication in games of complete in-

formation focus on 2-player games.[1,2] These papers assume that messages pointing at a Nash equilibrium are credible because they are self-committing: if believed, they create the incentive to play according to them. Grandjean and Mantovani [2014] have extended this notion of credibility to games with more than two players, and arbitrary communication structures. A profile of messages in a group of communicating players is self-committing if each player in the group is better-off by playing his message when he trusts the messages revealed by the other group members. A profile of messages in a group is credible if it is self-committing, and no group member can improve on it by following a self-committing agreement of another group. The structure of communication thus plays a key role in the determination of the credibility of agreements. Following this approach, we characterize the credible equilibria of the game, i.e. the perfect Bayesian equilibria of the game where it is assumed that agents trust credible messages and use the messages they observe to update their beliefs about the messages they do not observe.

In *Public*, an agreement to play the Pareto-dominant equilibrium is credible. The only credible equilibrium is such that each player reveals his intention to play the action associated to the Pareto-dominant equilibrium, and plays that action if each player has also communicated in this way. In *Private* and *Both*, an agreement to play the action associated to the Pareto-dominant equilibrium is not credible since any pair of agents could jointly deviate from it by coordinating privately. In these treatments, the players choose the action associated to the Pareto dominated equilibrium after every communication history in a credible equilibrium. In that sense, communication is uninformative and should be disregarded.

Our results show that the communication structure has a huge impact on play. Absent communication, miscoordination is initially high and play quickly converges to the dominated equilibrium. Coordination on pure strategy Nash equilibria is higher in the first rounds with communication. In *public*, the efficient equilibrium is played by more than half of the groups, steadily across rounds. A vast majority of the players announce their intention to play

---

[1]See Farrell [1988, 1995], Rabin [1994] and Farrell and Rabin [1996]. A review of the theoretical literature can be found in Crawford [1998].

[2]Here we focus on the communication of intentions of play. A different, though related, literature addresses communication about the state of the world in sender-receiver games with incomplete information [see Blume et al., 1993; Crawford and Sobel, 1982]. In particular, a number of studies have investigated information revelation under different communication structures. Farrell and Gibbons [1989] compared private and public communication in sender-receiver games with multiple audiences. Battaglini and Makarov [2014] provided an experimental test of the model.

the action corresponding to the Pareto-dominant equilibrium, and play accordingly only when others do so as well. When private communication is allowed, the dominated equilibrium ends up being played by most groups in the last rounds. Lying is prevalent: players try to convince one of their partners to play the Pareto-dominant equilibrium, while agreeing on a deviation with the other. This strategy is frequently successful because some agents trust messages that are not credible, especially in the earlier rounds. As a result, and contrary to the theoretical hypothesis, the outcome of communication affects beliefs even under *Private* and *Both*, where messages should be considered uninformative.

In two-player games, the only possible variation to the structure of communication distinguishes one-sided and two-sided communication. Cooper et al. [1992, 1989] found that one-sided communication is more effective in the battle of the sexes, where some symmetry-breaking device is needed, while two-sided is more effective in the stag hunt.[3,4]

Relatively few papers have studied the role of cheap talk in games with more than two-players.[5] Blume and Ortmann [2007] found that communication improves coordination and efficiency in the weak-link and the median game, which both feature Pareto-ranked equilibria and perfectly aligned interests. In Moreno and Wooders [1998], two players with common interests and opposed to those of the third one communicate to coordinate and correlate their actions. Castillo and Leo [2007] allowed the players to chat freely both publicly and privately. Some of their subjects use communication to organize deviations, betting on the partial sophistication of their partners. Choi and Lee [2014] studied a multi-player battle of the sexes, where structured communication takes place on different network structures. They showed that the level of coordination and the equilibria that are reached depend on the network structure. In their design, the players cannot send different messages to different audiences.

The paper is organized as follows. Section 2 presents the game and the theoretical framework. The experimental design and procedures follow in Section 3. Section 4 presents the results. Section 5 concludes.

---

[3]Related to this result, it has been pointed out that, in situations where agents face strategic risk, mutual communication of intentions may have a reassuring effect [see Brandts et al., 2012; Charness and Dufwenberg, 2006, 2011].

[4]Andersson and Wengström [2012] allow for intra-game communication, and show how this can work, through renegotiation, against efficient coordination in dynamic settings.

[5]Crawford [1998] surveyed the early literature on two-player games. A more recent one can be found in Charness and Dufwenberg [2007]

2 Credibility in the coalitional prisoner's dilemma

2.1 The game

FIGURE 1: THE COALITIONAL PRISONER'S DILEMMA.

	X	Y
X	8,8,8	0,0,0
Y	0,0,0	16,16,0

X

	X	Y
X	0,0,0	0,16,16
Y	16,0,16	4,4,4

Y

The game we investigate is depicted in Figure 1. The set of players is $N = \{i,j,k\}$, and the action space of any player $l \in N$ is $A_l = \{X,Y\}$. We denote by x, y the message announcing a player's intention to play X and Y. Let XYY be a shortcut notation for the profiles of actions where one player chooses X, and the others choose Y, and let XXX, XXY, and YYY be defined accordingly.

There are two pure strategy Nash equilibria: XXX and YYY. The game belongs to the class of coordination games with Pareto-ranked equilibria and XXX is the Pareto-dominant equilibrium (PDE). There also exists a mixed strategy Nash equilibrium where all players play X with probability $\frac{1+\sqrt{2}}{3} \cong$ .805, leading to an expected payoff of 5.18.

The structure of the game is similar to the example introduced in Bernheim et al. [1987] to motivate the notion of coalition-proof Nash equilibrium. In particular, the agreement to play XXX is not immune to the deviation YY by any two players, which is self-enforceable, since the agents who have initiated the deviation do not want to further deviate. Following the same argument, the deviation from YYY to XXX is not self-enforceable and YYY is the unique coalition-proof Nash equilibrium of the game.[6]

When groups of two agents may coordinate their moves, agreeing on XY or YX does not make sense since the agent playing X is sure of getting his worst payoff. Among the other possible strategies, the two agents are strictly better off by playing YY over XX. If the third agent anticipates this, he should play Y as well. Iterated elimination of 'coalitionally dominated strategies' leads to the selection of the Pareto-dominated equilibrium.[7] Because this feature recalls the prisoner's dilemma, except for defection being defined over

---

[6]Notably, Bernheim et al. [1987] explicitly mention "unconstrained communication" in their justification of the concept. Milgrom and Roberts [1996] proposed a form of (correlated) coalition-proof equilibrium where the set of feasible coalitions can be arbitrarily restricted.

[7]See Ambrus [2006] for a noncooperative solution concept based on this procedure.

coalitions rather than individuals, we label this game the *coalitional prisoner's dilemma* (CPD).

2.2 Credible communication

Grandjean and Mantovani [2014] analyzed the credibility of pre-play communication in games of complete information. They extended the notion of self-committing messages to n-player games, and arbitrary communication structures. Building on this, they proposed a notion of credible communication that is sensible to the communication structure, and a related solution concept – the credible equilibrium. We here describe informally how these concepts apply to our game, and use them to derive the experimental hypotheses. A formal account can be found in Appendix B.

A communication structure is a collection of groups of players. In each group, agents simultaneously announce their intended action to the other group members. We focus on the symmetric communication structures among three players: (i) no communication; (ii) (public) communication in the group of all players; (iii) (private) communication in each group of two players; (iv) communication in the group of all players, and in each group of two players.

Within each group, a message profile is *self-committing* if each player is better off by playing his message when he trusts the messages of the other group members, *for any belief he may hold about nonmembers*.[8] In groups of size two, the only message profile that is self-committing is (y,y). Only in that case agents prefer to stick to their announcement if they believe the other's message, whatever the choice of the third player. In the group of all players, the two message profiles corresponding to the pure strategy Nash equilibria of the game are self-committing.

In a group, a message profile that is not self-committing leads to a consensus to play an action profile other than the message profile if (i) the players who do not play according to their message choose a best-response to the message of those who stick to it, (ii) the message of the players who stick to it is a best-response to the play of the consensus, and (iii) there is no other action profile satisfying (i) and (ii). In a group of two players, the message profile

---

[8]Aumann [1990] noted that a player may have incentives to send a self-committing message even if he is planning to play something else. In that case, the message is not self-signaling and should not be credible. In our game, this tension does not occur. A player who convinces another to play Y would not do so if he intended to play X. Similarly in public, a player who is planning to play Y should not try to convince the others to play X. See Charness [2000] and Clark et al. [2001] for experimental tests of Aumann's conjecture.

5

where one player reveals his intention to play Y while the other player announces his plan to play X is not self-committing, but it leads to a consensus where the two players choose Y. The x-sender should play Y if he believes the y-sender (part (i)). Given this reaction, the y-sender should play his message (part (ii)). If the y-sender believes the x-sender, he will stick to his message for some beliefs about the play of the third player (part (iii)). In the group of three players, if two players communicate x and the remaining player communicates y, players would agree on the play of the PDE if the message of the x-senders is believed and the message of the y-sender is not. They would also agree on the dominated equilibrium if they believe the message y. Thus, no consensus is reached following this message profile since two action profiles satisfy part (i) and (ii) of the definition. If two players communicate y and one communicates x in the group of three players, a consensus to play the dominated equilibrium is reached. If the players believe at least one message y, they should play Y. On the other hand, if the y-senders only believe the x-sender, they could take both action, depending on their belief about the play of the third player.

A self-committing messages profile, or a consensus, is called an agreement. Agreements within a group are not credible per se, as one wants them to be robust to messages that are exchanged in other groups. An agreement is *credible* in a group if there are no members of that group that could gain by following the agreement reached in another group they also belong to. If players can communicate only in public or only in private, all agreements are credible. When agents may communicate both in public and in private, the public agreement (x, x, x) is not credible because any pair of agents could agree on Y by communicating privately. Agreements to play Y are, instead, credible.

A *credible equilibrium* is a perfect Bayesian equilibrium of the game with pre-play communication where agents condition their choices on the messages observed, *given that credible message profiles are believed*.

If there is only public communication, in a credible equilibrium the players play the PDE after observing (x, x, x) and play Y if at least two y are sent. If they observe only one y message, no agreement is reached and the players could play any equilibrium of the stage game. There are credible equilibria where agents send a message y with positive probability. However, communicating y is weakly dominated. If communicating y matters for the selection of the equilibrium, it leads to an equilibrium with a smaller payoff to each player. This occurs when the two other players reveal their intention to play X while players do not play the PDE following xxy. It is also the case when

one other player communicates y and players do not play the dominated equilibrium following xxy. In the other cases, the choice of the message of a player does not affect the outcome of the game.[9]

If there is only private communication, the play of an agent depends on the two messages he sends and on the two messages he receives. As there is a credible agreement in any couple where at least one y message has been sent, the players choose Y after observing at least one y message. After observing only x messages, a player updates his beliefs about the likelihood that the two other agents have exchanged at least one y message, and chooses a best-response. As shown in Appendix B, there does not exist a symmetric credible equilibrium such that agents play X with positive probability. Intuitively, for agents to play X with positive probabilities at a credible equilibrium, it must be that they send message x to both partners with high probability. If it were not the case, an agent observing only x messages would correctly anticipate that the others have exchanged at least a y with high probability, and his best-response would be to play Y. However, if the agents sent message x to both partners with high probability, and responded to the observation of only x messages by playing X, then an agent would have a profitable deviation in the communication stage by revealing the intention to play y to one of the others.

If there are both private and public communication, we have seen that agents cannot credibly reveal their intention to play X. An agent plays Y after any public or private credible agreement to do so. When an agent observes at most one public y message, and only private x message, he chooses an optimal action given his equilibrium beliefs. With an argument identical to the one used for private communication alone, we can show that the only symmetric credible equilibria of the game are such that all players choose Y with probability one, independently of the messages (see Appendix B).

As credible equilibria are babbling whenever private communication is present, any communication strategy can be part of an equilibrium. However, under our notion of credibility, communicating y to both of the other players, either with a public message or with two private ones, is weakly dominated: it is always weakly better to convince only one other player to play Y, using private messages.

By following the literature [e.g. Rabin, 1994], and assuming that multiple

---

[9]This occurs if the two others communicate y, if they communicate x and the credible equilibrium following xxy is such that the PDE is played, or if one communicates x, the other y, and the credible equilibrium following xxy is such that the dominated equilibrium is played.

stages are used to negotiate agreements, but that play is conditioned on the messages observed in the last stage of communication, our results carry over to the case of multiple stages of simultaneous communication.

We derive the following hypotheses from the theoretical analysis.

Hypothesis 1. *(i) If only public communication is allowed, the players always play X after observing (x, x, x), Y if at least two players send y. (ii) If private communication is allowed, the players play Y irrespective of the observed messages.*

Hypothesis 2. *(i) Players communicate x publicly. ii) Players do not communicate y to both partners privately.*[10]

These hypotheses imply the following comparative statics across communication structures. Coordination on pure strategy Nash equilibria is higher if communication is allowed with respect to no communication. Coordination on the PDE is higher when only public communication is allowed than under any other communication structure. Finally, public communication is also the only structure where communication affects beliefs.

3   Experimental design and procedures

We run four treatments. In each treatment, the subjects played eight rounds of the CPD depicted in Figure 1. The payoffs, in experimental currency units (ECU), were rescaled by a factor of ten. The treatment variable was the structure of pre-play communication. Each treatment implemented one of the four symmetric communication structures described in Section 2. Each subject took part only in one treatment.

In a baseline condition (*NoCom*), there was no pre-play communication. In *Public*, each player sent a single message to both of the other group members. In *Private*, each player sent a message to each of the other group members. In *Both*, each player sent a message to both, and a message to each of the other group members. The subjects had to send a message when given the opportunity – i.e. they could not opt out from communication.

When pre-play communication was allowed, each round included four communication stages. In every stage, the players exchanged structured messages. Each message was in the form "*I intend to choose action X (Y)*". All messages were received simultaneously. At the end of the communication phase,

---

[10]We note that sending y to everybody is no longer weakly dominated if the players care about promise keeping [see Ellingsen and Johannesson, 2004; Gneezy, 2005].

the subjects had to choose a pure action between *X* and *Y*. Agents could see the exchanged messages in that round when choosing a message or an action.

We elicited first order beliefs, asking the following question: *"Co-player 1 and Co-player 2 will choose between X and Y. What do you think they are going to choose? Please enter a number between 0 and 100, representing the probability that Co-Player 1(2) chooses X."*.

After answering, they were asked the following question, eliciting second order beliefs: *"Co-player 1 and Co-player 2 also answered the same question as you did. Regarding you, what number do you think they entered? Please enter a number between 0 and 100, representing the probability with which Co-player 1(2) thinks you are choosing X."* Thus, we elicit probabilistic first order beliefs, and point second order beliefs.[11]

Beliefs were incentivized using a quadratic scoring rule.[12] Let b_{ij}^{1st} and b_{ij}^{2nd} be the first and second order belief stated by player *i* relative to player *j*, and taking value in $[0,1]$. The chosen action of player *j* is $a_j \in \{0,1\}$, where $a_j = 0$ indicates that player *j*'s choice is *Y*, while $a_j = 1$ indicates that player *j*'s choice is *X*. The payoff for the belief tasks, expressed in ECU, were: $25 \cdot (1 - (a_j - b_{ij}^{1st})^2)$ and $25 \cdot (1 - (b_{ji}^{1st} - b_{ij}^{2nd})^2)$, for the first and second order beliefs, respectively.[13]

In *NoCom*, the belief tasks were taken after the action phase, before receiving feedback on the game. In all communication treatments we elicited the beliefs twice. The subjects took the tasks once before the communication phase, and a second time after they had chosen an action, before receiving feedback on the game. As the beliefs were stated before and after communication, we will refer to the first as prior beliefs, to the second as posterior beliefs. In total the players completed four belief tasks in each round under *NoCom*, eight tasks in the other treatments.

Summing up, the players stated their prior beliefs, communicated for four stages, chose their action, stated their posterior beliefs, and received feedback on the game and on the belief tasks. Only the last three steps were implemented in *NoCom*.

---

[11]Second order beliefs are expressed as probabilities, but not as a probability distribution over probabilities.

[12]As shown by Savage [1971] the quadratic scoring rule is proper, meaning that the expected utility is uniquely maximized when the stated probabilities are equal to the true probabilities, under some restrictions on the players' preferences.

[13]The subjects were not provided with these formulas, but with tables reporting the possible payoffs from different combinations of choices and stated beliefs. A copy of the tables can be found in Appendix A.

The players repeated this procedure eight times. We refer to these repetitions as rounds. We used a constrained stranger matching. The subjects were informed that they were being re-matched in every round with a new group of players. They were guaranteed they would not play twice in the same group with the same two co-players. We formed matching group of six people (at random). In every new round the players were re-matched with players in their own matching group. The sequence of matching was predetermined to ensure that no group appeared twice in the sequence. The assignments of subjects to terminals, of terminals to matching groups, and of labels within the matching group were random. As a consequence, so was the sequence of groups for each individual subject.

This procedure allowed us to collect one independent observation every six subjects, while preventing repeated game effects. We point out that all the information given to the subjects was correct.

The computerized experiment was run at the WZB-TU Experimental Lab in Berlin, in June 2013, and involved 210 subjects, distributed over 12 experimental sessions. Fifty-four subjects took part in *NoCom*, *Private* and *Both*; forty-eight subjects took part in *Public*.[14] Sessions took on average 45 minutes. The computerized program was developed using Z-tree [Fischbacher, 2007]. Table 1 summarizes sessions' details.

All sessions followed an identical procedure. After subjects were allowed to enter into the lab, instructions were read aloud.[15] Participants were asked to fill in a control questionnaire. The experiment started only when all the subjects had correctly completed the task.

After completing the eight rounds, subjects were asked to fill in a questionnaire. We gathered qualitative information about the expectations from the game, the opponent, and the strategy followed. We recorded their assessment on the credibility of communication, and its role in the game. Finally, we elicited self-reported quantitative measures of trust and risk preferences, using the questions of the SOEP German Panel.[16]

Subjects were told they would have been paid according to the ECU earned in one stage game, and in one belief task, selected at random by the computer. At the end of each round the computer selected one of the belief tasks to be relevant for payment in case that round was selected. After the last round the

---

[14]As some subjects did not show up in one session, we were forced to give up one matching group.

[15]The experiment was conducted in German. An English version of the experimental instructions is available in Appendix A.

[16]For the use of the risk questions to measure risk preferences, see Dohmen et al. [2011]

| Session	Participants	Matching groups (Ind. Obs)	Treatment
1	18	3	*Public*
2	18	3	*Private*
3	18	3	*Both*
4	18	3	*Private*
5	18	3	*Both*
6	12	2	*Public*
7	18	3	*Private*
8	18	3	*Both*
9	18	3	*NoCom*
10	18	3	*NoCom*
11	18	3	*NoCom*
12	18	3	*Public*

computer drew two numbers for each individual, between 1 and 8, without replacement. The stage game and the relevant belief task in the corresponding rounds were paid to the subjects, according to the exchange rate: $10ECU = 1€$. Thus subjects could earn between 0 and 16 Euros from the stage game, and between 0 and 2.50 Euros from the belief task. The subjects knew they could not be paid for the stage game and the belief task from the same round. This prevents them from hedging their action using the belief tasks. The average payment, including 5 Euros of show-up fee, was around 13.70 Euros.

4   Results

4.1   Aggregate outcomes

Figure 2 shows the aggregate outcomes for each treatment. Subjects coordinate more than half of the time on the PDE in *Public*, whereas they virtually never play it in the other treatments. On the other hand, the dominated equilibrium is played 23 percent of the time in *Public*, but more than half of the time in the other treatments. A Komogorov-Smirnov test (KS) confirms that the distribution of aggregate outcomes is different only when comparing *Public* with each of the other treatments. A Wilkoxon rank-sum test (WRS) con-

FIGURE 2: AGGREGATE OUTCOMES
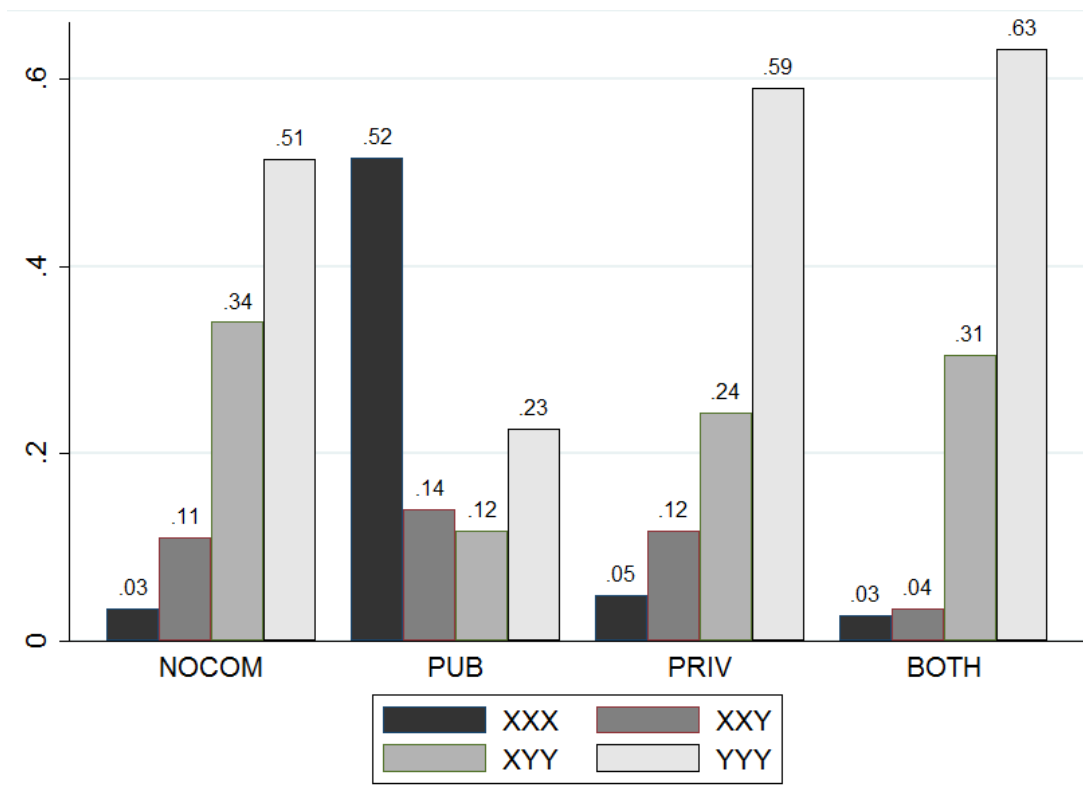
firms that the fraction groups playing the PDE is higher in *Public* than in any other treatment (see Table 2).[17,18]

Coordination on pure strategy Nash equilibria is higher when subjects communicate. The fraction of equilibrium outcomes ranges from 75 percent in *Public* to 55 percent in *NoCom*. Using a WRS test, the fraction of equilibrium outcomes in *NoCom* is significantly different than in *Public*, *Both*, and the aggregation of communication treatments. Coordination rates in any two communication treatments are not significantly different (see Table 3).

In Figure 3 the outcomes are decomposed in initial (1-3), intermediate (4-6) and final rounds.[19] In *NoCom*, *Private* and *Both*, coordination on the dominated equilibrium increases over rounds, peaking above 80 percent in the final rounds. In *Public*, the proportion of groups playing the PDE slightly increases

---

[17]All the tests that will be shown are based on one independent observation for each matching group.

[18]The variable 'outcome' is categorical. The variable for 'outcome PDE' is binary. However, the tests are run on averages over rounds for each matching group. These averages are continuous, which makes the reported tests applicable.

[19]We use this decomposition to disentangle first responses from limiting outcomes.

TABLE 2: AGGREGATE OUTCOMES ACROSS TREATMENTS

	Overall distr. KS test (P-val)			Frac. *XXX* WRS test (P-val)		
	PRIV	**BOTH**	**NOCOM**	**PRIV**	**BOTH**	**NOCOM**
PUB	0.778 (.00)	1.000 (.00)	0.889 (.00)	3.518 (.00)	3.552 (.00)	3.552 (.00)
PRIV		.222 (.96)	0.333 (.59)		0.607 (.54)	−0.506 (.62)
BOTH			0.222 (.96)			0.053 (.96)

TABLE 3: COORDINATION ACROSS TREATMENTS

	Rounds 1-8 WRS test (P-val)			Rounds 7-8 WRS test (P-val)		
	PRIV	**BOTH**	**NOCOM**	**PRIV**	**BOTH**	**NOCOM**
PUB	1.27 (.20)	1.10 (.27)	**−2.79 (.01)**	−0.76 (.45)	−1.40 (.16)	0.90 (.37)
PRIV		−0.10 (.92)	−1.59 (.11)		−0.54 (.59)	−0.10 (.92)
BOTH			**−1.91 (.05)**			−0.462 (.64)
COM[*]			**−2.58 (.01)**			0.226 (.82)

[*] aggregate communication treatments.

from 50 percent in the initial rounds to 63 percent in the final rounds.[20] Thus, the difference in the distributions between *Public* and the other treatments persists.

The lower coordination rate on pure strategy Nash equilibria in *NoCom* is explained by large miscoordination in the initial periods. In the final rounds, there are no statistical differences of coordination rates between any two treatment (see Table 3).[21]

Result 1. *Coordination on pure strategy Nash equilibria is higher with communication than without, but this difference shrinks with experience.*

---

[20]The slight decline we observe in the central rounds is not significant.

[21]The results on coordination rates also hold against a permutation test on means, as well as a binomial test on the original data – i.e. not averaging within each matching group.
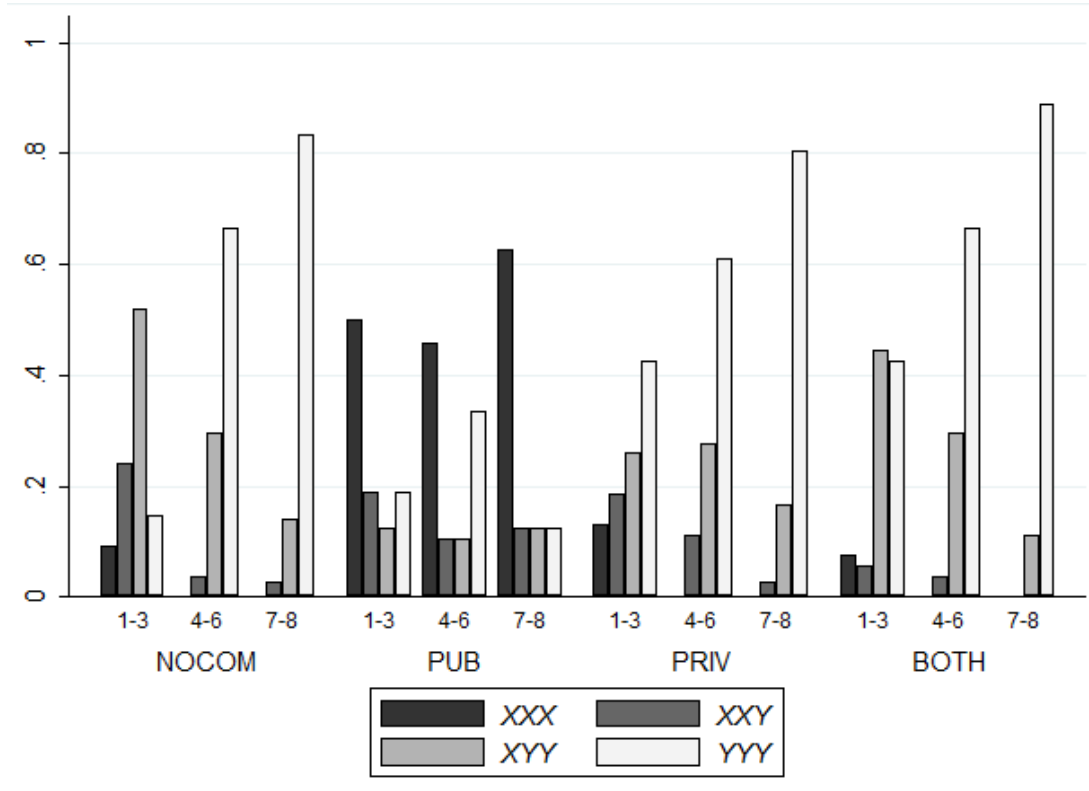
Result 2. *Coordination on the Pareto-dominant equilibrium is sustained only in Public. In the other treatments, play converges to the dominated equilibrium.*

These results match the implications of the theoretical analysis.

Figures 4 - 6 display the ex-post probability of reaching an outcome, given the profile of last-stage messages. In *Public* (Figure 4), the probability of achieving the PDE, given that an agreement to play it has been reached, is stable around .8 after the first round (solid black line). The decline in the occurrence of the PDE in the central rounds, and the recovery that follows, are explained by a decline in the number of agreements reached (dashed black line). The probability of playing the dominated equilibrium after a credible agreement to play it is one (solid dark gray line). If no agreement is reached, the probability of playing the dominated equilibrium oscillates between .2 and 1 (solid light gray line).[22]

With the exception of the first two rounds, the PDE is never reached under *Private* and *Both* (Figures 5 and 6), even when every player sends only message *x* (solid black lines). Conversely, if every player sends or receives at least one

---

[22]The dashed lines represent the probability of the corresponding message profiles.

FIGURE 4: OUTCOMES CONDITIONAL ON MESSAGES - *Public*

─✕─ P(*XXX*|agreement x)	─●─ P(*YYY*|no agreement)	─◆─ P(*YYY*|agreement y)
─ ✕ ─ P(agreement x)	─ ● ─ P(no agreement)	─ ◆ ─ P(agreement y)
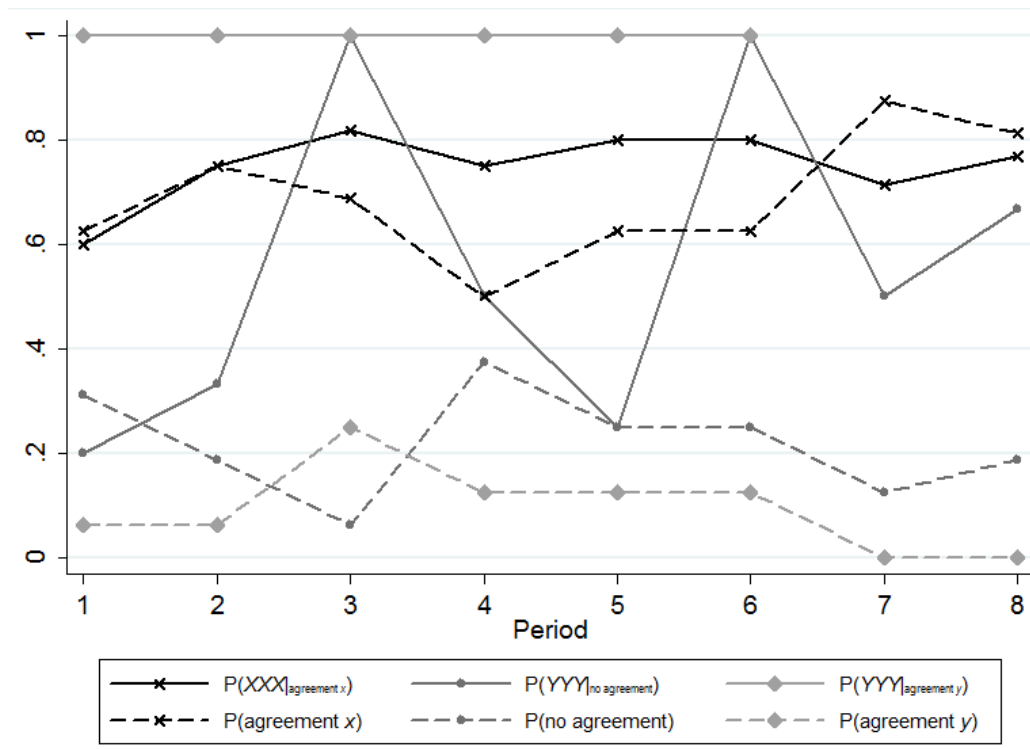
FIGURE 5: OUTCOMES CONDITIONAL ON MESSAGES - *Private*

|		
─✕─ P(*XXX*|all send & rec. x)	─●─ P(*XYY*|ONE sends & rec. x)	─◆─ P(*YYY*|agreement y)
─ ✕ ─ P(ALL send & rec. x)	─ ● ─ P(ONE sends & rec. x)	─ ◆ ─ P(agreement y)

15
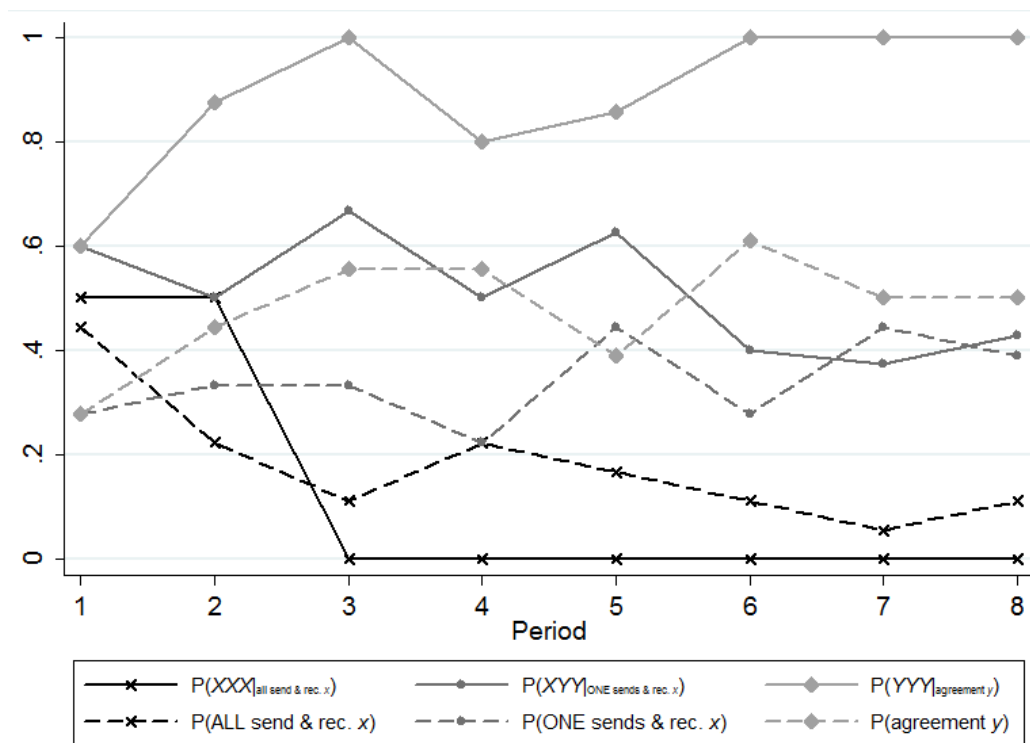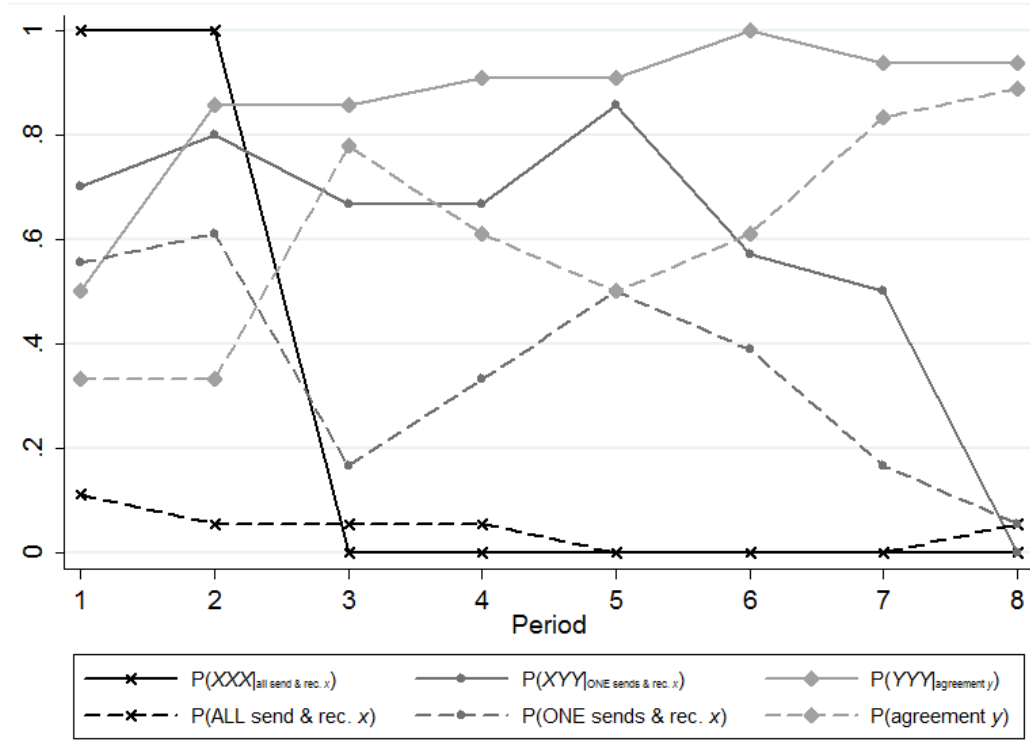
y message, a group is almost certain to play the dominated equilibrium (solid light gray lines). These message profiles happen more than half of the time in *Private*. In *Both* they represents more than 80 percent of the profiles in the last rounds (dashed light gray lines).

Exactly two players exchanged at least one *y* message privately with probability around one third, across all rounds, in *Private*. This figure peaks at .6 in the first two periods in *Both*, then declines, though still recording a .4 in the sixth round (dashed dark gray lines). Those messages profiles give two players the opportunity to let the third down by reaching the *XYY* outcome.

The probability of the *XYY* outcome conditional on these message profiles is between .5 and .6 in the first five rounds in *Private*, then declines to .4. In *Both*, the same figure is generally higher, peaking above .8 in round five, still above .5 in round seven (solid dark gray lines). Overall, it is remarkably high, and suggests that some agents believe the received messages even when they are not credible, while others try to exploit this tendency.

FIGURE 7: FRACTION OF X ACTIONS CONDITIONAL ON OBSERVED MESSAGES
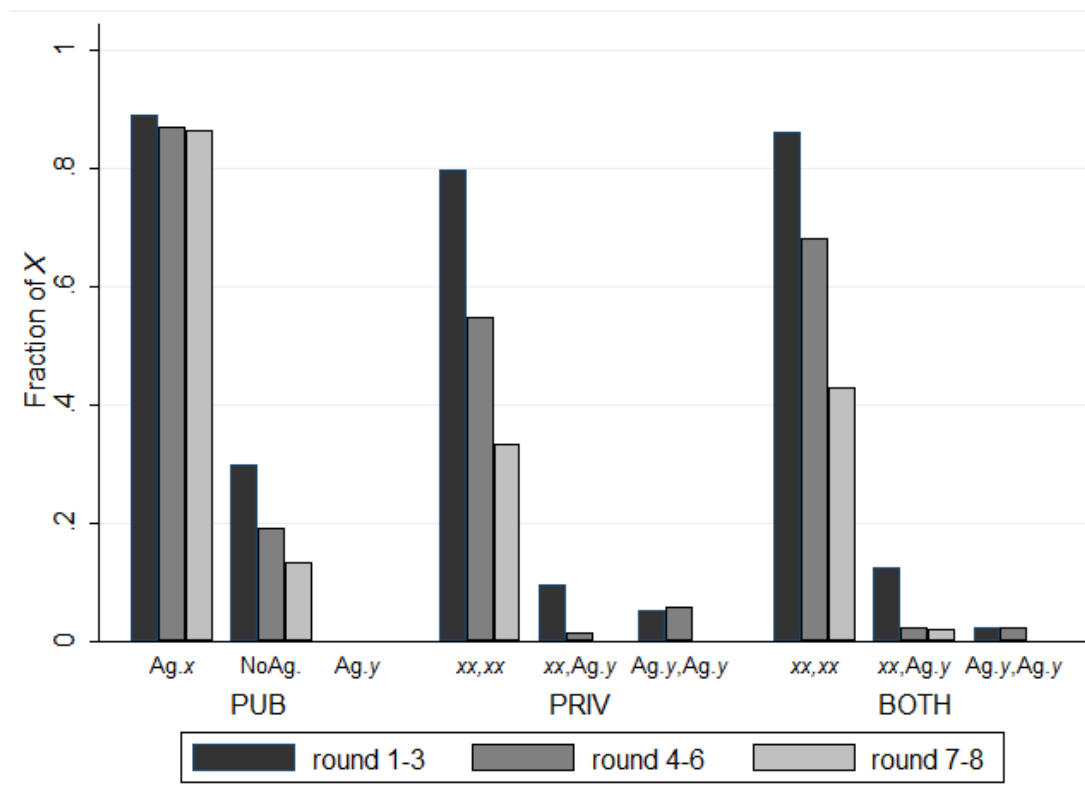
4.2 Actions, messages and beliefs

Action X is chosen 65 percent of the time in *Public*, 22, 21, and 15 percent in *NoCom*, *Private*, and *Both*, respectively. Figure 7 displays the fraction of X actions conditional on the observed messages in the last stage of communication, and for initial, intermediate and final rounds. In *Public*, subjects play X more than 80 percent of the time following an agreement to do so. If no agreement is reached, around one third of the subjects initially chooses X, but this fraction drops below 15 percent in the last rounds.[23] In *Private* and *Both*, the agreement to play Y with at least one group member is sufficient to induce action Y.[24] When players observe only message x, in the initial rounds they respond by playing X around 80 percent of the time, as in *Public*. As they gain experience, they learn not to trust these messages, but more than one third of the subjects still chooses X in the final rounds.

Table 4 shows how communication decisions evolved over rounds. It re-

---

[23]We already know from Figure 4 that everybody stick to agreements to play Y.

[24]For treatment *Both* we do not condition on the observed public messages. The reason is that, when it is observed in public, some message y is observed also in private.

TABLE 4: LAST MESSAGE SENT IN EACH TREATMENT

|	PUB		PRIV			BOTH					
	x	*y*	*xx*	*xy*	*yy*	*x-xx*	*x-xy*	*x-yy*	*y-xx*	*y-xy*	*y-yy*
1	.85	.15	.65	.26	.09	.44	.43	.02	.00	.02	.09
2	.90	.10	.48	.39	.13	.37	.50	.00	.00	.02	.11
3	.77	.23	.33	.52	.15	.20	.46	.06	.02	.09	.17
4	.77	.23	.46	.41	.13	.30	.61	.00	.00	.00	.09
5	.83	.17	.41	.52	.07	.19	.57	.04	.02	.06	.13
6	.83	.17	.35	.52	.13	.22	.61	.04	.00	.06	.07
7	.96	.04	.28	.63	.09	.13	.63	.09	.00	.04	.11
8	.94	.06	.30	.54	.16	.20	.52	.03	.02	.02	.20
Avg	.86	.14	.41	.47	.12	.27	.54	.03	.01	.04	.12

ports, for each treatment, the distribution of messages that are sent in the last stage of communication, round by round. In *Public*, *x* prevails across all rounds. In *Private*, the fraction of messages *xx* declines over rounds from above 60 percent to around 30. The converse happens to message *xy*, which steadily accounts for the majority of the messages in the second half of the experiment. In *Both* public messages *x*, combined with either *xx* or *xy* private messages account for between 75 and 90 percent of all messages. As in *Private*, the former declines steadily across rounds, accounting for around one fifth of the messages in the second half, while the latter increases up to around 60 percent from round 4 on.

The messages that are weakly dominated under our theoretical framework are rarely sent. The overall percentage ranges between 12 percent in *Private* and 19 percent in *Both*, with no significant difference across treatments, and no clear pattern across rounds. The presence of these messages may be interpreted as an instance of preferences for promise keeping. This motivation appears minor for our subjects, especially when one considers that half of the time subjects send different messages to their partners, thus necessarily lying to one of them.

That communication strategies reflected an attempt to affect partner's actions is confirmed by the distributions of first order beliefs.[25] In Figure 8 each circle represents a couple of beliefs, one for each partner. Beliefs are stated as the probability that a partner chooses X. The dimension of the circles represents the frequency of each couple of beliefs. We overimpose the frontier of

---

[25]We also collected data on second order beliefs. The analysis of those shows no difference with respect to first order ones. Thus, for brevity, we focus on first order beliefs in the main text. Results on second order beliefs are reported in Appendix C.

FIGURE 8: DISTRIBUTION OF FIRST ORDER BELIEFS
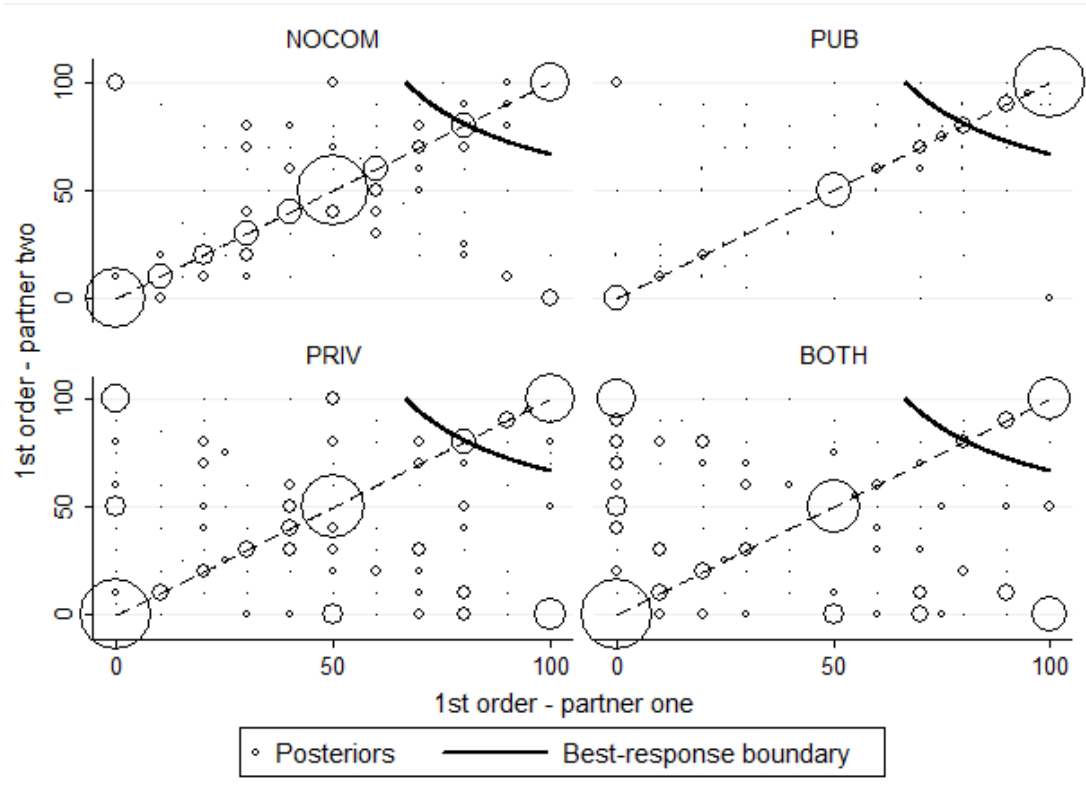
beliefs where the best response switches from *Y* (below the solid black line) to *X* (above it).[26]

As it should be expected at this point, in *Public* beliefs are concentrated on higher values for each partner with respect to the other treatments. A more subtle difference is present between *Public* and *NoCom*, on one side, and *Private* and *Both*, on the other. While in the former beliefs mainly fall on the bisector, where the same belief is reported for the two partners, in the latter a substantial share of players have divergent expectations regarding their partners. In particular, the highest off-the-diagonal frequencies correspond to expecting one partner to play *Y*, the other to play *X*, both with probability one. We check for the significance of this result, by testing across treatments for the difference in the absolute spread between each subject's first order beliefs (Table 5). A WRS rejects the null of equal differences in all pairwise comparisons, except for *Public* and *NoCom*.[27]

---

[26]In each treatment the fraction of best response to beliefs is above 80 percent.

[27]The joint distributions show a high frequency of beliefs around 50-50. The attraction toward such intermediate reports may be explained by risk aversion: 50-50 is a risk-free report. See also the analysis in Appendix C.

TABLE 5: INDIVIDUAL SPREAD IN FIRST ORDER BELIEFS

| | WRS test (P-val) | | |
	PRIV	BOTH	NOCOM
PUB	−**3.468**	−**3.466**	0.915
	(**.00**)	(**.00**)	(.360)
PRIV		−**2.250**	−**3.401**
		(.024)	(**.00**)
BOTH			−**3.311**
			(**.00**)

Figure 9 shows how the first order belief regarding one partner depends on the messages exchanged with him, for initial, intermediate, and final rounds. In *Public* the relation between belief and messages is rather stable: players trust messages if they reach an agreement, and are uncertain on what to expect if not. In *Private* and *Both* the players learn after the initial rounds to trust agreements to play Y, even when they descend from an implicit consensus, and independent of who is sending message y. If they both send x, they initially believe the message. Gradually, they learn this cannot be trusted. Still, in the final rounds, they report an average probability that their partner plays X close to 50 percent. These data suggest that communication affects beliefs in all treatments.

We run a GLS panel regression model with random effects to formally assess the effect of communication on beliefs.[28] The model can be written as:

$$B_{it} = \gamma_0 + \gamma_C C_{it} + \gamma_W W_{it} + \gamma_t t + u_{it} + \epsilon_{it}$$

where B_{it} is the first order posterior belief of individual i in round t, C_{it} is a set of regressors that refer to the outcome of the communication phase, W_{it} is a set of controls, including individual characteristic, such as attitudes toward trust and risk aversion, and statistics on the past observed play, u_{it} is the between-individual error, and ϵ_{it} the within-individual error.

Messages are different objects in different treatments. As the regressors are treatment specific, we first run treatment-by-treatment regressions. The communication regressors include all messages observed by the subject, and the interaction between the messages exchanged between the subject and the

---

[28]The opportunity to use time invariant regressors, a Hausman test, and the correlation between errors and regressors obtained through a diagnostic fixed effect estimation all suggested the random rather than the fixed effects model.
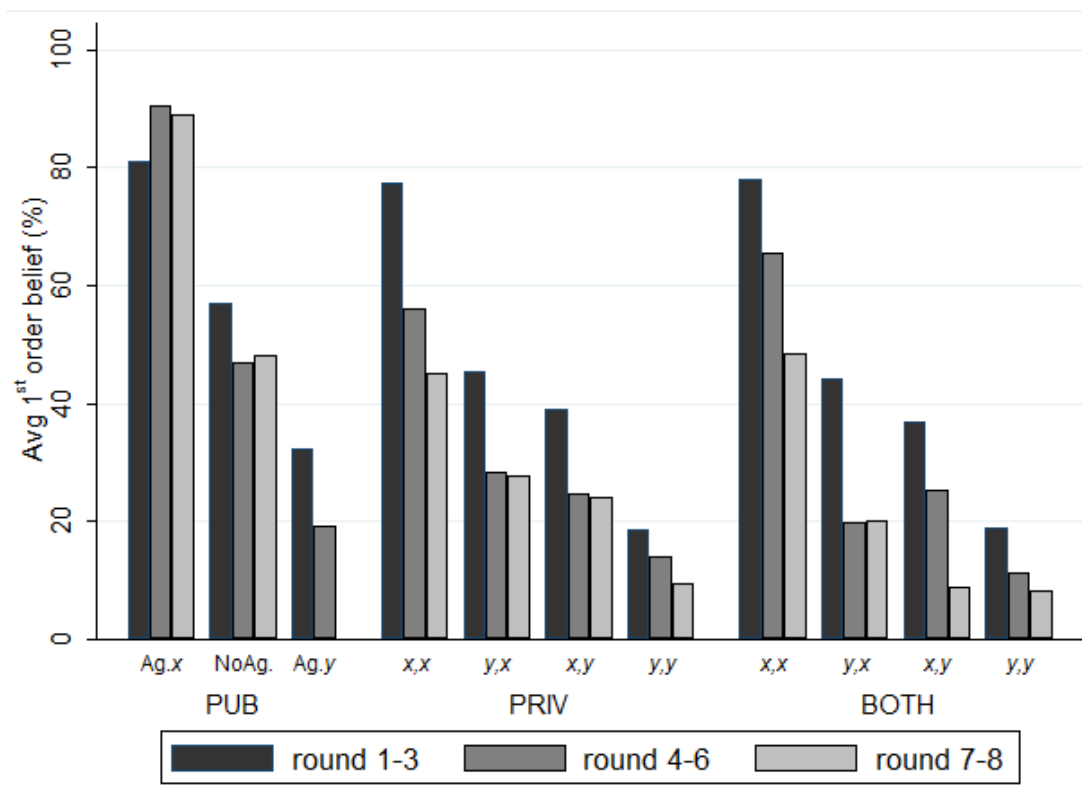
group member to which the belief refers.[29] For treatment *Both*, we collapse the outcome of public communication in one dummy taking value 1 whenever one or more public message y is sent (labeled $\tilde{y}_{PUB}^{BOTH}$ in Table 6). The standard errors are clustered at the matching group level.

One may be concerned that posteriors are the result of self-fulfilling priors: players' belief that induce communication strategies that end up confirming the prior.[30] In this case the causality would move from beliefs to communication, rather than the other way around. To address this concern we run a similar model where we use as dependent variable the change in first order beliefs, as measured by the difference between the posterior and the prior. A positive (negative) difference indicates that the subject is more (less) confident in his partner choosing action X after communication than he was before.

Table 6 reports the estimates for both models. The results indicate that messages communicating the intention to play Y induce beliefs to be significantly

---

[29]The main coefficients should then be interpreted as the effect of a y message when the interacted message is x.

[30]This concern is somewhat justified since we fail to find differences in the medians of priors and posterior beliefs in *Public* and *Private*. See Appendix C.

	PUB		PRIV		BOTH	
	(1)	(2)	(3)	(4)	(5)	(6)
	B_{ij}	ΔB_{ij}	B_{ij}	ΔB_{ij}	B_{ij}	ΔB_{ij}
t	0.71	-0.89	-2.03 ***	0.61	-3.27***	0.80
	(0.588)	(0.680)	(0.584)	(0.426)	(0.646)	(0.959)
$\bar{Y}_{\{j,k\}_{PAST}}$	-12.05***	8.73	-15.41***	2.34	-5.03**	10.29
	(3.995)	(4.67)	(1.368)	(3.20)	(2.295)	(6.354)
y_i^{PUB}	-36.26***	-23.88***				
	(6.204)	(7.007)				
y_j^{PUB}	-41.28***	-40.30***				
	(7.747)	(7.219)				
y_k^{PUB}	-22.30***	-20.61***				
	(3.007)	(3.257)				
$y_i^{PUB} * y_j^{PUB}$	26.05**	23.45				
	(12.68)	(15.36)				
$\tilde{y}_{BOTH}^{PUB}$					-9.18***	-9.91***
					(3.343)	(2.459)
y_{ij}			-24.74***	-23.11***	-31.90***	-20.27***
			(3.693)	(4.807)	(3.899)	(5.017)
y_{ji}			-29.06***	-28.30***	-38.23***	-37.80***
			(3.278)	(3.589)	(4.166)	(6.358)
y_{ik}			3.06	1.159	2.99	2.71
			(3.939)	(4.874)	(1.593)	(2.593)
y_{ki}			-2.73	-0.0795	-1.16	0.10
			(2.449)	(3.405)	(2.839)	(2.737)
$y_{ij} * y_{ji}$			10.66**	9.80	22.14***	25.25***
			(5.277)	(7.967)	(4.581)	(9.214)
γ_0	51.28	-28.48	85.61***	-20.81	65.31***	-15.07
	(31.34)	(24.31)	(14.49)	(11.30)	(7.783)	(16.25)
N	672	672	756	756	756	756
Matching groups	8	8	9	9	9	9
R-squared	0.548	0.323	0.493	0.263	0.503	0.235

Standard errors in parentheses (clustered at the matching group level)

** $p < 0.05$, *** $p < 0.01$

Controls include a set of individual characteristics, including risk aversion and trust measures.

revised downward. The highest coefficient in absolute terms is always that of the message sent by the group member to whom the belief refers (y_j^{PUB}, y_{ji}). Receiving a y from this player induces a 30-40 percentage points downward revision of the first order belief.[31] The message of the third group member has a significant effect only under *Public*, where it is observed by everyone, and not in the other treatments (y_k^{PUB}, y_{ki}). These results also hold when the dependent variable is the change in beliefs (ΔB_{ij}).

The coefficient of the interaction of exchanged messages is always significant and of opposite sign with respect to single messages ($y_i^{PUB} * y_j^{PUB}$, $y_{ij} * y_{ji}$). Thus it offsets the additive effect of the single messages when both players send y. The overall effect on posterior beliefs is around four thirds of the average effect of single y's in *Public* and *Both*, and three halves in *Private*. Results for the change in beliefs are similar, though some coefficients for the interaction are not significant.

A downward time trend on posterior beliefs is present in *Private* and *Both*. The average number of Y choices by past partners also negatively affects beliefs ($\bar{y}_{PAST}$). Both effects disappear if we look at changes in beliefs. A possible interpretation is that those two variables primarily affect prior beliefs, while communication determines the change between priors and posteriors.

The previous estimates show that communication affects beliefs in all treatments. To test whether they do so to a similar extent, we run a version of the previous models over all treatments. We collapse the communication variables into one dummy regressor (observes_y), taking value one whenever a player has observed at least one y in the last stage of communication, and we interact this dummy with the treatments (base treatment: *Private*; see Table 7).

The coefficients again show an impact of communication in all treatments, with a stronger effect in *Public*. In particular the posterior belief is on average 13.7 percentage points higher in *Public* with respect to *Private* after observing only message x, while *Both* does not differ significantly from the base. Observing at least one y, shifts beliefs downward of about 27 percent in *Private* and *Both*. The same figure is 45.8 percent in *Public* ($-26.9 - 18.9$), so that, summing up the effects, the average belief for those that observe y messages is about 5 percent lower in *Public* than in the other treatments.[32] Again, similar results are obtained by focusing on changes in beliefs. The insignificant main effect of *Public* tells there is no difference in how players update their beliefs after

---

[31]A change in beliefs of this magnitude is always sufficient to make Y a best response for the player.

[32]$(+13.7 - 26.9 - 18.9) - (-26.9)$.

TABLE 7: FIRST ORDER BELIEFS: ACROSS-TREATMENT GLS ESTIMATES

	(1) B_{ij}	(2) ΔB_{ij}
t	-1.42***	-0.311
	(0.511)	(0.435)
$\bar{Y}_{\{j,k\}PAST}$	-13.69***	6.07**
	(1.592)	(2.510)
observes_y	-26.90***	-24.69***
	(4.045)	(4.459)
Public	13.75**	-3.535
	(6.272)	(5.506)
Both	7.416	1.584
	(5.581)	(5.487)
observes_y*Public	-18.92**	-13.96**
	(8.754)	(6.794)
observes_y*Both	-8.22	-6.190
	(6.620)	(6.538)
γ_0	75.35***	2.852
	(9.104)	(9.176)
N	2184	2184
Matching groups	26	26
R-squared	0.417	0.123

Standard errors in parentheses (clustered at the matching group level)
** $p < 0.05$, *** $p < 0.01$
Controls include a set of individual characteristics,
including risk aversion and trust measures.

observing only x. Thus the corresponding positive coefficient found in model (1) is due to higher priors in *Public*. As before, observing y results in a larger revision of the beliefs in this treatment, but a significant revision happens in *Private* and *Both* as well.

Result 3. *Communication affects beliefs in all treatments. The effect is stronger in Public.*

Thus, despite the comparative statics across treatments are in line with theory, we reject the hypothesis that communication is disregarded in *Private* and *Both*.

5 Conclusion

When equilibria are Pareto-ranked, coordination problems can lead to inefficient outcomes. The fact that everybody would be happy to coordinate on Pareto-superior equilibria suggests that communication could help the agents to reach the Pareto-dominant outcome. Results in two-by-two games have generally supported this view. The paper presents an experiment on pre-play communication in a three-player coordination game with Pareto-ranked equilibria. We implement three different communication structures: subjects are allowed to send one public message to the two other players (*Public*), one private message to each other player (*Private*), or both public and private messages (*Both*).

Backed by theoretical analysis, we hypothesize that subjects should communicate and play the Pareto-dominant equilibrium only when communication is public. When pair of agents exchange messages privately, the players should play the Pareto-dominated equilibrium, and disregard communication.

The aggregate results conform to these predictions. In *Public*, communication sustains the Pareto-dominant equilibrium, which is never played when communication is absent. Thus, public communication alone replicates the efficiency-enhancing effect that communication has in two-player games [e.g. Charness, 2000]. If private messages are allowed, as in *Private* and *Both*, the players end up in the dominated equilibrium. This happens because lying is prevalent in these treatments, which makes efficient coordination impossible.

Thus, our results suggest caution in generalizing the effect of pre-play communication in two-player games to larger ones. The idea that efficient coordination may be reached through communication rests on equilibria being

Pareto-ranked. We have shown the importance of also considering how the interests of various groups of players are aligned outside equilibrium, and whether the available communication structure allows those groups to coordinate.

In particular, this latter element – i.e. who can talk with whom – plays a central role. The credible equilibrium notion, proposed by Grandjean and Mantovani [2014] as an extension of standard models of credibility to games with more players, precisely predicts the aggregate outcomes under different communication structures, as well as the communication strategies of the subjects.

The mechanics of credibility, however, are not straightforward to grasp, and subjects do not realize immediately when they can trust messages. In the first rounds, players respond similarly to the observed messages in all treatments. Even in the last ones, beliefs react to messages also in *Private* and *Both*, though to a smaller extent with respect to *Public*. Thus, the effects of pre-play communication, and of its structure, will generally depend also on how experienced the players are. As the widespread and persistent tendency to believe to messages that we detect is hardly matched by the subjects' preferences for truth-telling, groups of players use communication to cheat on their partners, and exploit naivety for their own interests.

References

AMBRUS, A., "Coalitional rationalizability," *The Quarterly Journal of Economics* (2006), 903–929.

ANDERSSON, O. AND E. WENGSTRÖM, "Credible communication and cooperation: experimental evidence from multi-stage games," *Journal of Economic Behavior & Organization* 81 (2012), 207–219.

AUMANN, R., "Nash equilibria are not self-enforcing," *Economic Decision Making: Games, Econometrics and Optimisation* (1990), 201–206.

BATTAGLINI, M. AND U. MAKAROV, "Cheap talk with multiple audiences: An experimental analysis," *Games and Economic Behavior* 83 (2014), 147–164.

BERNHEIM, B. D., B. PELEG AND M. D. WHINSTON, "Coalition-proof nash equilibria i. concepts," *Journal of Economic Theory* 42 (1987), 1–12.

BLUME, A., Y.-G. KIM AND J. SOBEL, "Evolutionary stability in games of communication," *Games and Economic Behavior* 5 (1993), 547–575.

BLUME, A. AND A. ORTMANN, "The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria," *Journal of Economic theory* 132 (2007), 274–290.

BRANDTS, J., G. CHARNESS AND M. ELLMAN, "Let's talk: How communication affects contract design," *mimeo* (2012).

CASTILLO, M. AND G. C. LEO, "From Coordination to Double-Crossing: Experiments on the Strategic Behavior of Groups," *mimeo* (2007).

CHARNESS, G., "Self-serving cheap talk: A test of Aumann's conjecture," *Games and Economic Behavior* 33 (2000), 177–194.

CHARNESS, G. AND M. DUFWENBERG, "Promises and partnership," *Econometrica* 74 (2006), 1579–1601.

———, "Message Personalization and Effectiveness in Cheap-talk Games," *mimeo* (2007).

———, "Participation," *The American Economic Review* 101 (2011), 1211–1237.

CHOI, S. AND J. LEE, "Communication, coordination, and networks," *Journal of the European Economic Association* 12 (2014), 223–247.

CLARK, K., S. KAY AND M. SEFTON, "When are Nash equilibria self-enforcing? An experimental analysis," *International Journal of Game Theory* 29 (2001), 495–515.

COOPER, R., D. V. DEJONG, R. FORSYTHE AND T. W. ROSS, "Communication in coordination games," *The Quarterly Journal of Economics* 107 (1992), 739–771.

COOPER, R., D. V. DEJONG, R. E. FORSYTHE AND T. W. ROSS, "Communication in the Battle of the Sexes Game: Some Experimental Results," *RAND Journal of Economics* 20 (1989), 568–587.

CRAWFORD, V., "A survey of experiments on communication via cheap talk," *Journal of Economic theory* 78 (1998), 286–298.

CRAWFORD, V. P. AND J. SOBEL, "Strategic information transmission," *Econometrica: Journal of the Econometric Society* (1982), 1431–1451.

DOHMEN, T., A. FALK, D. HUFFMAN, U. SUNDE, J. SCHUPP AND G. G. WAGNER, "Individual Risk Attitudes: Measurement, Determinants, And Behavioral Consequences," *Journal of the European Economic Association* 9 (2011), 522–550.

ELLINGSEN, T. AND M. JOHANNESSON, "Promises, Threats and Fairness*," *The Economic Journal* 114 (2004), 397–420.

FARRELL, J., "Communication, coordination and Nash equilibrium," *Economics Letters* 27 (1988), 209–214.

———, "Talk is cheap," *The American Economic Review* 85 (1995), 186–190.

FARRELL, J. AND R. GIBBONS, "Cheap Talk with Two Audiences," *American Economic Review* 79 (1989), 1214–23.

FARRELL, J. AND M. RABIN, "Cheap talk," *The Journal of Economic Perspectives* 10 (1996), 103–118.

FISCHBACHER, U., "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics* 10 (2007), 171–178.

GNEEZY, U., "Deception: The role of consequences," *The American Economic Review* 95 (2005), 384–394.

GRANDJEAN, G. AND M. MANTOVANI, "Structure of communication and credibility of messages," *mimeo* (2014).

MILGROM, P. AND J. ROBERTS, "Coalition-proofness and correlation with arbitrary communication possibilities," *Games and Economic Behavior* 17 (1996), 113–128.

MORENO, D. AND J. WOODERS, "An experimental study of communication and coordination in noncooperative games," *Games and Economic Behavior* 24 (1998), 47–76.

RABIN, M., "A Model of Pre-game Communication," *Journal of Economic Theory* 63 (1994), 370–391.

SAVAGE, L. J., "Elicitation of personal probabilities and expectations," *Journal of the American Statistical Association* 66 (1971), 783–801.

A   Experimental instructions

Welcome to this experiment in decision-making. You will receive 5 €as a show-up fee. Please, read carefully these instructions. The amount of money you earn depends on the decisions you and other participants make. In the experiment you will earn ECU (Experimental Currency Units). At the end of the experiment we will convert the ECU you have earned into Euro according to the rate: 1 Euro = 10 ECU. You will be paid your earnings privately and confidentially after the experiment. Throughout the experiment you are not allowed to communicate with other participants in any way. If you have a question please raise your hand. One of us will come to your desk to answer it.

The experiment consists of 8 identical rounds. At the beginning of every round the computer will pair you with two other participants in this room (labeled Partner 1 and Partner 2). Thus, you will form a new group in every new round and you will never play twice with the same group. In each round you will choose between 2 options, labeled A and B. You will have 30 seconds to make your choice. After you and your partners have chosen, you will be informed about the choice of your partners and of the ECUs you earned for the round. Your earnings will be determined according to the following four cases:

- Case 1: you and your partners all choose A. Everybody earns 80 ECU

- Case 2: you and your partners all choose B. Everybody earns 40 ECU

- Case 3: among you and your partners, two persons choose A and one person chooses B. Everybody earns 0 ECU.

- Case 4: among you and your partners, one person chooses A and two persons choose B. The person that chooses A earns 0 ECU. The persons that chose B earn 160 ECU.

[Treatment PUBLIC] Before you make your choice, you will be asked to communicate with your partners about which choice you intend to make. You will complete the message "I intend to choose:" with A or B. You will send one message to both of your partners. When you and your partners have sent the messages, you will be informed about the content of the messages that were sent to you. You will repeat this procedure 4 times; thus there will be 4 stages of communication, and you will have two minutes to complete all of them.

After that you will make your choice as explained before. The messages that are sent do not limit your choice.

[Treatment PRIVATE] Before you make your choice, you and your partners will be asked to communicate with your partners about which choice you intend to make. You will complete the message "I intend to choose:" with A or B. You will send one message to each of your partners; thus you will send two messages. You can send the same message or different messages. When you and your partners have sent the messages, you will be informed about the content of the messages that were sent to you. You will repeat this procedure 4 times; thus there will be 4 stages of communication, and you will have two minutes to complete all of them. After that, you will make your choice as explained before. The messages that are sent do not limit your choice.

[Treatment BOTH] Before you make your choice, you and your partners will be asked to communicate with your partners about which choice you intend to make. You will complete the message "I intend to choose:" with A or B. You will send one message to each of your partners and one message to both of your partners; thus, you will send three messages. You can send the same message or different messages. When you and your partners have sent the messages, you will be informed about the content of the messages that were sent to you. You will repeat this procedure 4 times; thus there will be 4 stages of communication, and you will have two minutes to complete all of them. After that you will make your choice as explained before. The messages that are sent do not limit your choice.

[All treatments] You will be also asked to state your beliefs both about your partners choice and about their beliefs regarding your own choice. For each of your partners you will be asked to complete two tasks, stating: a) a number between 1 and 100, representing the probability with which you think he will choose action A ; b) a number between 1 and 100, representing the probability with which you think he thinks you will choose action A. Thus you will be asked to state four numbers, corresponding to four belief tasks. The ECU you earn from the belief tasks depend on the accuracy of your beliefs. In particular, for task a) you earn more ECU the closer you are to the actual decision of your partner; for task b) you earn more ECU the closer you are to the number inserted by your partner in task a). The number of ECU you can earn ranges between 0 and 25. We enclose two tables displaying the ECU you earn in tasks a) and b), depending on your accuracy. Note that the ECU are chosen so that it is in your best interest to always state what you truly believe.

[Treatment NOCOM] You will state your beliefs after you made your choice.

[Treatments PUB,PRIV,BOTH] You will state your beliefs twice: before the communication phase and after you made your choice.

[All treatments] You will be paid only for one of the 8 rounds of the game and for only one single belief task you answered. At the end of the experiment, one number between 1 and 8 will be selected at random by the computer, and the corresponding game will be paid. The computer will select at random another round, different from the previous, and one single belief task in that round: the corresponding points will be paid. You will be informed of the chosen rounds, of your final payoff in ECU and of the corresponding Euro. You have reached the end of the instructions. It is important that you understand them. If anything is unclear to you or if you have questions, please raise your hand. To ensure that you understood the instructions we ask you to answer a few control questions. The experiment will start after everyone has answered these control questions correctly.

Earnings in the belief tasks

TABLE 8: EARNINGS (ECU) FOR TASK a)

		Partner's actual choice	
		A	**B**
	0	0	25
	10	4.75	24.75
	20	9	24
	30	12.75	22.75
	40	16	21
Your belief	**50**	18.75	18.75
	60	21	16
	70	22.75	12.75
	80	24	9
	90	24.75	4.75
	100	25	0

TABLE 9: EARNINGS (ECU) FOR TASK b)

		Partner's belief in task a)										
		0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
	0	25	24.75	24	22.75	21	18.75	16	12.75	9	4.75	0
	10	24.75	25	24.75	24	22.75	21	18.75	16	12.75	9	4.75
	20	24	24.75	25	24.75	24	22.75	21	18.75	16	12.75	9
	30	22.75	24	24.75	25	24.75	24	22.75	21	18.75	16	12.75
	40	21	22.75	24	24.75	25	24.75	24	22.75	21	18.75	16
Your belief	**50**	18.75	21	22.75	24	24.75	25	24.75	24	22.75	21	18.75
	60	16	18.75	21	22.75	24	24.75	25	24.75	24	22.75	21
	70	12.75	16	18.75	21	22.75	24	24.75	25	24.75	24	22.75
	80	9	12.75	16	18.75	21	22.75	24	24.75	25	24.75	24
	90	4.75	9	12.75	16	18.75	21	22.75	24	24.75	25	24.75
	100	0	4.75	9	12.75	16	18.75	21	22.75	24	24.75	25

B Credible communication: formal model

B.1 Credibility

Let G be a normal form game defined by $< N, (A_j)_{j \in N}, (u_j)_{j \in N} >$, where N is the player set of cardinality n, A_i is the finite set of action of player i, and $u_i : \times_{j \in N} A_j \to R$ is player i's von Neumann-Morgenstern utility function. We refer to G as the stage game.

Let a communication structure be $\mathbb{C} = \{C_1, C_2, ..., C_m\}$, where $C_k \subseteq N$. $\Gamma^{\mathbb{C}}(G)$ denote the game G preceded by one round of preplay communication, according to the coalition structure $\mathbb{C}$. In each coalition C, each player $i \in C$ communicates a message about his intended action $m_{i,C} \in A_i$ to the members of coalition C. All messages are sent simultaneously.

Take a group of agents $C \in \mathbb{C}$ and a message profile in that group $m_C \in \times_{j \in C} A_j$. The message profile m_C is self-committing if each group member is better off by playing his message when he trusts the message of the others, for every possible belief about the play of nonmembers.

Definition 1. *The message profile $m_C \in \times_{j \in C} A_j$ is **self-committing** if $u_i(m_C, \alpha_{N \setminus C}) \geq u_i(a_i, m_{C \setminus \{i\}}, \alpha_{N \setminus C})$ for all $i \in C$, for all $a_i \in A_i$, for all $\alpha_{N \setminus C} \in \times_{j \in N \setminus C} \Delta A_j$.*

A message profile that is not self-committing leads to a consensus to play an action profile if (i) the players who do not play according to their message choose a best-response to the message of those who stick to it, (ii) the message of the players who stick to it is a best-response to the play of the consensus, and (iii) there is no other action profile satisfying (i) and (ii).

Definition 2. *A message profile $m_C \in \times_{j \in C} A_j$ that is not self-committing leads to the **consensus** $a_C \in \times_{j \in C} A_j$ if*

(i) $a_C = (m_{C'}, a_{C \setminus C'})$ for some $C' \subset C$ with $C' \neq \emptyset$, where a_i satisfies $u_i(a_i, m_{C'}, \alpha_{N \setminus (C' \cup \{i\})}) \geq u_i(a_i', m_{C'}, \alpha_{N \setminus (C' \cup \{i\})})$ for all $i \in C \setminus C'$, for all $a_i' \in A_i$, for all $\alpha_{N \setminus (C' \cup \{i\})} \in \times_{j \in N \setminus (C' \cup \{i\})} \Delta A_j$.

(ii) $u_i(m_i, a_{C \setminus \{i\}}, \alpha_{N \setminus C}) \geq u_i(a_i', a_{C \setminus \{i\}}, \alpha_{N \setminus C})$ for all $i \in C'$, for all $a_i' \in A_i$, for all $\alpha_{N \setminus C} \in \times_{j \in N \setminus C} \Delta A_j$.

(iii) There is no $\overline{a_C} \in A_C \setminus \{a_C\}$ satisfying (i) and (ii).

A message profile leads to an agreement if it is self-committing or if it leads to a consensus.

Definition 3. *Following the message profile $m_C \in \times_{j \in C} A_j$, an **agreement** to play $a_C \in \times_{j \in C} A_j$ is reached in the group $C \in \mathbb{C}$ if $a_C = m_C$ and m_C is self committing, or if m_C leads to the consensus a_C.*

For $C \in \mathbb{C}$, and $m_C \in \times_{j \in C} A_j$, let $Ag(m_C)$ be the agreement reached following the message profile m_C, and let $Ag(C) = \cup_{m_C \in A_C} Ag(m_C)$. An agreement is credible in a group if no members of that group could gain by following an agreement reached in another group they also belong to.

Definition 4. *An agreement to play a_C in the group $C \in \mathbb{C}$ is **credible** if for all $C' \in \mathbb{C}$ such that $C \cap C' \neq \emptyset$ and for all $b_{C'} \in Ag(C')$, $u_i(a_C, b_{C' \setminus C}, \alpha_{N \setminus (C \cup C')}) \geq u_i(a_{C \setminus C'}, b_{C'}, \alpha_{N \setminus (C \cup C')})$ for some $i \in C \cap C'$ such that $a_i \neq b_i$, for all $\alpha_{N \setminus (C \cup C')} \in \times_{j \in N \setminus (C \cup C')} A_j$.*

For $C \in \mathbb{C}$, and $m_C \in \times_{j \in C} A_j$, let $Ag^*(m_C)$ be the credible agreement reached following the message profile m_C.

B.2 Credibility in the Coalitional Prisoners Dilema

We consider three types of coalition structure: $\mathbb{C}^{PUB} = \{\{i,j,k\}\}$, $\mathbb{C}^{PRIV} = \{\{i,j\}, \{i,k\}, \{j,k\}\}$ and $\mathbb{C}^{BOTH} = \{\{i,j\}, \{i,k\}, \{j,k\}, \{i,j,k\}\}$. We let $\Gamma^{PUB}(G)$ denote the coalitional prisoners dilema G preceded by one round of public preplay communication, and we let $\Gamma^{PRIV}(G)$ and $\Gamma^{BOTH}(G)$ be defined accordingly.

In $\Gamma^{PUB}(G)$, the message profiles $m = (x,x,x)$ and $m' = (y,y,y)$ are self-committing since they are associated to the pure Nash equilibria of the game. The message profile (x,y,y) leads to a consensus to play (y,y,y) while the profile (x,x,y) does not lead to a consensus. Agreements are credible since there are no conflicting agreements that could be reached in other groups.

In $\Gamma^{PRIV}(G)$, the only self-committing message profile in a group C is $m_C = (y,y)$. The message profiles $m_C \in \{(x,y), (y,x)\}$ lead to the consensus (y,y) while the message profile $m_C = (x,x)$ does not lead to a consensus. Players may only agree privately to play Y. As a result, every agreement to play Y is credible.

In $\Gamma^{BOTH}(G)$, the public message profile $m_{\{i,j,k\}} = (x,x,x)$ leads to an agreement to play the Pareto dominant equilibrium. However, it is not credible since any group of two agents could agree on the play of (Y,Y), which is payoff improving for those two agents. Every agreement to play Y in a group of two agents is credible since the other agreements that could be reached either

involve the play of Y, or involve the play of X by the three players. In that case, the two agents are better off under their initial agreement. Every agreement to play Y in a group of three players is credible since every other agreement involve the play of Y.

B.3 Credible equilibrium

In $\Gamma^{\mathbb{C}}(G)$, let $c_i = \#\{C \in \mathbb{C} \mid i \in C\}$. Let $m_{i,C} \in A_i$ be the message of player i in the group C. Let $m_i = (m_{i,C})_{C \in \mathbb{C}|i \in C} \in A_i^{c_i}$ be a message profile for player i. For $C \in \mathbb{C}$, let $m_C = (m_{j,C})_{j \in C} \in \times_{j \in C} A_j$ be a message profile in group C. Let $m = (m_C)_{C \in \mathbb{C}} \in \times_{C \in \mathbb{C}} (\times_{j \in C} A_j)$ be a message profile in each group. Let $m^i = (m_C)_{C \in \mathbb{C}|i \in C} \in \times_{C \in \mathbb{C}|i \in C} (\times_{j \in C} A_j)$ be a message profile in each group C involving player i. The strategy of a player i is $\sigma_i = (p_i, f_i)$, where $p_i : A_i^{c_i} \to [0,1]$ is a probability distribution over his messages profiles such that $\sum_{m_i \in A_i^{c_i}} p_i(m_i) = 1$, and $f_i : \times_{C \in \mathbb{C}|i \in C} (\times_{j \in C} A_j) \to \Delta A_i$ is a mapping from the set of messages observed by player i to a probability distribution over his action. The expected utility of player i given the strategy profile $\sigma = (\sigma_1, \sigma_2, ..., \sigma_n)$ is $U_i(\sigma) = \sum_{a \in A} u_i(a) [\sum_{m \in \times_{C \in \mathbb{C}} (\times_{j \in C} A_j)} p(m) \Pi_{j \in N} f_j(a_j \mid m^j)]$, where $p(m) = \Pi_{j \in N} p_j(m_j)$.

A credible equilibrium is a Perfect Bayes Nash equilibrium of the game with pre-play communication where it is common knowledge that credible agreements are believed and played.

Definition 5. *The strategy profile $\sigma = (p, f)$ is a credible equilibrium of $\Gamma^{\mathbb{C}}(G)$ if for each player $i \in N$,*

(i) $p_i(m_i) > 0 \Rightarrow m_i \in \arg\max_{m'_i \in A_i^{c_i}} U_i((m'_i, f_i), \sigma_{-i})$

(ii) *For all $m^i \in \times_{C \in \mathbb{C}|i \in C} (\times_{j \in C} A_j)$.*

a. *If $a_C \in Ag^*(m_C)$ for some $C \in \mathbb{C}$ such that $i \in C$, then $f_i(a_i \mid m^i) = 1$,*

b. *Otherwise,*

$$f_i(a_i \mid m^i) > 0 \Longrightarrow a_i \in \arg\max_{a'_i \in A_i} \sum_{a_{-i} \in A_{-i}} u(a'_i, a_{-i})$$
$$\prod_{j \neq i} \sum_{\widetilde{m}^j \in \times_{C \in \mathbb{C}|j \in C} (\times_{k \in C} A_k)} f_j(a_j \mid \widetilde{m}^j) \mu_i(\widetilde{m}^j \mid m^i)$$

where for all j and $\widetilde{m}^j$, $\mu_i(\widetilde{m}^j \mid m^i) = \prod_{k \in N} \sum_{\widehat{m}_k \in A_k^{c_k}|\widehat{m}_{k,C} = \widetilde{m}_{k,C} \forall C|j,k \in C} p_k(\widehat{m}_k \mid m^i)$ is player i's assessment of the probability player j observes $\widetilde{m}^j$, given the equilibrium choice of messages by the agents p and his observation m^i, computed using Bayes rule when possible.

Condition (i) of Definition 5 requires that the choice of message of each player is optimal. Condition (ii.a) imposes that credible agreement are played. Condition (ii.b) requires that when a credible agreement is not reached, a player chooses an optimal action given the expected play of others, by deriving the observation they might get from the optimal choice of messages of each player, updated by his own observation according to Bayes rule. Notice that Condition (ii) should hold for all observation a player might get, not only on the equilibrium path.

B.4 Credible equilibrium in the Coalitional Prisoner's dilemma

In $\Gamma^{PUB}(G)$, $m_l \in \{x,y\}$, $p_l(x) + p_l(y) = 1$, $m^i = m^j = m^k = (m_i, m_j, m_k) = m$. Since each player's observation is identical, player i's assessment of the probability player j observes $\widetilde{m}^j$ is

$$
\mu_i(\widetilde{m}^j \mid m^i) = \begin{cases} 1 & \text{if } m^i = \widetilde{m}^j \\ 0 & \text{if } m^i \neq \widetilde{m}^j \end{cases}
$$

For each player $l \in \{i,j,k\}$, credibility implies $f_l(x \mid (x,x,x)) = 1$ and $f_l(y \mid m) = 1$ for $m \in \{(y,y,y),(x,y,y),(y,x,y),(y,y,x)\}$. Players coordinate on a Nash equilibrium of the CPD, if $m \in \{(x,x,y),(x,y,x),(y,x,x)\}$. The following matrix represents the payoff of the agents for each message profile, given the role of communication on play. In the matrix, $z_1 = 8$ if $f_l(x \mid (y,x,x)) = 1$, $z_1 = 5,18$ if $f_l(x \mid (y,x,x)) = \frac{1+\sqrt{2}}{3}$ and $z_1 = 4$ if $f_l(x \mid (y,x,x)) = 0$, and z_2 and z_3 are defined accordingly. This matrix thus defines 27 one shot game. Each Nash equilibrium of this one-shot game is associated to a credible equilibrium, as shown hereafter.

FIGURE 10: PAYOFFS IN THE CONTINUATION GAME

	x	y
x	8,8,8	z_2,z_2,z_2
y	z_1,z_1,z_1	4,4,4

x

	x	y
x	z_3,z_3,z_3	4,4,4
y	4,4,4	4,4,4

y

There are four classes of credible equilibria. For each candidate we identify the choice of message by each player and the choice of action when an agreement is not reached. When an agreement is reached, the players play it.

(i) Each player sends a message y and the players coordinate on any of the Nash equilibria of the game if they do not reach an agreement. Formally,

$p_l(x) = 0$; $f_l(x \mid m) = f$ for $l \in \{i,j,k\}$, where $f \in \{0, \frac{1+\sqrt{2}}{3}, 1\}$ and $m \in \{(x,x,y);(x,y,x);(y,x,x)\}$.

(ii.a) Players 1 and 2 send y while player 3 sends x with positive probabilities. If they do not reach an agreement and player 1 or player 2 are the players who have announced their intention to play Y, players coordinate on the dominated equilibrium. If the players do not reach an agreement and player 3 reveals y, they coordinate on any of the Nash equilibria.

Formally, $p_1(x) = p_2(x) = 0; p_3(x) \in]0,1]$, and for $l \in \{1,2,3\}$, $f_l(x \mid (y,x,x)) = 0$; $f_l(x \mid (x,y,x)) = 0$, $f_l(x \mid (x,x,y)) = f$ where $f \in \{0, \frac{1+\sqrt{2}}{3}, 1\}$.

(ii.b) and (ii.c) are symmetric to (ii.a).

(iii) Each player sends a message x and players coordinate on any of the Nash equilibria of the game if they do not reach an agreement. Formally, $p_l(x) = 1$; $f_l(x \mid m) = f$ for $l \in \{1,2,3\}$, where $f \in \{0, \frac{1+\sqrt{2}}{3}, 1\}$ and $m \in \{(x,x,y);(x,y,x);(y,x,x)\}$.

(iv.a) Players 1 and 2 send x while player 3 sends y with positive probabilities. If they do not reach an agreement and player 3 has announced his intention to play Y, players coordinate on the Pareto dominant equilibrium. If the players do not reach an agreement and players 1 (2) reveals y, they coordinate on any of the Nash equilibria. Formally, $p_1(x) = p_2(x) = 1; p_3(x) \in [0,1[$, and for $l \in \{1,2,3\}$, $f_l(x \mid (x,x,y)) = 1$; $f_l(x \mid (x,y,x)) = f_l(x \mid (y,x,x)) = f$ where $f \in \{0, \frac{1+\sqrt{2}}{3}, 1\}$.

(iv.b) and (iv.c) are symmetric to (iv.a).

The other candidates are not immune to individual deviations.[33]

Communicating y is a weakly dominated strategy. If communicating y affects the payoff of a player, it leads to a lower payoff for that player. Sending the message y leads to a strictly smaller expected payoff than sending the message x if others communicate both messages with positive probabilities.

In $\Gamma^{PRIV}(G)$, credibility implies $f_l(y \mid m^l) = 1$ when $m^l = (m_{l,\{lk\}}, m_{k,\{lk\}}, m_{l\{lj\}}, m_{j,\{lj\}}) \neq (x,x,x,x)$. A strategy (p_l, f_l) for player l such that $p_l(y,y) > 0$ is weakly dominated. Indeed, player l would get a payoff of 4 by sending (y,y) while he would get $4\alpha + 16(1 - \alpha)$ by sending (x,y) where α is the probability that a player observes only messages x and react to

---

[33] 1°) Suppose $p_1(x) = 0$, $p_2(x) \in]0,1]$, $p_3(x) \in]0,1[$. Then, we should have $f_l(x \mid (y,x,x)) = 1$ for $l \in \{1,2,3\}$ to prevent the deviation by player 1 from $p_1(x) = 0$ to $p'_1(x) = 1$. It follows that player 3 could profitably deviate to $p'_3(x) = 1$.

2°) Suppose $p_1(x) = 1$, $p_2(x) \in [0,1[$, $p_3(x) \in]0,1[$. Then, we should have $f_l(x \mid (x,y,x)) = 1$ for $l \in \{1,2,3\}$ to prevent the deviation by player 2 from $p_2(x)$ to $p'_2(x) = 1$. It follows that player 3 could profitably deviate to $p'_3(x) = 1$.

this by playing X. To simplify the analysis, we only discuss the credible equilibria involving strategies that are not weakly dominated, and we assume that if a player sends exactly one message y, he sends it randomly to one of the two partners. We show that the only symmetric credible equilibria of the game in this class of candidates are such that players play the dominated equilibrium.

Proposition 1. *The symmetric credible equilibria of the game such that $p_l^*(y,y) = 0$ and $p_l^*(x,y) = p_l^*(y,x)$ for $l \in \{i,j,k\}$ are such that a) $p_l^*(x,x) \in]0,1]$, $f_l^*(x \mid (x,x,x,x)) = 0$, or b) $p_l^*(x,x) = 0$, $f_l^*(x \mid (x,x,x,x)) \in [0,1]$.*

Proof. Proof. Let σ^* be a credible equilibrium. For $l \in \{i,j,k\}$, let $p_l^*(y,y) = 0$, $p_l^*(x,y) = p_l^*(y,x) = \pi/2$, and $f_l^*(X \mid (x,x,x,x)) = \phi$. Credibility imposes $f_l^*(X \mid o^l) = 0$ for $o^l \neq (x,x,x,x)$. Using Bayes rules, we have $\mu_l(o^{l'} = (x,x,x,x) \mid o^l = (x,x,x,x)) = (\frac{1-\pi}{\pi/2+(1-\pi)})^2$ for all $l \neq l'$.

Let $\lambda(\pi) = (2(\frac{1-\pi}{\pi/2+(1-\pi)})^2 - \sqrt{(\frac{1-\pi}{\pi/2+(1-\pi)})^2 + (\frac{1-\pi}{\pi/2+(1-\pi)})^4})/3(\frac{1-\pi}{\pi/2+(1-\pi)})^2$ be the value of ϕ such that a player observing (x,x,x,x) is indifferent between playing X or playing Y.[34] There are three classes of equilibrium candidates for play in the stage game given π: (i) $\phi = 1$ when $\lambda(\pi) \leq 1$, (ii) $\phi = 0$ when $\lambda(\pi) \geq 0$ and (iii) $\phi = \lambda(\pi)$ when $\lambda(\pi) \in]0,1[$. Notice that $\phi(0) = (2 - \sqrt{2})/3$, $\partial\phi(\pi)/\partial\pi < 0$, and $\phi((20 - \sqrt{48})/22) = 0$. Thus, the only equilibrium candidates satisfying $\pi > (20 - \sqrt{48})/22$ are such that $\phi = 1$. When $\pi \leq (20 - \sqrt{48})/22$, the three candidates satisfy sequential rationality at information sets where agents observes (x,x,x,x).

a) Suppose $\pi \in [0,1[$.

(a.i) Suppose $\phi = 0$, then $U_l((p_l', f_l^*),(\sigma_{-l}^*)) = 6\pi^2 - 18\pi + 16 > U_l(\sigma^*) = 11\pi^3 - 23\pi^2 + 8\pi + 8$ for any $l \in \{i,j,k\}$, for $p_l'(x,y) = 1$, contradicting σ^* is a credible equilibrium.

(a.ii) Suppse $\phi = 1$, then $U_l(\sigma^*) = 4 = U_l(\sigma_l', \sigma_{-l}^*)$ for all $l \in \{i,j,k\}$, for all $\sigma_l' \neq \sigma_l^*$.

(a.iii) Suppose $\phi = (2(\frac{1-\pi}{\pi/2+(1-\pi)})^2 - \sqrt{(\frac{1-\pi}{\pi/2+(1-\pi)})^2 + (\frac{1-\pi}{\pi/2+(1-\pi)})^4})/3(\frac{1-\pi}{\pi/2+(1-\pi)})^2$. Then, $U_l((p_l', f_l^*),(\sigma_{-i}^*)) = 16 - 12\phi - 18\pi + 18\pi\phi + 6\pi^2 - 6\pi^2\phi > U_l(\sigma^*) = \pi^3(11 + 9\phi - 56\phi^2 + 36\phi^3) + \pi^2(-23 - 37\phi + 168\phi^2 - 108\phi^3) + \pi(8 + 52\phi - 168\phi^2 + 108\phi^3) + 8 - 24\phi + 56\phi^2 - 36\phi^3$ for any $l \in \{i,j,k\}$, for all $\pi < (20 - \sqrt{48})/22$, where $p_l'(x,y) = 1$, contradicting σ^* is a credible equilibrium..

---

[34]$\phi(\pi)$ is the solution to $(\frac{1-\pi}{\pi/2+(1-\pi)})^2(1 - \phi)^2 8 = 2(\frac{1-\pi}{\pi/2+(1-\pi)})^2\phi(1 - \phi)16 + (\frac{1-\pi}{\pi/2+(1-\pi)})^2\phi^2 4 + (1 - (\frac{1-\pi}{\pi/2+(1-\pi)})^2)4,$

b) Suppose $\pi = 1$, then $U_l(\sigma^*) = 4 = U_l(\sigma'_l, \sigma^*_{-l})$ for all $l \in \{i, j, k\}$, for all $\sigma'_l \neq \sigma^*_l$, for all $\phi \in [0, 1]$.

□

In $\Gamma^{BOTH}(G)$, we have seen that agents cannot credibly reveal their intention to play x. Following a private message y, an agent plays Y. When an agent observes at most one public message y and only private messages x, he should choose an optimal action given his equilibrium beliefs. As in the case of private communication, the only symmetric equilibria of the game are such that players choose Y.

C Further results

Figures 11-12 and Table 10 report for second order beliefs the same statistics we have discussed in the main text for first order ones. Results are similar. In particular, in *Private* and *Both* more players think their partners have divergent expectations on their action.
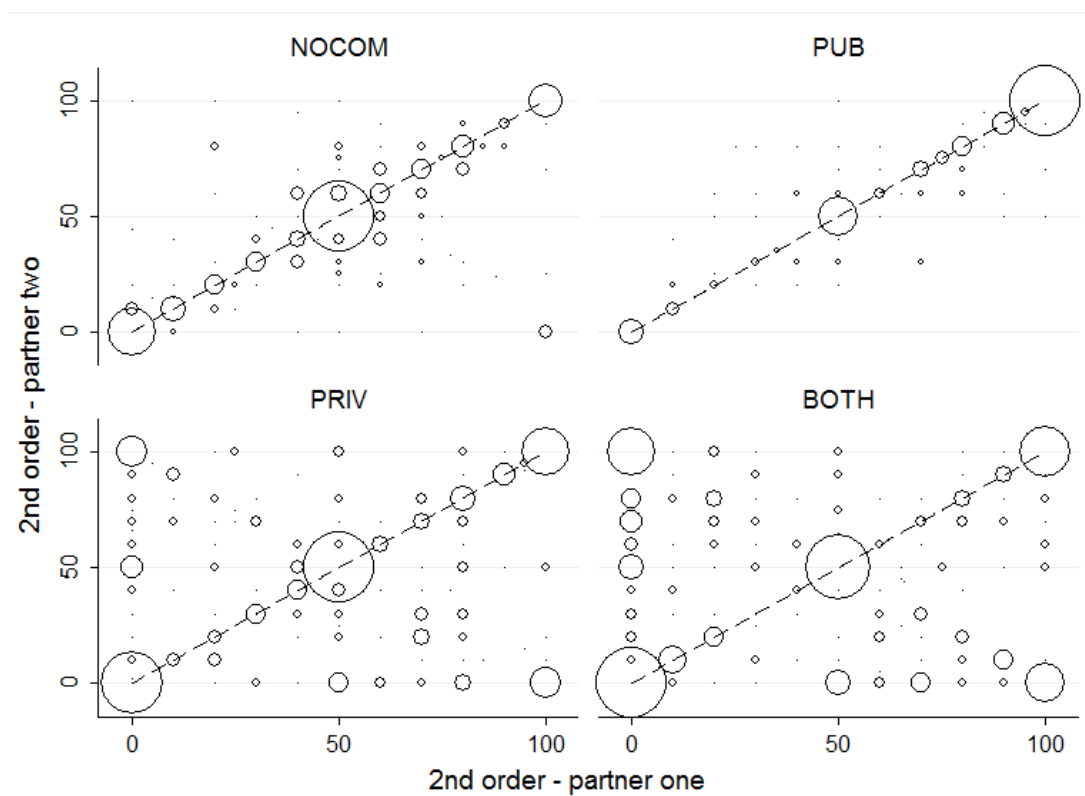
FIGURE 11: DISTRIBUTION OF SECOND ORDER BELIEFS

Table 10 shows the related tests. The null hypothesis is that the (median) absolute spread between each subject's second order beliefs is the same across a pair of treatments. We reject this hypothesis for all pairwise comparisons.

Figure 12 shows how the second order belief regarding one partner depends on the messages exchanged with him, for initial, intermediate, and final rounds. The results mimic those of first order beliefs. In *Public* players expect others to trust credible agreements. In *Private* and *Both*, if they both send x, they initially think their message is believed. Gradually, they learn this is not necessarily the case. When they send y, players expect themselves to be trusted. However, when they send x, and receive y, they are on average more uncertain on the belief of their partner. Thus, after an implicit consen-

TABLE 10: INDIVIDUAL SPREAD IN SECOND ORDER BELIEFS

| | WRS test (P-val) | | |
	PRIV	BOTH	NOCOM
PUB	**−3.466** (.00)	**−3.466** (.00)	2.505 (.01)
PRIV		**−2.163** (.03)	**−3.578** (.00)
BOTH			**−3.225** (.00)

sus, the identity of the sender, irrelevant for first order beliefs, makes some difference for second order ones. In particular, the players seem to overstate how much their messages are trusted.

FIGURE 12: SECOND ORDER BELIEF CONDITIONAL ON EXCHANGED MESSAGES
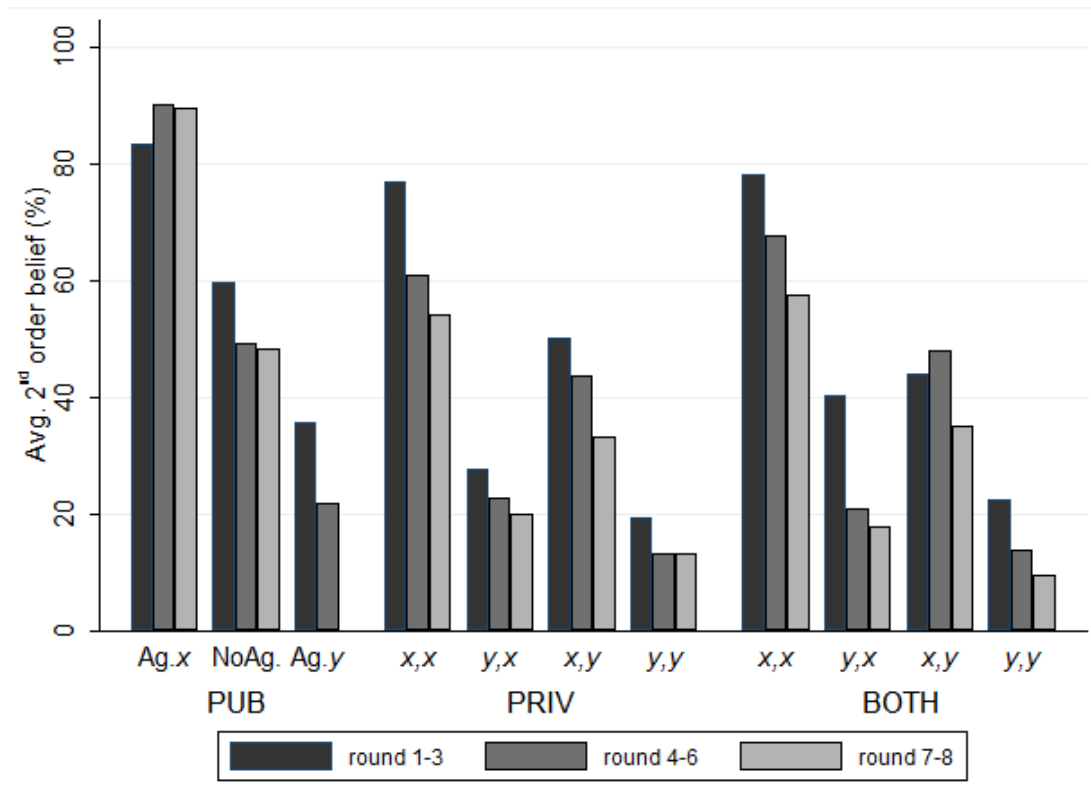
Tables 11 and 12 report the tests for differences in the distributions of first and second order beliefs. Across treatments, the only significant differences are recorded between *Public* and any of the other treatments. This holds for both priors and posteriors. The table also reports on the differences between

42

TABLE 11: ACROSS AND WITHIN TREAT. DIFFERENCES IN **FIRST** ORDER BELIEFS

| | Before comm. (KST) | | | After comm. (KST) | | | After VS before | |
	PRIV	BOTH	NOCOM	PRIV	BOTH	NOCOM	WSRT	ST
PUB	.779	0.875	0.875	.779	1.000	0.875	−0.980	
	(.005)	(.001)	(.001)	(.005)	(.000)	(.001)	(.327)	(.289)
PRIV		0.333	0.222		0.222	0.222	−1.362	
		(.593)	(.960)		(.960)	(.960)	(.173)	(.508)
BOTH			0.333			0.222	−**2.667**	
			(.593)			(.960)	(**.008**)	(**.004**)

prior and posterior beliefs within each treatment. Both the Wilkoxon signed-ranks test (WSR) and the sign test (ST)[35] fail to find a significant difference in treatments *Public* and *Private*. This may be due to the fact that communication shifts beliefs toward more extreme values, so that positive and negative differences between priors and posteriors compensate each other, leaving the median unchanged.

TABLE 12: ACROSS AND WITHIN TREAT. DIFFERENCES IN **SECOND** ORDER BELIEFS

| | Before comm. (KST) | | | After comm. (KST) | | | After VS before | |
	PRIV	BOTH	NOCOM	PRIV	BOTH	NOCOM	WSRT	ST
PUB	.779	0.779	1.000	.779	1.000	1.000	0.560	
	(.005)	(.005)	(.000)	(.005)	(.000)	(.000)	(.575)	(1.00)
PRIV		0.222	0.222		0.222	0.222	−1.007	
		(.960)	(.960)		(.960)	(.960)	(.314)	(.508)
BOTH			0.333			0.222	−**2.667**	
			(.593)			(.960)	(**.008**)	(**.004**)

Theoretically, communication should not have a role on the choice of the subjects beyond its effect through beliefs: for a given level of posterior beliefs, having observed a y message should not affect the actual choice. To test for this, we run a panel linear probability model (LPM) with random effects on the actions chosen. The model can be written as:

$$P(a_{it} = Y \mid, B_{it}, C_{it}, Z_{it}) = \beta_0 + B_{it}\beta_B + C_{it}\beta_C + Z_{it}\beta_Z + \nu_{it} + \eta_{it}$$

where Z_{it} is a set of controls, B_{it} refers to the beliefs of individual i in round t, C_{it} to the outcome of communication, and ν_{it} and η_{it} are the between and the within error term, respectively.

To understand if and how communication affects actual play at different levels of posterior beliefs, we first build a single measure for the level of

---

[35]We use the sign test, on top of WSR test, because the distributions look hardly unimodal.

first order beliefs, defined as the product between the two stated posteriors.[36] Second, we transform it into a categorical variable that takes value low (L), medium-low (ML), medium-high (MH) and high (H), when the level of beliefs belongs, respectively, to the first, second, third and fourth quartile of its distribution. High beliefs (approximately) correspond to action X being the best response. The best response is Y at lower levels of beliefs.[37] We include the already used dummy for the outcome of the communication phase, and interact it with the treatments, and with the level of beliefs.

Table 13 reports the (aggregate) coefficients for each possible combination of the interacted terms. The baseline combination corresponds to treatment *Private*, MH beliefs,[38] and observing no y message. Each coefficient represents the spread in the probability of taking action Y with respect to this base combination.

The comparison of coefficients in the same column represents the effect of a change in the level of beliefs, given the treatment, and the outcome of communication. The comparison of adjacent coefficients in the same row represents the effect of communication, within one treatment, for a given level of beliefs. The treatment effect is measured by comparing coefficients in the same row, given the content of communication. Apparently, communication affects actions directly, on top of its effect through beliefs, and this happens consistently across treatments.[39]

We plot the mean response of the dependent variable against different combinations of the interacted variables in Figure 13. Under best response to stated beliefs, the gray and the black points would overlap. There would be no treatment differences, and the mean response would be 0 at high levels of beliefs, 1 otherwise. The gap between the gray and the black lines then represents the direct effect of communication on actions.

The across-treatment comparisons, conditional on having observed some y (gray), show hardly any difference, qualitatively matching the best response

---

[36]The results are robust to different specifications of this measure (e.g. the minimum or the sum between the posteriors). The product captures the idea that, for choosing X, a player must believe that both partners are playing X with a sufficiently high probability; it has also the advantage that, at some level, it approximates pretty well the locus of beliefs where the players should be indifferent between action X and Y.

[37]We choose to use a categorical variable for beliefs precisely because, theoretically, the relation between beliefs and actions is not linear or continuous.

[38]The arithmetic mean of the original continuous variable falls in category MH.

[39]A probit version of the model gave similar results, but the interpretation and the development that follows are much more immediate using a LPM. The estimates from this model, from other specification of the beliefs variables, as well as those from treatment specific models that include a richer set of communication variables are available upon request.

TABLE 13: ACTIONS: COEFFICIENTS FOR THE INTERACTION BELIEF QUAR-TILE*TREATMENT*COMMUNICATION DUMMY

Belief level	PUB Observes y No	PUB Observes y Yes	PRIV Observes y No	PRIV Observes y Yes	BOTH Observes y No	BOTH Observes y Yes
L	.265* (.1261385)	.279* (.128)	.297** (.112)	.297* (.120)	.242* (.122)	.304* (.121)
ML	-.043 (.176)	.343** (.129)	.319* (.127)	.301* (.120)	.357** (.121)	.320* (.123)
MH	-.313* (.148)	.210 (.133)	BASE	.301* (.121)	-.236 (.159)	.246 (.133)
H	-.538** (.136)	-.135 (.154)	-.585** (.127)	.053 (.273)	-.560** (.134)	-.304 (.188)

Standard errors in parentheses (clustered at the matching group level)

* $p < 0.05$, ** $p < 0.01$

FIGURE 13: INTERCATION PLOT: MEAN RESPONSES FROM MODEL WITH CATEGORI-CAL BELIEFS (QUARTILES)
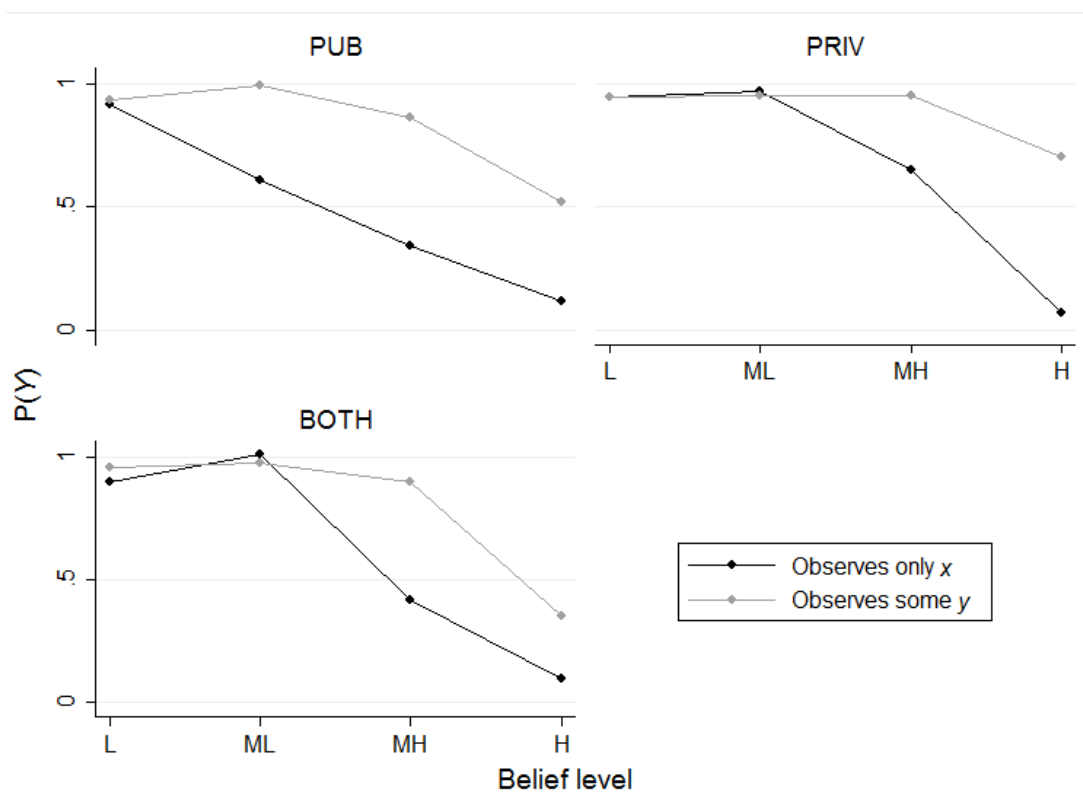
benchmark.[40]

The black points differ from the gray ones at intermediate levels of beliefs (MH level in *Private* and *Both*, ML and MH in *Public*). They are close to the best response benchmark at extreme levels. Thus, the content of communication affects actions, given beliefs, when players report uncertainty on the others' choices.

Incorrectly reported beliefs or departure from best response are alternative explanations of this result. We have however no way to disentangle the two. Players may incorrectly report beliefs if they are risk averse, for instance. In that case, agents would report beliefs biased towards 50 percent, so that in intermediate categories we would have subjects with different "true" beliefs. If the latter are related to the content of communication, then this may separate between different "true" levels of beliefs, even when the reported ones fall in the same category. Departures from best response may occur, for instance, for reasons of guilt or regret aversion. A player holding the beliefs that others will not play X with sufficiently high probability, may nevertheless end up playing X if he observes only x messages, because he does not want to deceive the others in case they prove sincere.

---

[40]The only exception arises at high beliefs (H). However, note that those estimates come from subjects declaring they are almost sure their partners will both play X, after having observed some y in the communication phase. They are based on a tiny number of observations, from subjects with either a peculiar updating process or an incorrect belief reporting strategy. Thus, we think they should not be taken too seriously.