The Power and Limits of Sequential Communication in Coordination Games

Simin He*, Theo Offerman†, and Jeroen van de Ven‡

January 11, 2019

Abstract

We study theoretically and experimentally the extent to which communication can solve coordination problems when there is some conflict of interest. We investigate various communication protocols, including one in which players chat sequentially and free-format. We develop a model based on the 'feigned-ignorance principle', according to which players ignore any communication unless they reach an agreement in which both players are (weakly) better off. With standard preferences, the model predicts that communication is effective in Battle-of-the-Sexes but futile in Chicken. A remarkable implication is that increasing players' payoffs can make them worse off, by making communication futile. Our experimental findings provide strong support for these and some other predictions.

Keywords: mixed-motive games, sequential communication, feigned-ignorance principle.

JEL codes: C72, C92.

---

*Key Laboratory of Mathematical Economics and School of Economics, Shanghai University of Finance and Economics, 111 Wuchuan Rd, 200433 Shanghai, China. E-mail: he.simin@mail.shufe.edu.cn.

†CREED, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, Netherlands; and Tinbergen Institute, 1082 MS Amsterdam, Netherlands. E-mail: t.j.s.offerman@uva.nl.

‡Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, Netherlands; and Tinbergen Institute, 1082 MS Amsterdam, Netherlands. E-mail: j.vandeven@uva.nl. We benefited from financial support by the Research Priority Area Behavioral Economics of the University of Amsterdam. We gratefully acknowledge the helpful comments of the editor and three referees.

1 Introduction

"Speak only if it improves upon the silence."

–Mahatma Gandhi

Humans have achieved astonishing successes in creating ideas and developing technologies. Most of these accomplishments required some form of coordination between people. Communication is undoubtedly at the heart of such successful coordination. Yet, exactly how and under what kind of conditions people manage to coordinate effectively is still largely an open question, both theoretically and empirically. A main obstacle is that people's objectives are usually not fully aligned. There will often be disagreement over what to coordinate on, even if there are potential benefits of coordination and people can communicate with each other. While it has since long been recognized that communication can help (e.g., Farrell 1987, 1988; Farrell and Rabin 1996), existing theories of communication are focused on restrictive and somewhat unnatural communication forms, and therefore give little guidance over the outcomes we can expect to occur in this class of settings, and whether those outcomes will be efficient.

Our contribution is to develop a theoretical model in which players can alternatingly send messages to each other before making their decisions. A main way in which our model differs from most of the existing literature is the way in which players send messages. Whereas in most models players send messages simultaneously or only one of the players can send a message,[1] in our setting players alternate in sending messages. As discussed in the concluding section of Rabin (1994), there are several advantages to our approach. Simultaneous communication is at variance with how people normally communicate. It also introduces another coordination problem since messages can be conflicting. One-way communication gives too much power to the player that can send a message. Especially in coordination games with partially conflicting objectives, it seems reasonable to allow both players to express their agreement or disagreement. The experimental evidence shows that these concerns are potentially very relevant (see later). Our model also differs from many other models in that we assume communication is costly and players can choose when to end the communication process. This captures the opportunity cost of time spent communicating, and prevents players from talking forever.

To study how people will communicate, we need to make some behavioral assumptions about the link between communication and actions. The main assumption we introduce is

[1]See for instance Farrell (1987, 1988); Rabin (1994), and Costa-Gomes (2002).

Figure 1: Payoff structures

	H	L
H	0,0	200,50
L	50,200	0,0

A: Battle-of-the-Sexes

	H	L
H	0,0	200,50
L	50,200	150,150

B: Chicken

the '*feigned-ignorance principle*'.[2] According to this principle, an agreement reached in the communication stage is only effective if following the agreement is a Pareto-improvement compared to the expected payoffs without communication. If at least one player would be better off by ignoring (or pretending to ignore) the conversation, both players will play according to the outcome that is focal without communication. Which point is focal may depend on the structure of the game, and can be empirically determined.

Our model identifies conditions under which communication will help. With standard preferences, the main prediction of the model is that communication will result in successful coordination in a Battle-of-the-Sexes game (like in Figure 1A), but will be futile in a Chicken game (like in Figure 1B). The reason is that the pure strategy equilibria in a Battle-of-the-Sexes game yield higher expected payoffs to both players compared to the mixed strategy equilibrium, which in this game we expect to be the focal equilibrium without communication (see Farrell 1995). By contrast, in a game of Chicken, the payoff of a player's least preferred pure strategy equilibrium is worse than the expected payoff of the mixed strategy equilibrium. Thus, in a Battle-of-the-Sexes game, players will want to listen to each other, but in a game of Chicken at least one of the players will want to ignore the conversation. We also examine the case in which players are lying-averse (e.g., Vanberg 2008). In that case, we show that players in a game of Chicken may agree on playing the "both chicken strategy" (both playing strategy "L" in figure 1B), but they will not always conform to the agreement.[3]

We put our theoretical predictions to an experimental test. The results support the main predictions. In a Battle-of-the-Sexes game, communication is very effective and

[2]We thank Gary Charness for suggesting this terminology.

[3]Ellingsen et al. (2018) pursue a different methodological approach. They suppress psychological motivations like lying aversion by encouraging subjects to pursue their self-interest. An advantage of that approach is that it may shed more light on the accurateness of underlying game-theoretic solution concepts that assume that the researcher knows subjects' preferences. Our approach has the advantage that it allows us to investigate the effectiveness of communication in the natural situation where subjects pursue their own goals.

helps players to coordinate. Typically, players coordinate immediately on the first sender's preferred equilibrium. This result resonates well with the existing experimental evidence using different communication formats (Cooper et al. 1989 and Duffy and Feltovich 2002). More surprising and novel is our finding that communication is largely ineffective in the game of chicken. Subjects also appear to anticipate the ineffectiveness of sending messages, and frequently forgo the option to communicate at all. As a consequence, and consistent with the theoretical predictions, higher values of the payoffs associated with (L, L) can make subjects worse off by making communication futile.

As subjects in our experiments can send free-form messages, we are also able to analyze their contents in more detail.[4] The analysis tells us that first-senders frequently express an intention to play H in the Battle-of-the-Sexes, and this happens much less often in the game of Chicken. This is consistent with the comparative statics predictions of our model. With higher (L, L) payoffs, we find that subjects frequently reach an agreement to both play L. In agreement with the equilibrium that allows for lying aversion, we find that subjects play L more often after agreeing on (L, L), but the effect is small and subjects still often choose H. The data also support the prediction that more players conform to the agreement when the payoffs of (L, L) are larger.

In a follow-up experiment, we implement two alternative communication formats: one-sided costless communication and free-form costless chat. For one-sided communication, we can derive theoretical predictions under the 'feigned-ignorance' principle, i.e., assuming that players will ignore any messages that, if followed, do not yield a Pareto-improvement over the focal strategies without communication. Also for this case the model predicts that with standard preferences communication is very effective in a Battle-of-the-Sexes game but not in a Chicken game. The experimental results are in line with these predictions. Communication is also very effective in a Battle-of-the-Sexes game if players can chat free-format. This shows that the effectiveness of sequential communication is not driven by the imposed asymmetry that results from our assignment of the first sender. We also find that free-format chat is somewhat effective in a Chicken game. With free-format chat, participants communicate more often and the conversations are more

[4]We used free-format communication, because in different contexts it has been found that this is more effective in changing behavior than pre-coded messages (Charness and Dufwenberg, 2010; Palfrey et al., 2015). Wang and Houser (2015) find that free-form simultaneous two-way communication is more effective than restricted communication in coordination games, as it allows the possibility to signal attitudes besides signaling intentions. Closest to our work in this respect is Cason and Mui (2015), who find that the possibility of free-form messages is critical for coordinated resistance in a "resistance game."

lengthy.

The importance of our results go beyond a better understanding of how and when communication works. When faced with multiple Nash equilibria, many theorists focus on the set of efficient equilibria. The rationale is that communication would help players to coordinate on an efficient equilibrium (cf. Rabin 1994), even though the communication stage is often not explicitly modeled. Our results provide support for this approach.[5]

The feigned-ignorance principle provides an alternative to the approach taken in, e.g., Farrell and Rabin (1996). They argue that if people talk about intentions, messages that are both self-signaling and self-committing seem especially credible.[6] On the other hand, experimental evidence shows that communication is effective even when the messages are not self-signaling (Charness 2000; Clark et al. 2001; Blume and Ortmann 2007; Brandts and Cooper 2007; Avoyan and Ramos 2016), so that it is an open question when messages need to be self-signaling.[7]

Our theoretical predictions are also quite different from those of the theory developed in Ellingsen and Östling (2010). They study the effect of communication in both the Battle-of-the-Sexes game and the Chicken game by using a level-k model. They predict that one-way communication will powerfully resolve the coordination problem in such coordination games if players have some depth of thinking, even in games like Chicken where our approach predicts communication to be ineffective unless lying aversion plays a sufficient role.

In terms of communication structure, Santos (2000) is closest to our approach. He provides a model of finite sequential cheap talk communication in coordination games. In his game, the two players alternate making costless announcements that may be accepted or followed up by a counterproposal before they make their choices in the coordination stage. With a commonly known final round of communication, all the negotiation power is essentially given to the player who can make the last announcement. In this sense, this model is quite similar to a model of unilateral communication where only one player can

[5] For example, Tirole (1988) comments on equilibrium selection in tacit collusion models in the following way: "The multiplicity of equilibria is an embarrassment of riches. We must have a reasonable and systematic theory of how firms coordinate on a particular equilibrium if we want the theory to be predictive comparative statics. One natural method is to assume that firms coordinate on an equilibrium that yields a Pareto- optimal point in the set of the firms' equilibrium profits" (p.253).

[6] A message specifying an intention is self-committing if the sender wants to follow up on it when she thinks the receiver believes it. A message is self-signaling if the sender wants to send it if and only if the message is true (see Farrell and Rabin 1996).

[7] For some more discussion on this approach, see Ellingsen et al. (2010).

make an announcement. In most actual cases , it is not commonly known on beforehand who will have the possibility to say the last word. Our model seems a better approximation of such conversations.

In terms of experimental work, our results shed light on the existing literature that shows the effectiveness of communication in Batte-of-the-Sexes games with different communication formats (see e.g., Cooper et al. 1989). Cooper et al. (1989) show that one-way communication increases the coordination rate dramatically from 0.48 to 0.95. Two-sided communication is much less effective though. One round of two-way communication raises the coordination rate only to 0.55, and three rounds yields a coordination rate of 0.63. Our findings suggest that the earlier mentioned concerns with the previous communication forms are valid. Our more natural form of communication increases the coordination rate from 0.43 to 0.80. Thus, one way-communication may overstate the effect of communication while two-sided communication underestimates its effect. In the context of Chicken games, the only work we are aware of is that by Duffy and Feltovich (2002), who investigate how one-way pre-coded cheap talk and observations of previous play affects behavior. They find that observations of previous play are more effective than cheap talk to increase coordination in the Chicken game. Duffy and Feltovich (2006) extend the analysis by investigating how results change when subjects' messages can contradict previous actions.

The remainder of the paper is organized in the following way. Section 2 describes the game and the theory. Section 3 presents the experimental design of the sequential communication protocol. Section 4 discusses the experimental results for sequential communication. Section 5 presents the design and results for other communication protocols. Section 6 provides and discussion and Section 7 concludes.

2 Theoretical Background

We investigate the impact of pre-play communication in a two-player simultaneous-move normal-form game G. Each player chooses some action $A_i \in \{H, L\}$, with payoffs $u_i(A_i, A_j)$. In Section 2.1 we consider situations where players talk sequentially. In Section 2.2 we deal with one-sided communication which is a common protocol in the previous literature.

2.1 Two-sided Sequential Communication

Before choosing their actions, there is a pre-play communication stage C. In this section the focus is on two-sided sequential communication. In the corresponding experiment, communication is free-format. Here, we assume a more restricted message space, but one that is rich enough to capture the most important messages. Possible messages are hh, hl, lh, and ll, matching the four strategy profiles, where the first letter indicates player 1's action and the second letter player 2's action. We assume that messages are interpreted as indicating the sender's own intended action and the expected action of the other player. A player can also terminate the communication stage by sending an empty message $\varnothing$. Thus, the set of possible messages for each player is $M_i = \{hh, hl, lh, ll\} \cup \{\varnothing\}$. Players send messages in turns, starting with player 1, where it is randomly determined which of the players is player 1. The communication stage ends as soon as the players reach an agreement or if one of the players sends $m = \varnothing$. The message $m = \varnothing$ is costless, sending any other message costs $\gamma > 0$ to *each* player.[8]

We refer to game G including the communication stage C as the extended game GC. Strategies in the game GC are messages in the communication stage (possibly mixed and contingent on time and the opponent's messages) combined with probability distributions (possibly degenerate) over H and L, where the probabilities can depend on the messages sent in the communication stage. Payoffs in GC are $U_i = u_i(A_i, A_j) - \gamma T$, where T is the total number of non-empty messages sent by both players, and A_i and A_j are the actions chosen in game G by players i and j, respectively. We set $U_i = -\infty$ if the communication stage goes on forever.

We next define what constitutes an agreement. We assume that the players reached an agreement if the communication stage ends with non-conflicting messages. This means that the last two messages (excluding the empty messages that terminates the communication stage) are identical. If the communication stage is terminated after a single message then we also assume that there is an (implicit) agreement. If the communication is terminated immediately (the first player sending message $\varnothing$ before any other message), we assume that there is no agreement.

[8]By making communication costly, our environment has some similarities to bargaining models with costly bargaining, such as Rubinstein (1982). A crucial difference is that in those models the division of the surplus is predetermined once an agreement is reached, while in our setup any agreements reached in the communication stage are not binding. Avoyan and Ramos (2016) also investigate the case where communication involves commitment power. Another possibility would have been to make players pay per time-unit of communication instead of per message (e.g., see Embrey et al. 2014).

In what follows, we use *conversation* to describe all non-empty messages in the communication stage until the communication stage is ended. When a player copies the previous player's message, the communication stage ends automatically and the conversation includes all messages sent. If the communication is terminated by a player sending the empty message, the conversation includes all messages prior to the empty message.

Definition 1. *A* ***conversation*** *is a sequence of non-empty messages* $(m_1, m_2, ... m_T)$ *in the communication stage.*

Definition 2. *Suppose the communication stage is terminated after T non-empty messages. An* ***agreement*** *is reached if the communication stage (i) contains a single non-empty message, or (ii) it has length* $T \geq 2$ *and messages* m_{T-1} *and* m_T *are identical. The agreement is* ***credible*** *if it induces equilibrium strategies in game G.*

Note that an agreement is not binding. Notice also that we only allow agreements on pure-strategy outcomes. In principle, players could also agree on randomizing their strategies.

Our first behavioral assumption is that players will play their focal strategies if no messages are sent or no agreement is reached. In general, it depends on the specifics of the game what the focal strategies are. In some games, it seems natural to assume that the mixed-strategy equilibrium is focal if players end the conversation without reaching an agreement, since they have no way of coordinating (see Farrell 1987). In other games, specific features may make another strategy-pair focal when players cannot communicate.

Our second behavioral assumption is that players will only act in accordance with an agreement if (1) it is credible, and (2) it gives at least as high utility to the players as when they disregard the conversation and play according to the focal outcome of the game G. Thus, they will ignore the conversation if at least one of the players is better off by not listening or pretending not to listen. We label this the 'feigned-ignorance principle.'[9]

We now make the above more precise and summarize the main elements in Assumption 1. Let μ_i represent player i's expected payoff from playing the focal strategy in game G in the absence of communication.

[9] We thus assume that players may ignore an agreement even if it constitutes a pure-strategy equilibrium and one of the players explicitly agreed (by copying the other player's message). The set of equilibrium outcomes is unaltered if we instead assume that players act in accordance with the agreement, as in that case players would refrain from reaching such an agreement.

Assumption 1 (Feigned-Ignorance Principle)**.** *Each player ignores the conversation and believes that the other player will play according to the focal outcome of the game G in the absence of communication, unless the players reach a credible agreement on an outcome* (A_1, A_2) *in which each player earns at least as much as in the focal outcome:* $u_i(A_i, A_j) \geq \mu_i$ *for each player. In the latter case, each player follows the equilibrium strategies* (A_1, A_2) *specified in the agreement.*

Now we zoom in on the specific class of games that we study in our experiments. The payoffs are given in Table 1. We assume that $a > b > 0$ and $a > c$. For $c = 0$, it is a "Battle-of-the-Sexes" game. For $c > b$, it has the structure of a "Chicken" game. The game G has two pure-strategy Nash equilibria, (H, L) and (L, H), and a mixed strategy Nash-equilibrium in which each player randomizes between H and L, playing H with probability $p = (a-c)/(a+b-c)$. A player's expected payoff in the mixed-strategy equilibrium is $ab/(a + b - c)$.

Table 1: Payoff matrix of Game G

		Player 2	
		H	L
Player 1	H	0, 0	a, b
	L	b, a	c, c

Notes: $a > b > 0$ and $a > c \geq 0$.

We call an agreement **demanding** if player i proposed the outcome and it is player i's most preferred outcome. We call an agreement **conceding** if player i proposed the outcome and it is player i's least preferred equilibrium outcome. An agreement to play (L, L) is called a **compromise**. Finally, we say that an **agreement is immediate** if it is not preceded by any other messages in the conversation.

2.1.1 Equilibrium

In this section we characterize the set of equilibrium strategies if players have standard preferences. To narrow down the set of equilibria, we assume a natural language interpretation of messages, such that *hl* corresponds to the intention to reach the outcome (H, L) etc. Given the symmetry of the game, we assume that the mixed-strategy equilibrium is focal if players end the conversation without reaching an agreement. This assumption is largely supported by the data. Thus, $\mu_i = ab/(a+b-c)$ for each player i.

Many potential equilibrium strategies are eliminated under this assumption. In particular, it rules out correlated equilibria, in which the random assignment of players to roles is used as a coordination device.[10] Assumption 1 also restricts admissible beliefs and thereby the set of equilibria. Under the assumption, players always interpret certain (sequences of) messages as an agreement or disagreement. For instance, when a player sends *hl* and the other player responds with the same message, an agreement is reached, and players will not interpret this sequence of messages as mere babbling or disagreement.

The following proposition presents the subgame perfect equilibria (SPE) in which the players use pure strategies in the communication stage (see Appendix A for proofs). The cost of sending a message determines the type of equilibrium. Define the thresholds as $\gamma_1 = \frac{b(b-c)}{a+b-c}$, $\gamma_2 = \frac{a(a-c)}{a+b-c}$, and $\gamma_3 = \min\{\frac{1}{2}\gamma_1, \frac{1}{3}\gamma_2, \frac{1}{2}(a-b)\}$.

Proposition 1 (Pure strategy equilibria). *Under Assumption 1, if players do not randomize in the communication stage, the only SPE outcomes of game GC for $b \geq c$ are: (i) immediate concession ($m_1 = lh, m_2 = \varnothing$) followed by (L,H) exists if and only if $\gamma \leq \gamma_1$; (ii) immediate demanding ($m_1 = hl, m_2 = \varnothing$) followed by (H,L) exists if and only if $\gamma \leq \gamma_2$; (iii) delayed demanding ($m_1 \in \{hh, ll\}, m_2 = hl, m_3 = hl, m_4 = \varnothing$) followed by (H,L) exists if and only if $\gamma \leq \gamma_3$, (iv) immediate termination ($m_1 = \varnothing$) followed by the mixed strategy equilibrium of game G exists if and only if $\gamma \geq \gamma_1$. When $b < c$ or $\gamma > \gamma_2$, communication is ineffective and players refrain from sending costly messages.*

Note that immediate termination and delayed demanding are Pareto-dominated by immediate demanding in terms of expected payoffs.

A property of the above equilibrium conversations is that they quickly result in agreement. Under some conditions, there also exists an equilibrium in which players potentially take a long while before they reach an agreement. In the communication stage of such an equilibrium, players are indifferent between conceding (and get the low payoff b) and demanding in the hope that the other player will concede (possibly getting the high payoff

[10]For instance, the first player could terminate the communication stage immediately, and both players could then believe that the first player will choose H and the other player will choose L. Assumption 1 rules this out by specifying that the mixed-strategy equilibrium is played after immediately terminating the communication stage (something that is supported by the data).

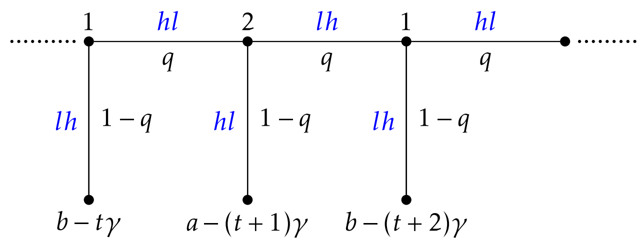

Figure 2: Part of a possible communication stage tree when period ≥ 4.

a but at the cost of sending more messages). In the following proposition, we characterize this equilibrium in which players use mixed strategies in the communication stage.

Let $q_1 = \frac{a-b}{a-b+2\gamma}$, $q_2 = \frac{a-b-2\gamma}{a-b+\gamma}$, $q = \frac{a-b-\gamma}{a-b+\gamma}$ and $N = 1 + 2q_1^2 + \frac{q_1^2 q_2}{1-q}$.

Proposition 2 (Mixed strategy equilibrium)**.** *Under Assumption 1 and for $\gamma \leq \gamma_1$, there exists a mixed strategy equilibrium in which players randomize in the communication stage C. In C, players mix between a demanding message and a conceding message in each period. In the first two periods, both players are demanding with probability q_1. In the third period, player 1 is demanding with probability q_2. From the fourth period, each player is demanding with probability q whenever it is her turn to send a message.*[11] *The player that concedes plays L in game G, the other player plays H in game G. The expected length of the conversation is N messages.*

We do not think that players will literally randomize at each instance where they can send a message. Instead, the equilibrium described in Proposition 2 may approximate a situation where players at the start of the communication stage decide to test whether the other will concede, and play a mixed strategy with regard to the maximum number of periods in which they are willing to send a message *H* before they concede themselves. They can determine this maximum before they start communicating. If the mixed strategy for the maximum agrees with the randomization process described in Proposition 2, an equilibrium results in which players test whether the other will concede.

The following corollary is an immediate implication of the above propositions.

[11] The initial phase of periods 1, 2 and 3 differs from the remainder of the game because player 2 in period 2 has the possibility to concede by simply terminating the communication. In the other periods, to avoid conflicting messages, players concede by sending the costly message *lh* (for player 1) or *hl* (for player 2). We assume that player 1 mixes with probability q_1 in the first period, though any probability is supported in equilibrium.

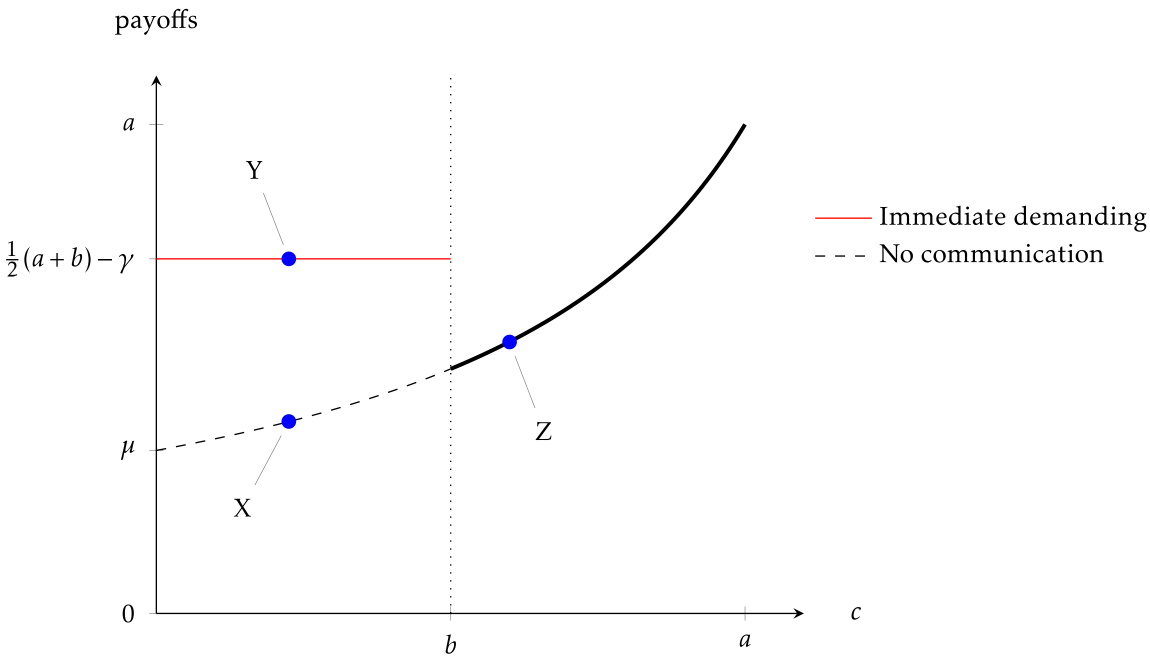

Figure 3: mean payoffs without communication and in the 'immediate demanding' equilibrium. Notes: the figure is drawn for $a-b>\gamma$ (implying that $\gamma<\gamma_2$ at $b=c$). For $c>b$ there are no equilibria in which players send messages.

Corollary 1. *For $c>b$, costly communication cannot be supported in equilibrium.*

The reason behind the result in this corollary is that players will anticipate that the player who is worse off after communication prefers to ignore the communication and to play the mixed-strategy equilibrium ($\mu_i>b$ for any $c>b$).

Given that communication can help for $b\geq c$ but not otherwise, it is possible that players can be worse off for a higher value of c, because communication becomes futile. This is illustrated in Figure 3, which shows the expected payoffs in the equilibrium with immediate demanding and without communication for different values of c. For relatively low values of c, the equilibrium in which sender 1 is demanding exists, increasing mean payoffs compared to a situation where communication is not possible (from X to Y, for instance). For high values of c, no equilibria exist in which players communicate. Without communication, payoffs are higher for higher values of c (compare Z to X). With communication, an increase in c can decrease mean payoffs, because communication becomes ineffective (compare Z to Y).

2.1.2 Extension: lying aversion

So far we only considered the direct costs of sending messages. Several studies show that there can be psychological costs related to talking. In particular, many people do not break promises because of lying aversion or guilt aversion.

In this section we analyze the effects of lying aversion and show that it expands the set of credible agreements. We assume that a player experiences a cost of lying when the player deviates from an agreement that will not be ignored by the other player (in accordance with the 'feigned-ignorance principle'). The cost of lying is independent of the other player's actions; a player incurs lying cost after deviating from an agreement that satisfies the feigned-ignorance principle even if it later turns out that the other player also deviated from the agreement.

Following Ellingsen and Johannesson (2004), we model the cost of lying as a cost k_i that is subtracted from a player's payoff u_i.[12] We allow for heterogeneity in lying costs. We assume that k_i is drawn independently from a continuous and strictly increasing cumulative distribution function $F(\cdot)$ that has full support on $[0,\bar{k}]$, and this is common knowledge. Each player i is only privately informed of her own k_i at the start of the game. We refer to this imperfect information version of game G as game BG and to the extended game with the communication stage as game BGC. A player i's payoff in BGC is given by $U_i = V_i - \gamma T$, where V_i is player i's payoff in BG that is potentially affected by costs of lying. We again set $U_i = -\infty$ if the communication state goes on forever.

A pure strategy in game BG given an agreement on outcome (O_1, O_2) from the set $\{(H,H),(H,L),(L,H),(L,L)\}$ is a function $A_i(k_i)$ from $[0,\bar{k}]$ into $\{L,H\}$. Player i's payoff in BG given the actual choices (A_1, A_2) and the agreement on (O_1, O_2) is given by:

$$V_i = u_i(A_i, A_j) - \mathbb{1}_{A_i \neq O_i} k_i,$$

where $\mathbb{1}_{A_i \neq O_i}$ is an indicator function taking value 1 if there is an agreement that satisfies feigned ignorance and $A_i \neq O_i$, and it takes value 0 otherwise. A Bayesian equilibrium is a pair of strategies $(A_1^*(\cdot), A_2^*(\cdot))$ such that for each player i and k_i strategy $A_i^*(k_i)$ maximizes player i's expected payoff $\tilde{V}_i(k_i) = \mathbb{E}_{k_j} V_i(A_i, A_j^*(k_j), k_i, (O_1, O_2))$.

To characterize the equilibrium set in the extended game with communication (game BGC), we modify the 'feigned-ignorance principle' to reflect the uncertainty.

[12]We thus model the cost of lying as a fixed cost, independent from expectations held by the other player. Expectation-based models include, for instance, Charness and Dufwenberg (2006). Direct tests favor the fixed cost approach, see e.g. Vanberg (2008) and Di Bartolomeo et al. (2018).

Assumption 1′ (Feigned-Ignorance Principle in Bayesian games)**.** *Each player ignores the conversation and believes that the other player will play according to the focal outcome of the game BG in the absence of communication, unless the players reach an agreement on an outcome* (O_1, O_2) *that is supported by a unique Bayesian Nash equilibrium* $(A_1^*(\cdot), A_2^*(\cdot))$ *of game BG and in which* $\tilde{V}_i(k_i) \geq \mu_i$ *for each type of each player. In the latter case, each player follows the equilibrium strategies* $(A_1^*(\cdot), A_2^*(\cdot))$ *of game BG that correspond to the outcome on which the agreement was made.*

Like before we assume that the focal outcome in the absence of communication is the mixed strategy equilibrium. Notice that we define feigned ignorance for a class of Bayesian games where each agreement on one of the outcomes is supported by a unique Bayesian Nash equilibrium. The definition can readily be extended to Bayesian games in which an agreement on an outcome is supported by multiple equilibria. In such cases, an agreement should not only specify the outcome but also the equilibrium on which players coordinate (which would require a larger message space).

In this setting, players understand that not all types of players will necessarily respect an agreement. Players with a relatively low cost of lying may deviate, while those with a high cost of lying will conform to the agreement. This is a feature of the most interesting case, in which players agree on the outcome (L, L). Introducing a cost of lying does not affect any of the equilibria previously characterized for the case of complete information, as in those equilibria players act in accordance with the agreement reached.[13]

We now show that there can be an equilibrium in which all types of sender 1 immediately propose ll and all types of sender 2 accept the agreement by terminating the communication stage. After such an agreement on the outcome (L, L), the equilibrium strategy in game BG on whether to conform or deviate from the agreement is characterized by a threshold strategy, such that players with a cost of lying lower than some threshold k^* deviate from the agreement (choose H) and players with a cost of lying higher than k^* conform to the agreement (choose L). After reaching an agreement, players then expect the other player to deviate from the agreement with probability $F(k^*)$. A player i who has a lying cost k_i is then indifferent between L and H when:

$$F(k^*)b + (1 - F(k^*))c - \gamma = F(k^*)(-k_i) + (1 - F(k^*))(a - k_i) - \gamma \tag{1}$$

[13] Once an agreement is reached, the equilibria that we previously derived can still be supported with lying costs as players have no incentives to deviate even without lying costs. The threshold values of γ are also unaffected, as they are determined by possible deviations to other messages in the communication stage, and players have no lying costs from sending any message as long as no agreement exists.

The LHS of equation (1) is the expected payoff of playing L after an agreement on (L, L), the RHS of equation (1) is the expected payoff of playing H after an agreement on (L, L). The threshold level k^* must satisfy (1) with equality. This gives:

$$F(k^*) = \frac{a - c - k^*}{a + b - c} \tag{2}$$

Note that $F(0) < \frac{a-c}{a+b-c}$ and $F(\bar{k}) > \frac{a-c-\bar{k}}{a+b-c}$. Given that both functions are continuous, $F(k^*)$ is increasing in k^*, and $\frac{a-c-k^*}{a+b-c}$ is strictly decreasing in k^*, there exists a unique solution k^*.

A condition that must be fulfilled is that players find it more profitable to agree on (L, L) than to avoid any conversation. Since players with $k_i > k^*$ will not lie, and players with $k_i < k^*$ have lower costs of lying than a player with $k_i = k^*$, the player with $k_i = k^*$ has the largest incentive to deviate. A sufficient condition is therefore that a player with $k_i = k^*$ in the role of first sender does not want to deviate to sending no message:

$$F(k^*)b + (1 - F(k^*))c - \gamma \geq \frac{ab}{a + b - c}. \tag{3}$$

Note also that an agreement on (L, L) cannot occur in equilibrium for $b > c$. The first sender can secure a payoff of b in game BG by sending the conceding message lh. For $b > c$, this payoff is strictly larger than the expected payoff of conforming to an agreement, which is between b and c.

Finally, we need to verify that players will not ignore an agreement. Whether or not an agreement is ignored is again dictated by the 'feigned ignorance principle'. This assumption is automatically satisfied when condition 3 holds.

The following proposition identifies the conditions under which an immediate agreement to play (L, L) can be sustained in equilibrium.

Proposition 3. *Under Assumption 1′, an immediate agreement to play (L, L) can be supported in a Perfect Bayesian Equilibrium conversation outcome of game BGC if (i) $c > b$, and (ii) $\gamma \leq c - \frac{ab}{a+b-c} - (c - b)F(k^*)$, where k^* is the solution to $F(k^*) = \frac{a-c-k^*}{a+b-c}$. After reaching an agreement to play (L, L), players with a cost of lying lower than k^* choose H and players with a cost of lying above k^* choose L.*

What is the effect of c on the probability that players deviate from the agreement? From equation (2), it follows that k^* decreases as c increases. Thus, given any agreement on (L, L), players are less likely to deviate. However, the impact of c on the likelihood of reaching an agreement is ambiguous; *a priori* it is not clear whether a larger c relaxes constraint (3).

Corollary 2. *Given any agreement on (L,L), a larger value of the joint concession payoff c decreases the likelihood that players deviate. The effect of c on reaching an agreement is ambiguous.*

2.1.3 Summary

Below we list the main predictions of the model with two-sided, sequential communication.

1. Without communication, payoffs are increasing in c. With communication, players can on average be better off with lower values of c.
2. When players communicate, the conversation does not end in disagreement.
3. When players do not reach an agreement, or when the (expected) agreement payoff is worse than the mixed strategy equilibrium payoff for at least one of the players, they play in accordance with the mixed-strategy Nash equilibrium.
4. In the Battle-of-the-Sexes game, communication allows players to coordinate on a pure equilibrium. The conversation may be short and then either the first or second mover is favored, or it may be long and then the potential gains of communication are partially wasted.
5. In the Chicken games, communication is either ineffective and not used or players agree on (L,L). In the latter case, players will sometimes conform to the agreement and the extent to which they deviate from the agreement decreases with the joint concession payoff c.

2.2 One-sided Communication

In this subsection we study the effects of one-sided communication, which was a common communication protocol in previous experiments. In this case, one of the players is the sender and the other player is the receiver. The sender can choose a message from the same set as before, that is, one of the four strategy profiles or the empty message. The communication stage ends after the sender's message. In order to stay close to protocols used in other experiments, we assume that sending a message is costless.

We again apply the logic of the 'feigned-ignorance principle' to derive model predictions. Without having the option to send a message, the receiver cannot explicitly signal agreement or disagreement. We assume that the sender's message still counts as an agreement.

Under the above assumptions, it is straightforward to show that (without a cost of lying), the sender will send *hl* whenever $b \geq c$. When $b < c$, the receiver will ignore any message, and the sender's message is undetermined.

Proposition 4. *Under one-sided communication with costless messages, when $b \geq c$ the sender will be demanding and this is followed by the Nash-equilibrium outcome (H, L). If $b < c$, any message by the sender is ineffective and players follow the mixed-strategy equilibrium of the game G.*

Lying aversion does not change the predictions for one-sided communication. When $b \geq c$, the sender secures the best possible outcome by sending a message that does not require a lie. When $b < c$, a lying averse sender may consider to propose the outcome (L, L). However, given that receiver has no chance to express disagreement in any way, it is natural to assume that he does not experience lying aversion when he plays H. Given that no type of receiver will suffer costs of lying, senders will not benefit from making the proposal. Receivers will ignore the proposal unless a sufficient mass of senders chooses L after the proposal. In the latter case, receivers will play H for sure, and the senders will be worse off compared to no communication.

3 Experiment I: two-sided sequential communication

3.1 Treatment design

In the experiment, we implemented the payoff matrix of Table 1. Table 2 summarizes the different treatments of Experiment I. Payoffs were presented in points. We always set $a = 200$ and $b = 50$, and varied the value of c. For $c = 0$, the game reduces to a Battle-of-the-Sexes game (treatment *BoS*). For $c = 75$ (treatment *C-Small*) and $c = 150$ (treatment *C-Large*) it has the structure of a Chicken game. Subjects simultaneously made a choice between H and L.

In the communication condition, the game was played after one of the players ended the conversation. In that condition, subjects could send free-form messages to each other. Each subject in a pair had to pay a cost $\gamma = 2$ for every message that was sent, no matter who sent the message. It was randomly determined which subject in a pair would be the first sender. After that, they alternated. They could only end the communication if it was their own turn to send a message.[14]

Each subject participated in only one of the treatments. They played the game for 20 rounds: 10 in the condition without communication and 10 in the condition with communication. We changed the communication condition every five rounds, balancing the condition in which they started. This gives a 3x2 design: three treatments (between-subject) and two communication conditions (within-subject).

Subjects were rematched to a different opponent in every round, and were informed that they would never meet the same opponent twice within each communication condition. At the end of each round, each subject received feedback about the decision of the other person and her own payoff.

At the end of the experiment, we administered a short survey, collecting some background information. 4 out of the 20 rounds were then randomly selected for payment. Subjects also received a starting capital of 300 points to cover any possible losses. Every point was worth €0.025.

Table 2: Overview of Treatments

Treatment	Parameter values			
	a	b	c	γ
BoS	200	50	0	2
C-Small	200	50	75	2
C-Large	200	50	150	2

Notes: a, b, and c correspond to the payoff matrix in Table 1. γ is the cost per message (to each sender).

3.2 Procedures

Experiment I was conducted in the CREED laboratory of the University of Amsterdam. A total of 288 subjects were recruited from the CREED database. We conducted 13 sessions

[14]Before running the main experiment, we ran a few pilot sessions (with 48 subjects) of *BoS*. In those sessions, we had a higher cost per message ($\gamma = 5$ instead of 2) or a lower value of a ($a = 75$ instead of 200). Coordination rates were high and subjects sent very few messages. To make sure that these results were not driven by high message costs or small losses of coordinating on one's least preferred equilibrium, we adjusted the values.

with 12 or 24 subjects each. Treatments were randomized at the session level. Each treatment had 96 subjects. Subjects were divided into matching groups of 12 subjects, so that we have 8 independent matching groups per treatment. 48% of subjects were female, and approximately 68% were majoring in economics or business.

The experiment was computerized using PHP/MySQL and was conducted in English. Subjects were randomly assigned to a cubicle. Instructions were given on their screen (see Appendix B for the instructions). They also received a hardcopy sheet with a summary of the instructions. Subjects could not continue until they correctly answered a set of test questions. The same experimenter was always present during the experiment.

Subjects received their earnings in private. Mean earnings were €16.40. A session lasted between 45 and 65 minutes.

3.3 Coding of messages

Three research assistants independently coded the messages on several dimensions. Coders were asked to code if a subject expressed an intention to play H or L, making a distinction between strong and weak expressions of intentions. An expression is considered strong if the sender emphasizes that this is definitely what he or she will do. We also asked coders if a pair of subjects reached an explicit agreement on the outcome (L, L) or any of the outcomes (L, H) or (H, L). To classify a conversation as an explicit agreement, we used the criteria that: (*i*) senders were aware of each other's intentions, and (*ii*) they showed some approval or confirmation. The exact coding instructions can be found in Appendix B.

Coders were not informed of the hypotheses that we were testing. Each coder coded all 1009 conversations. At the end, 50 randomly selected conversations were shown again and recoded, to check each coder's individual consistency. The intra-rater consistency is very high. If we combine the weak and strong expressions into a single category, then each rater gives the same assessment in the retest question as in the original question in at least 48 out 50 cases. The inter-rater consistency is also very high. The values of kappa (a measure of inter-rater consistency) is between 0.89 and 0.93 for the different categories, which is commonly regarded as excellent. In our analysis, we classify messages according to the majority of coders. If all coders disagreed with each other, we treat the conversation as missing value. This is the case for 32 out of 1009 cases.

It took coders roughly eight to ten hours of work to complete the task. They worked at their own pace, taking breaks as they saw fit, and were paid a flat amount of €120.

4 Experimental Results: Sequential Communication

4.1 Effects of Sequential Communication

Without sequential communication, the mean proportion of *H*-choices and corresponding payoffs are fairly close to the mixed-strategy equilibrium outcome in all three games. Figure 4 shows the actual payoffs (solid line) and the theoretically predicted payoffs if subjects play the mixed-strategy equilibrium (orange dots). As expected, mean payoffs are increasing in the value of c: In *BoS* ($c = 0$) mean payoffs are 54, in *C-Small* ($c = 75$) mean payoffs are 73, and in *C-Large* ($c = 150$) mean payoffs are 111.

While communication is very effective in *BoS*, it is futile in *C-Large* and *C-Small*.[15] For the Chicken games, mean payoffs remain the same when subjects have the opportunity to communicate (see the dashed line in Figure 4). In *BoS*, the opportunity to communicate increases payoffs considerably, from 54 to 98, an increase of 82 percent ($p < 0.001$, two-sided Mann-Whitney test).[16] This increase is so large that with communication subjects are on average better off in *BoS* than in *C-Small* ($p = 0.003$). Consequently, mean payoffs are no longer monotonically increasing in the value of c when subjects have the option to communicate. These results are in line with the theoretical prediction that communication is effective in *BoS* (where $c < b$) but not in the Chicken games (where $c > b$), at least if lying costs are sufficiently small.

[15] Strictly speaking, it is not necessarily communication *per se* that is effective, but having the *option* to communicate. As a shorthand we will not make this distinction in most of the text.

[16] Unless specified otherwise, tests reported are based on taking the matching group as the independent unit of observation.

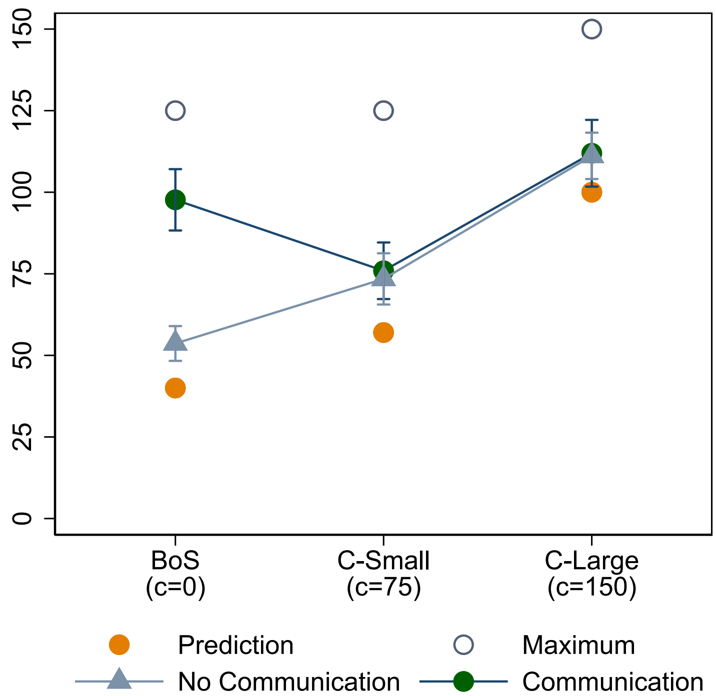

Figure 4: Mean payoffs by treatment and two-sided sequential communication. "Prediction" is the expected payoff if subjects play the mixed-strategy Nash equilibrium of game G. "Maximum " is the maximum mean payoff. Bars are the 95% confidence intervals.

Table 3: Percentage of times positive payoffs and efficiency

Treatment	No Communication		Communication		Test Difference
	Payoffs> 0	Efficiency	Payoffs> 0	Efficiency	(Payoffs> 0)
BoS	43	43	80	78	$p < 0.001$
C-Small	65	59	67	61	$p = 0.673$
C-Large	83	74	84	75	$p = 0.710$

Notes: Communication protocol is two-sided sequential communication. Payoffs> 0 indicates the percentage of times that actions led to positive net payoffs (including communication costs). Efficiency is percentage of maximum joint payoffs. p-values are based on two-sided Mann-Whitney tests.

Communication increases mean payoffs and efficiency in *BoS* because it allows subjects to coordinate their actions on outcomes with positive payoffs, i.e., outcomes (H, L) or (L, H).

This is shown in Table 3. Without communication, subjects end up with positive payoffs roughly 43 percent of the time. With the option to communicate, they coordinate on outcomes with positive payoffs 80 percent of the time. By contrast, coordination rates on positive outcomes in the Chicken games (i.e., (H,L), (L,H), or (L,L)) are unaffected by the option to communicate. Similar patterns apply to the achieved efficiency (earnings in a pair relative to the maximum joint payoffs). The efficiency is highest in *BoS* with communication, although it is not significantly higher than in *C-Large* ($p = 0.529$).[17]

Table 4 illustrates how payoffs depend on role and treatment. In *BoS*, only first-senders benefit from coordination, as subjects tend to coordinate on (H,L). In fact, first-senders in *BoS* obtain the highest payoffs (143, on average) of all subjects in all roles. Although this is not the preferred equilibrium for second-senders, the decrease in coordination failures ensures that they are not made worse off. First-senders in the Chicken games benefit from communication at the expense of second-senders, although the difference between roles is only significant in *C-Small*. Second-senders in *C-Small* are significantly worse off with communication than without (58 vs 73, $p = 0.003$), because the subjects coordinate somewhat more on the first-sender's preferred equilibrium. This is not so surprising given that in *C-Small* the second-sender earns only slightly less in the worst equilibrium (50) compared to the mixed equilibrium (56).

Table 4: Mean payoffs of first and second sender

Treatment	Prediction	No Communication	Communication			Test Difference
			All	Sender 1	Sender 2	(sender 1 vs 2)
BoS	40	54	98	143	53	$p < 0.001$
C-Small	57	73	76	94	58	$p = 0.002$
C-Large	100	111	112	117	107	$p = 0.142$

Notes: "Prediction" is the expected payoff if subjects play the mixed-strategy Nash equilibrium of game G. Communication protocol is two-sided sequential communication. Payoffs are net payoffs, including cost of messages. Reported p-values are based on two-sided Mann-Whitney tests.

Result 1 (Earnings and the option to communicate)**.** *With two-sided sequential communication, the option to communicate substantially increases mean earnings in the Battle-of-the-Sexes*

[17] Table 9 in Appendix C presents details about the frequencies of outcomes in the different treatments.

game but is futile in the Chicken games. Increasing the payoffs when both concede (c) sometimes makes subjects worse off by making communication futile.

4.2 What causes the (in)effectiveness of sequential communication?

What explains the differences in effectiveness of sequential communication between the games? In this section we investigate the extent to which subjects play in accordance with equilibrium, and if they do, with which one. First we focus on whether our assumptions about agreements hold. That is, we investigate whether conversations result in agreements and we consider what happens when no agreement is reached. Are conversations without agreement indeed ignored and do subjects then play according to the mixed strategy equilibrium?

4.2.1 The role of agreements

Based on the results from the coders, we group conversations into four main categories. In the first group are conversations in which the first sender is conceding (i.e., expresses an intention to play L) while the other player silently or explicitly agrees. In the second group are conversations in which the first sender is demanding (i.e., expresses an intention to play H) while the other player is silent or explicitly agrees. In the third group are conversations in which both players are demanding, such that there is no agreement. In the fourth group are conversations in which the first sender suggests to play (L, L) and the other player is silent or explicitly agrees.[18] In our experiment, most agreements are implicit, in the sense that the other player does not explicitly agree by sending an affirmative message nor explicitly disagrees. Behavior is quite similar when explicit agreements are reached (compared to when agreements are implicit). Sometimes our subjects actively avoided the costs of an explicit agreement by saying that if the other agreed, there was no need for another message.

Table 5 reports the frequency of the different types of conversations, and how often the players choose H. The first notable regularity in our data is that behavior is fairly similar when no messages are sent as when communication is impossible. The data are consistent with our assumption that subjects play the mixed-strategy equilibrium when they immediately end the conversation without talking.

[18] In a few cases, coders coded that a sender in Chicken indicated to play L. In these cases, we assume that the sender suggested to play (L, L).

A second finding is that subjects rarely explicitly disagree by both being demanding. This still happens somewhat regularly in *BoS* (in 20% of the cases), but it is more rare in *C-Small* (10%) and *C-Large* (4%). When both players are demanding, they tend to choose H. In *BoS* the rate with which they choose H coincides with the mixed-strategy equilibrium, as assumed in the theory section. In the Chicken games the rate is substantially higher than the mixed-strategy equilibrium, but disagreement occurs relatively rarely in those games.

Result 2 (Disagreements and behavior in the absence of agreements)**.** *With two-sided sequential communication, subjects do not often explicitly disagree by both being demanding. When they do, they tend to choose H. When they do not talk, they choose H in accordance with the mixed-strategy equilibrium.*

Table 5: Detailed contents of conversations and behavior

Treatment	Condition	% of conversations	% Choosing H:	
			Sender 1	Sender 2
BoS	No Communication		70	70
	Communication			
	No messages	8	78	69
	Sender 1 conceding	3	0	100
	Sender 1 demanding	69	100	7
	Both senders demanding *H*	20	73	77
	Suggestions to both play L	0	–	–
C-Small	No Communication		60	60
	Communication			
	No messages	44	60	50
	Sender 1 conceding	2	45	100
	Sender 1 demanding	28	98	28
	Both senders demanding	10	79	85
	Suggestions to both play L	14	62	53
	Not classified	2	–	–
C-Large	No Communication		44	44
	Communication			
	No messages	38	39	40
	Sender 1 conceding	2	38	50
	Sender 1 demanding	8	100	39
	Both senders demanding	4	81	81
	Suggestions to both play L	45	36	28
	Not classified	3	–	–

Notes: Without communication, there is no distinction between the two players. Communication protocol is two-sided sequential communication.

4.2.2 Conversations and Equilibrium

There seems to be an understanding that communication favors the first sender. First senders concede only very rarely in all three treatments. Instead, we find that demanding behavior by the first sender is very prevalent in *BoS* (69% of the time). It is much less often observed in the Chicken games, though (28% and 8% of the time). Remember that the theory predicts that only in *BoS* we should observe demanding behavior.

Even when the content of the messages is sometimes the same between games, behavior can be very different. When sender 1 is the only demanding player in *BoS*, there is almost perfect coordination: sender 1 always chooses H and sender 2 almost always chooses L (93% of the time). In the Chicken games, Sender 2 takes Sender 1's demand with a grain of salt and is substantially less likely to play L in such a case (72% of the time in *C-Small* and 61% in *C-Large*). The difference between *BoS* and *C-Small* is significant ($p = 0.002$, two-sided MWU test), but that between *BoS* and *C-Large* is not ($p = 0.161$).

Result 3 (Demanding and conceding behavior)**.** *In BoS, first senders tend to be demanding and second senders concede with two-sided sequential communication. In the Chicken games, there is substantially less to no demanding behavior and second senders are less likely to concede than in BoS.*

Next, we zoom in on the conversations in which subjects suggest to play (L,L). This does not happen in *BoS* and is still relatively rare in *C-Small* (14%), but quite common in *C-Large* (45%). Interestingly, when a conversation ends with the suggestion to play (L,L), a large fraction of subjects still choose H, roughly 57% in *C-Small* and 32% in *C-Large*.[19] Consistent with the theoretical prediction, there are less deviations from the agreement in *C-Large* ($p = 0.012$, two-sided MWU test). The rate to play H in *C-Large* after (L,L) is suggested, 32%, is well below the rate when there is no communication, 44%, but still far above zero.

It remains a question whether an agreement on (L,L) is effective, and makes subjects more likely to play L, or only captures a self-selection of subjects who would otherwise also have chosen L at the same rate. To examine this, we make a within-person comparison, comparing the likelihood of choosing H after an agreement on (L,L) with the likelihood of choosing H when they are allowed to communicate but no message is sent. We find a

[19] At the pair level, we find that in only 44% of cases both subjects behave in conformity with the agreement in *C-Large*. If subjects within a pair would independently deviate from the agreement, we would expect that 46% of pairs ends up choosing (L,L). The actual percentage is 44, suggesting that subjects do not have a way of telling if their opponent will stick to the agreement.

small and insignificant increase of 2 percentage points in the likelihood of choosing H after an agreement on (L,L) in *C-Small* ($p = 0.834$, two-sided Wilcoxon signed-rank test), and a significant decrease of 12 percentage points in *C-Large* ($p = 0.035$). This suggests that in *C-Large*, an agreement on (L,L) does not just reflect subjects' type, but changes behavior to some extent.

Result 4 (Compromises)**.** *With two-sided sequential communication, suggestions to play* (L,L) *are absent in BoS and most common in C-Large. Consistent with the equilbrium prediction, many subjects deviate from the agreement and more subjects deviate from the agreement when the joint concession payoff c increases.*

Table 6 presents information on the number of messages in the conversations. The mean number of messages is 0.9 in *C-Small* and 1.0 in *C-Large*, against 1.4 in *BoS*. The conversations are much shorter than in the haggling equilibrium. More importantly, subjects in the Chicken games are much less likely to send any messages at all: 44% of pairs in *C-Small* and 38% in *C-Large* do not communicate. In *BoS*, only 8% of pairs do not communicate at all. Subjects seem to understand that communication is rather ineffective in the Chicken games, and therefore avoid sending costly messages. This result is consistent with the theory. More specifically, it is predicted that players always communicate in *BoS*, but that they only talk in the Chicken games if they are averse to lying.

Remarkably, the ineffectiveness of communication in the Chicken games does not appear to be driven by the fact that fewer pairs sent messages. Even among the pairs that do send messages, mean earnings are not higher than without communication. This is illustrated in Figure 5, that shows mean earnings by the length of the conversation. In *BoS*, mean earnings are highest when only one message is sent. Sending more messages is associated with lower mean earnings, probably because longer conversations are a consequence of disagreement. In the Chicken games, mean earnings are close to the mean earnings without communication when two or fewer messages are sent. For three or more messages, mean earnings drop, but this happens relatively rarely.

Table 6: Distribution of messages

Treatment	Mean	# messages (%)			
		0	1	2	3+
BoS	1.4^b	8	58	23	12
C-Small	0.9^a	44	32	16	7
C-Large	1.0^a	38	27	30	4

Notes: Entries with different superscripts are significantly different at the 5% level (two-sided Mann-Whitney test).

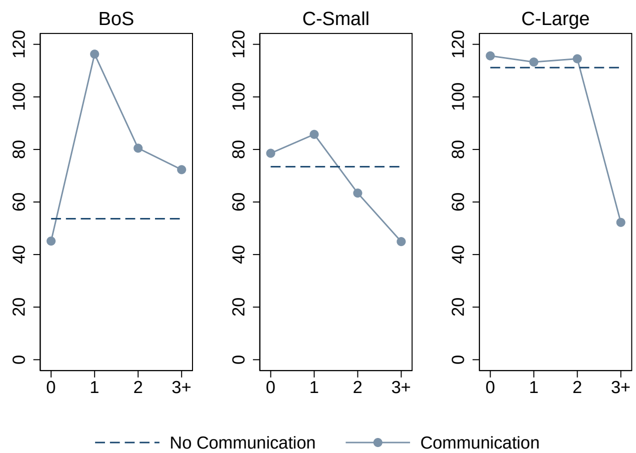

Figure 5: Mean earnings by number of messages sent. The horizontal (dashed) lines are reference lines showing the mean earnings without communication. Communication protocol is two-sided sequential communication.

Result 5 (Communication length)**.** *In BoS, subjects use the option to communicate with two-sided sequential communication. In the Chicken games, subjects appear to understand that communication is futile and often forgo the possibility to communicate. In all cases, conversations are short.*

Overall, the results line up well with the theoretical predictions listed in Section 2.1.3: (1) subjects are sometimes worse off as c increases as it makes communication futile,

(2) explicit disagreements are rare, (3) subjects tend to play the mixed-strategy after no agreement, (4) subjects quickly coordinate on the first sender's preferred equilibrium in *BoS*, (5) subjects often do not communicate in the Chicken games, or they agree on (L,L) but then often deviate especially in *C-Large*. The major deviations from the predictions are that: (1) first senders are often demanding in *C-Small*, and (2) subjects choose H at a higher rate than expected after explicit disagreements in the Chicken games. Such disagreements are, however, not very common.

4.3 Learning

Although the game itself is simple, subjects may have to learn how to interpret others' messages and how their own messages are interpreted by others. In this section, we briefly look at learning effects.

We do not find much evidence of learning. The length of communication and the effect of communication on earnings change little over time. Figure 6 shows the length of communication over time in the different treatments. There is no discernible time trend in any of the treatments. Almost right from the start, the mean number of messages is higher in *BoS* than in the Chicken games (left panel), and the percentage of cases where subjects do not send messages is lower in *BoS* than in the Chicken games (right panel). In terms of earnings, we find that in *BoS* there is a stable earnings gap between the communication and no-communication conditions, while communication is ineffective in the Chicken games in all rounds (see Figure 7). Although earnings vary somewhat over time, the effectiveness of communication does not.

5 Experiment II: Alternative Communication Protocols

Our communication protocol differs from the most commonly used protocols in the literature. In a follow-up experiment, we implemented two different communication protocols. In *one-sided*, only one of the players could send a message. In *chat*, both players could use a chat box to communicate in a less structured manner. These treatments allow us to make a good comparison with the existing literature and provide a stress test of our

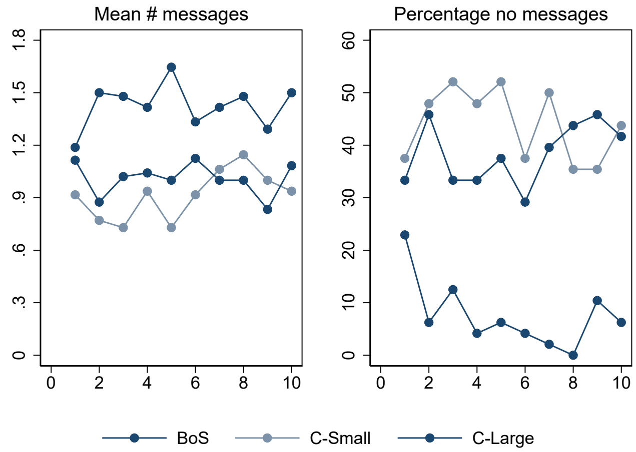

Figure 6: Mean number of messages over rounds. Communication protocol is two-sided sequential communication.

model. We tested both alternative communication protocol in *BoS* and *C-Large*, which we believe are the most interesting treatments.[20]

5.1 Experimental Design and Procedures

In "*one-sided*", there was a single round in which one of the players could send a message. To stay close to the existing literature on one-sided communication, the possible messages were costless and pre-coded. Messages had the form "I intend to choose ___ and propose that you will choose ___" where each of the blanks could be replaced by "H" or "L". Senders also had the option not to send a message. It was randomly determined which of the two players could send a message.

In "*chat*", both players could send free-form messages using a chat box. Messages were costless. There was no sequential talk order enforced. Players decided themselves who started the conversation, and after that they could communicate in any order. Each player could at any point indicate the wish to end the conversation. This would freeze the chat box. The other player could then agree to end the conversation or to decline the request and continue with the chat. In case the other player continued, both players could send messages again. After three minutes, a message appeared on the screen asking the players to conclude the chat, but players were not forced to stop.

[20]We are very grateful to anonymous referees for suggesting these treatments. We did not run *C-Small* for practical reasons (a limited number of participants in the database).

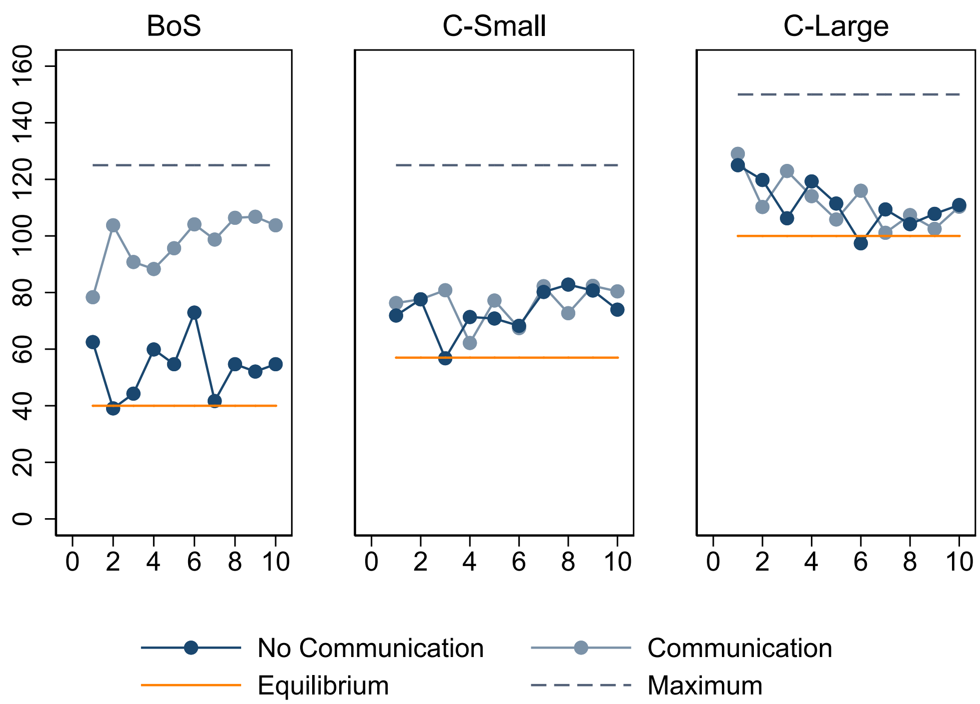

Figure 7: Mean earnings over rounds. "Equilibrium" is the expected payoff if subjects play the mixed-strategy Nash-equilibrium of game G. "Maximum" is the maximum mean efficiency. Communication protocol is two-sided sequential communication.

A total of 312 subjects (55 percent females) participated in one of 15 sessions. Subjects who had participated in Experiment I were excluded. We have 6 independent matching groups in each of the treatments with one-sided communication and 7 independent matching groups in each of the chat treatments. The other design features and procedures are similar to Experiment I. The same three research assistants coded the chat contents. This time we did not ask to make a distinction between weak and strong expressions. Coders also indicated which player was the first to indicate a strategy. The intra-rater consistency is good. The percentage of identically coded messages in the retest questions varies between 84 and 94. The inter-rater reliability is also good, with values of kappa ranging between 0.81 and 0.83.

5.2 Effects of one-sided communication

For *One-sided* communication we have clear theoretical predictions. In *BoS*, senders will propose (H, L) and players coordinate on that outcome. In *C-Large*, communication is ineffective and players follow the mixed-strategy equilibrium.

Table 7 summarizes the earnings. Note first that the mean payoffs without communication closely replicate those of Experiment I. We cannot reject that the distributions are the same ($p = 0.897$ for *BoS* and $p = 0.561$ for *C-Large*).

The main predictions are supported by the data. In *BoS*, senders propose to play outcome (H, L) 94 percent of the time. Following this message, senders always play H and in 95 percent of cases receivers play L. Overall, the coordination rate on outcomes (L, H) and (H, L) increases from 43 percent without communication to 93 percent with communication. The high coordination rate with one-sided messages results in substantially higher mean payoffs (116) compared to mean payoffs without communication (54). Not surprisingly, it is exclusively the sender that benefits from communication.

In *C-Large*, communication is again ineffective. Mean payoffs with communication (105) are even a bit lower than without communication (115), but the difference is not statistically significant ($p = 0.200$, two-sided MWU test). The coordination rate on outcomes (L, H) and (H, L) increases only modestly from 47 percent without communication to 54 percent with communication. Although senders frequently propose (L, L) in this case, the actual coordination on this outcome is not higher than without communication. Those that propose (H, L) tend to play H, but cannot convince receivers to play L. On average, senders are equally well off with communication, while receivers earn somewhat less.

Our model predicts that for *BoS*, mean earnings under one-sided communication are possibly higher (but not lower) than under sequential communication. The reason is that only one (pure-strategy) equilibrium exists under one-sided communication, while under sequential communication there is also a mixed-strategy equilibrium that yields lower payoffs. This is indeed the case: mean earnings with one-sided communication (116) are higher than with sequential communication (98) and the difference is significant ($p = .007$). The reverse is true for *C-Large*. There, a mixed-strategy equilibrium exists for both communication protocols, but only under sequential communication can (L, L) be an equilibrium agreement (provided subjects are sufficiently lying averse). Mean earnings are indeed higher with sequential communication, although the difference is modest and not significant (105 with one-sided communication versus 112 with sequential communication, $p = 0.519$).

Result 6 (One-sided communication)**.** *With one-sided communication, results agree with the feigned ignorance principle. Communication effectively solves the coordination problem in favor of the sender in BoS. Communication is ineffective in C-Large.*

Table 7: Mean earnings by treatment and communication

Treatment	Prediction	Communication					Test Difference (*p*-values)	
		No	Yes					
			Protocol	All	(First) Sender	Other		
		(1)		(2)	(3)	(4)	(1)=(2)	(3)=(4)
BoS	40	54	One-sided	116	182	50	0.004	0.004
C-Large	100	115	One-sided	105	116	94	0.200	0.037
BoS	40	54	Chat	93	107	80	0.002	0.018
C-Large	100	114	Chat	128	128	127	0.085	0.944
BoS (Exp I)	40	54	Sequential	98	143	53	< 0.001	< 0.001
C-Large (Exp I)	100	111	Sequential	112	117	107	0.815	0.142

Notes: "Prediction" is the expected payoff if subjects play the mixed-strategy Nash equilibrium of game G. Statistical tests report p-values of two-sided Mann-Whitney tests using matching group level data as the independent unit of observation ($N = 6$ for each treatment with *one-sided*, $N = 7$ for each treatment with *chat*). In the *Chat* treatments, the determination of sender 1 and 2 is based on the classification of coders. The bottom two rows reproduce the results of Experiment I with sequential communication.

5.3 Effects of free-form communication

Our model of sequential communication in section 2.1 captures situations in which people communicate in a structured way, such as in email exchanges. In this subsection, we report the results of our *chat* treatment, in which people can communicate in a less structured manner.

We think that the chat environment provides an interesting robustness check. Players are more symmetric when they can chat. Instead of randomly assigning a player to be the first sender, both players can start the conversation. This can potentially make it harder to coordinate in *BoS*. Players cannot rely on the random assignment of first senders to select which player should get his or her preferred outcome. If coordination with chat in *BoS* is lower than with sequential messages, this could indicate that the random assignment of first senders is important. The chat treatment is also interesting because it not only allows players to express disagreement but also to interrupt. Furthermore, players cannot unilaterally terminate the conversation in our chat treatment; both players need to agree to end the communication stage. This could avoid ambiguities in interpreting the other player's planned strategy.

The results are reported in Table 7. The mean payoffs without communication again closely resemble those of Experiment I. We again find that communication is very effective in *BoS*. Mean payoffs increase from 54 to 93. The increase is of a similar magnitude as with sequential communication. Thus, the success of sequential communication in *BoS* does not appear to be driven by the random assignment of the first sender. We do see, however, that in *Bos* the first sender (as coded by our raters) is much less powerful than in the treatments with sequential or one-sided communication. The difference between first and second senders is on average 27, much less than in the other communication conditions (90 with sequential, 132 with one-sided).

Unlike sequential and one-sided communication, chat increases mean payoffs in *C-Large*: from 114 without communication to 128 with chat ($p = 0.085$, two-sided MWU test). The increase is only marginally significant and covers around 39 percent of the gap between the maximum average earnings and the average earnings without communication. While we can only speculate about the reasons for why chat is more effective than sequential communication in *C-Large*, we note that there are differences in frequency, length, and contents of communication. Whereas with sequential communication subjects did not send any messages in 38 percent of the cases, this happens rarely (less than 2 percent) with the chat. With chat, a large majority of 88 percent of the subjects suggests to play (L, L), against 45 percent with sequential communication. Conditional on an agreement to play (L, L), the (L, L) outcome materializes in 54 percent in chat and 44 percent in sequential. The average length of conversations also differs: with chat, a conversation averages 23 words, against 12 words with sequential communication. Possibly this is the result of making communication free in the chat treatments and of making it impossible to end the conversation unilaterally with chat. Interestingly, the richer communication in chat boosts the rate of (L, L) messages, while the propensity to stick to the agreement is not much affected.

We also note another similarity between sequential communication and free form chat: senders that start the conversation have an advantage in *BoS* but not in *C-large*. We believe that in *BoS* it is reasonable for a subject to demand his or her preferred equilibrium payoff.

This is less reasonable in *C-Large*, where (L,L) seems a fairer outcome. Consequently, many subjects in *C-Large* propose to coordinate on (L,L). Starting from this symmetric outcome, both subjects deviate at equal rates and first senders don't have an advantage.

Result 7 (Free chat communication)**.** *With free chat, communication solves the coordination problem to a large extent in BoS. Communication is partially effective in C-Large, where it increases the rate of* (L,L) *messages compared to sequential, but not so much the propensity to adhere to the proposal. With free chat there is a smaller advantage for the (endogenous) first mover than in the other protocols in BoS, and it is non-existent in C-large.*

6 Discussion

In this Section we discuss how effective the approach based on the feigned-ignorance principle is in comparison with two other approaches when it comes to explaining the effects of pre-play communication in coordination games.

Our finding that communication is very effective in *BoS* and much less so in *Chicken* is also in agreement with the alternative view that cheap talk messages are only credible when the sender's message is both self-committing and self-signaling (Aumann 1990; Farrell and Rabin 1996). Self-commitment requires that if the message is believed, and the receiver optimizes on the basis of this belief, the sender wants to fulfill it. Self-signaling demands that if the receiver believes the message, the sender only wants to send this message if she plans to play in agreement with it.[21] Previous experimental evidence does, however, not support the idea that a message needs to be self-signaling to be credible. Communication tends to remain very effective in situations where self-signaling does not apply. Table 8 shows one of the Stag-Hunt games in which Charness (2000) finds a huge effect of one-sided communication on subjects' willingness to choose the risky cooperative action even though the message to cooperate is not credible. Notice that each player prefers the other player to play B for each of her own choices (80>70 and 90>50). Thus a message stating "I intend to play B" is not self-signaling. Nevertheless, subjects find the message very credible, and the rate of B-play increases from 0.40 to 0.90 when one-sided

[21] More formally: Consider the case with one-sided communication and let $BR_j(A_i)$ be the receiver's unique best response to action A_i by the sender. Suppose the sender sends a message m_i^* claiming to play A_i^*. If the sender thinks that her message is believed, the message is self-committing if $u_i(A_i^*, BR_j(A_i^*)) \geq u_i(A_i, BR_j(A_i^*))$ for any $A_i \in A$. The message claiming to play A_i^* is self-signaling if for every $A_i \neq A_i^*$, there is a message followed by some action A_j by the receiver such that $u_i(A_i, BR_j(A_i^*)) < u_i(A_i, A_j)$. See also Baliga and Morris (2002).

communication is allowed. Interestingly, Farrell and Rabin (1996) already foreshadowed this result when they conjectured “although we see the force of Aumann’s argument, we suspect that cheap talk will do a good deal to bring Artemis and Calliope to the stag hunt” (p.114).

Table 8: Stag-Hunt game

		Player 2	
		A	B
Player 1	A	70, 70	80, 50
	B	50, 80	90, 90

Notes: this is game 1 of Charness (2000).

The feigned-ignorance principle predicts an effect of communication in Stag-Hunt games even when the message to cooperate is not self-signaling. Unless players coordinate on the cooperative equilibrium without communication, in which case there is no scope for improvement, an agreement on mutual cooperation is supported as a Nash equilibrium that makes both players better off. In the game of Table 8, subjects choose B in only 40 percent of the cases without communication. Risk aversion may encourage subjects to choose the risky option B less often than in the risk neutral symmetric mixed strategy equilibrium, which predicts a rate of 2/3 B-play. In any case, there is ample room for mutual improvement compared to the focal outcome without communication, and feigned ignorance predicts that one-sided communication helps players to coordinate on the cooperative equilibrium in which both players are substantially better off.

Similar patterns are observed in related experiments. Clark et al. (2001) find a lesser but still substantial effect of two-sided communication in the Stag-Hunt game.[22] Blume and Ortmann (2007), Brandts and Cooper (2007) and Avoyan and Ramos (2016) investigate the effects of communication on the extent to which subjects cooperate in the minimum effort game. Also in this game, a message to choose the cooperative action is not self-signaling, since players (weakly) prefer other players to choose higher actions. Cheap talk is also very effective in helping subjects to cooperate in these studies with larger groups. That

[22]Burton and Sefton (2004) extend the results to 3x3 games. Existing experimental studies on social dilemmas establish the positive effect of costless pre-play communication on cooperation (see e.g. Bicchieri and Lev-On 2007). In the Prisoner’s Dilemma, some players may feel guilty to play defect when others play cooperatively. For them, the Prisoner’s Dilemma is in essence a Stag-Hunt game, and pre-play communication may help because both players can gain compared to the mixed equilibrium without communication.

is why we prefer our less demanding assumption that communication is effective when it helps all players to reach a better equilibrium outcome than without communication. The feigned-ignorance principle is in line with the positive effect of communication in all these studies.

Ellingsen and Östling (2010) propose a very different approach based on the idea that players differ in their level of rationality. Among other contributions, they develop the predictions of the level-k model for one-sided communication in the Battle of the Sexes and the Chicken games that we study. In their model, Level-0 players, referred to as T_0, choose randomly, with equal probability for each choice, and they ignore any message that they receive. Level-0 players may not actually exist; it is sufficient that they exist in the minds of level-1 players. Higher level players send messages that serve them best. Players are truthful when they are indifferent between messages. For positive integers k, a T_k player chooses a best response to the behavior that the T_k player expects from a T_{k-1} opponent. Players will sometimes observe unexpected messages. If that happens, they assume that it comes from the highest of the lower types T_{k-j} that makes T_k's inference consistent with the observed message.

In the Battle of the Sexes and Chicken games, players who perform more than one thinking step behave in the same way. Thus, it is convenient to let T_{k+} denote player types that think at least k steps ahead. If H is the risk dominant action, as in our case it is when $c = 0$ and $c = 75$, then T_{1+} senders send the message that they intend to play H and they play H whereas T_{1+} receivers optimally respond to this message. When $c = 150$, no action risk dominates the other, which means that T_1 senders can either say that they intend to choose L and choose L, or say that they intend to choose H and choose H. T_{2+} senders continue to communicate that they will play H and act in agreement with the message. In either case, T_{1+} receivers optimally respond to the message that they receive. One-way communication therefore implies that T_{1+} players always coordinate on the equilibrium preferred by the sender, unless $c = 150$ and the sender is of type T_1, in which case coordination is on either of the two pure equilibria. Only level-0 players are predicted to make uncoordinated choices, but empirically such players are found to be fairly rare (see e.g., Georganas et al. 2015). Ellingsen and Ostling conclude that one-way communication resolves the coordination conflict in these games.

The predictions of the level-k model are supported by the data on the effects of one-sided communication in the Battle-of-the-Sexes, where subjects almost always coordinate on the equilibrium preferred by senders. In the Chicken games with $c = 150$, results are at odds with the level-k predictions. There, one-sided communication does not help subjects to coordinate more often than they do without communication, and subjects earn similar amounts as in the inefficient mixed strategy equilibrium. Arguably, the level-k model is intended to capture the behavior of inexperienced subjects. Even in the first round, however, where subjects are still inexperienced, the coordination rate on outcomes (L, H) and (H, L) is only 39 percent in *C-Large*, far from perfect coordination.

7 Conclusion

Under various communication protocols, we investigated how people play coordination games with conflicting interests when they have the possibility to send non-binding messages. We developed a theory based on the feigned-ignorance principle, which says that players ignore communication if it harms any of them. Theoretically, we found that the effectiveness of sequential communication and one-sided communication depends crucially on the comparison of c (the joint concession payoff) and b (the payoff of the disadvantaged player in a pure equilibrium). If $c \leq b$, as in the Battle-of-the-Sexes, then either communication protocol is predicted to solve the coordination problem. With sequential communication, it may happen that agreement is immediate and that either first or second sender is advantaged. It may also happen that people haggle for a long time and dissipate a substantial part of the available pie. With one-sided communication, agreement will be on the equilibrium that favors the sender.

If on the other hand $c > b$, as in the Chicken games that we studied, the prediction with standard preferences is that communication is ineffective, both with sequential and one-sided communication. Notice that this prediction is quite surprising. It deviates for instance from Ellingsen and Östling (2010) who predict that communication will powerfully resolve the coordination problem if players have some depth of thinking.

Theoretically, with the sequential protocol communication may also be effective in the Chicken games when players are lying-averse. If players suffer a cost when they deviate from an agreement, they may agree to both concede. Players with a high cost of lying will then conform to the agreement, while other players deviate. The higher c, the more conforming behavior is to be expected.

In the experiment, we find that communication is very effective in the Battle-of-the-Sexes. With sequential as well as one-sided communication, there appears to be a common understanding that play should favor the first sender. Subjects do not lose much time to coordinate on this equilibrium.

In agreement with the feigned-ignorance principle, the possibility of communication is ineffective under the sequential and one-sided protocols in the Chicken games. In the aggregate, it does not allow subjects to benefit and subjects often simply forgo the possibility to talk. With the sequential protocol, when they do talk, they often agree on the outcome that gives them both c. As predicted, such agreements are only partly followed, and the extent to which they are followed responds positively to c. Even though our theory predicts that communication is completely ineffective in the Chicken games, we still sometimes observe that subjects focus on the outcome that benefits the first sender. Demanding the good outcome is not without risk though, since it may easily happen that both subjects are demanding, after which the bad outcome frequently occurs.

As a robustness check, we also investigate how effective unstructured free-chat communication is. With this protocol, subjects decide endogenously who starts the conversation, and the end of the conversation has to be mutually agreed on. This communication protocol again solves the coordination problem in the Battle-of-the-Sexes. Free-chat communication is partially effective in the Chicken game. Compared to the other communication protocols, there is a smaller advantage for the (endogenous) first mover.

An interesting finding is that many subjects deviate from an agreement to both play L. In other experiments using different games, subjects are often very cooperative and trustworthy after communicating with one another (e.g., Bicchieri and Lev-On 2007; Balliet 2009). Why doesn't that happen in our environment? One possibility is that chatting is less forceful than face-to-face communication (e.g. Jensen et al. 2000; Brosig et al. 2003). Another reason might be that the wording in the conversations is different. Many of our subjects do not make explicit promises, but instead only make a suggestion or a statement about what would be fair to do. In other experiments, subjects often make promises and those are a reliable sign of a person's trustworthiness (e.g. see Charness and Dufwenberg 2006; Belot et al. 2010; He et al. 2016). Of course, that begs the question why subjects are more reluctant to make promises in our game than in other games. We do not have a clear answer to this, but possibly subjects feel no need to make promises because it is obvious what the desired course of action is; (L, L) gives high payoffs to both in *C-Large*, and the gains from deviating is relatively small. In other games, subjects have more to gain from deviating from an agreement, and might be more compelled to make a convincing case that they can be trusted.

An open question is how the alternative communication protocols (one-sided communication and free chat) would affect behavior in *C-Small*. Free chat, in particular, may have a larger effect on earnings in *C-Small* than in *C-Large*. The reason is that relative to the maximum earnings, mean earnings without communication are already relative high in *C-Large* (75 percent) compared to *C-Small* (61 percent). Thus, the modest effect of free chat in *C-Large* may be driven by a ceiling effect. On the other hand, the effect of communication may be higher in *C-Large*, because an agreement on (L, L) becomes more attractive and the gains from deviating from that agreement are smaller. Which force is stronger is an empirical question, that is interesting to examine in future research.

References

Aumann, R. (1990). Nash Equilibria Are Not Self-Enforcing. In *Economic Decision Making: Games, Econometrics, and Optimization: Contributions in Honor of Jacques H. Dreze*, pages 201–206. Amsterdam: North-Holland, j.j.gabsze edition.

Avoyan, A. and Ramos, J. (2016). What It Takes to Coordinate: Road to Efficiency Through Communication and Commitment.

Baliga, S. and Morris, S. (2002). Co-ordination, Spillovers, and Cheap Talk. 468:450–468.

Balliet, D. (2009). Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review. *Journal of Conflict Resolution*, 54(1):39–57.

Belot, M., Bhaskar, V., and van de Ven, J. (2010). Promises and cooperation: Evidence from a TV game show. *Journal of Economic Behavior & Organization*, 73(3):396–405.

Bicchieri, C. and Lev-On, A. (2007). Computer-mediated communication and cooperation in social dilemmas: an experimental analysis. *Politics, Philosophy & Economics*, 6(2):139–168.

Blume, A. and Ortmann, A. (2007). The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria. *Journal of Economic Theory*, 132(1):274–290.

Brandts, J. and Cooper, D. J. (2007). It's What You Say, Not What You Pay: An Experimental Study of Manager Employee Relationships in Overcoming Coordination Failure. *Journal of the European Economic Association*, 5(6):1223–1268.

Brosig, J., Weimann, J., and Ockenfels, A. (2003). The Effect of Communication Media on Cooperation. *German Economic Review*, 4(2):217–241.

Burton, A. and Sefton, M. (2004). Risk, pre-play communication and equilibrium. *Games and Economic Behavior*, 46(1):23–40.

Cason, T. N. and Mui, V. L. (2015). Rich communication, social motivations, and coordinated resistance against divide-and-conquer: A laboratory investigation. *European Journal of Political Economy*, 37:146–159.

Charness, G. (2000). Self-Serving Cheap Talk: A Test Of Aumann's Conjecture. *Games and Economic Behavior*, 33(2):177–194.

Charness, G. and Dufwenberg, M. (2006). Promises and Partnership. *Econometrica*, 74(6):1579–1601.

Charness, G. and Dufwenberg, M. (2010). Bare promises: An experiment. *Economics Letters*, 107(2):281–283.

Clark, K., Kay, S., and Sefton, M. (2001). When Are Nash Equilibria Self-Enforcing? An Experimental Analysis. *International Journal of Game Theory*, 29:495–515.

Cooper, R., Dejong, D. V., Forsythe, R., and Ross, T. W. (1989). Communication in the battle of the sexes game: some experimental results. *RAND Journal of Economics*, 20(4):568–587.

Costa-Gomes, M. A. (2002). A Suggested Interpretation of Some Experimental Results on Preplay Communication. *Journal of Economic Theory*, 104(1):104–136.

Di Bartolomeo, G., Dufwenberg, M., Papa, S., and Passarelli, F. (2018). Promises, expectations & causation. *Games and Economic Behavior*, 1:1–10.

Duffy, J. and Feltovich, N. (2002). Do Actions Speak Louder Than Words? An Experimental Comparison of Observation and Cheap Talk. *Games and Economic Behavior*, 39(1):1–27.

Duffy, J. and Feltovich, N. (2006). Words, deeds, and lies: Strategic behaviour in games with multiple signals. *Review of Economic Studies*, 73(3):669–688.

Ellingsen, T. and Johannesson, M. (2004). Promises , Threats and Fairness. *Economic Journal*, 114(495):397–420.

Ellingsen, T., Johannesson, M., Tjøtta, S., and Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, 68(1):95–107.

Ellingsen, T. and Östling, R. (2010). When Does Communication Improve Coordination? *American Economic Review*, 100(4):1695–1724.

Ellingsen, T., Östling, R., and Wengström, E. (2018). How does communication affect beliefs in one-shot games with complete information? *Games and Economic Behavior*, 107:153–181.

Embrey, M., Fréchette, G. R., and Lehrer, S. F. (2014). Bargaining and reputation: An experiment on bargaining in the presence of behavioural types. *Review of Economic Studies*, 82(2):608–631.

Farrell, J. (1987). Cheap talk, coordination, and entry. *The RAND Journal of Economics*, 18(1):34–39.

Farrell, J. (1988). Communication, coordination and Nash equilibrium. *Economics Letters*, 27(3):209–214.

Farrell, J. (1995). Talk is Cheap. *The American Economic Review*, 85(2):186–190.

Farrell, J. and Rabin, M. (1996). Cheap Talk. *Journal of Economic Perspectives*, 10(3):103–118.

Georganas, S., Healy, P. J., and Weber, R. A. (2015). On the persistence of strategic sophistication. *Journal of Economic Theory*, 159:369–400.

He, S., Offerman, T., and van de Ven, J. (2016). The Sources of the Communication Gap. *Management Science*, (Articles in Advance):1–15.

Jensen, C., Farnham, S. D., Drucker, S. M., and Kollock, P. (2000). The effect of communication modality on cooperation in online environments. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*, 2(1):470–477.

Palfrey, T., Rosenthal, H., and Roy, N. (2015). How cheap talk enhances efficiency in threshold public goods games. *Games and Economic Behavior*.

Rabin, M. (1994). A Model of Pre-game Communication. *Journal of Economic Theory*, 63(2):370–391.

Rubinstein, A. (1982). Perfect Equilibrium in a Bargaining Model. *Econometrica*, 50(1):97–109.

Santos, V. (2000). Alternating-announcements cheap talk. *Journal of Economic Behavior & Organization*, 42(3):405–416.

Tirole, J. (1988). The Theory of Industrial Organization. *The MIT Press*.

Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations1. *Econometrica*, 76(6):1467–1480.

Appendices

Appendix A: Proofs

Proof of Proposition 1. First note that if the first sender does not immediately terminate the communication stage, then strategies can only be part of a subgame perfect equilibrium profile in pure strategies if the resulting outcome of the communication stage is a credible agreement (i.e., an agreement on strategy pairs (H,L) or (L,H) of game G). If the communication stage would not end in a credible agreement, then the first sender does strictly better by immediately terminating the communication stage. It follows that players will only send non-empty messages if $b \geq c$. If players reach an agreement, and $b < c$, then by the feigned ignorance principle the agreement will be ignored and expected payoffs are μ for each player. Sending $m_1 = \varnothing$ achieves the same expected payoffs at lower communication costs.

Note also that if the communication stage ends in an agreement on (L,H), then the agreement must be immediate. If it occurs after more than one message, sender 1 can deviate to sending lh in the first period. Sender 2's best-response is then to terminate the communication stage. This gives sender 1 the same payoff b in game G with fewer costly messages.

If the communication ends in an agreement on (H,L), then the conversation length can be at most 3 messages. To see this, note first that if sender 2 sends a conceding message (i.e., hl), copying the message is a strictly dominant strategy for sender 1. Copying the message ends the conversation with an agreement on (H,L), resulting in the highest possible payoff a in game G. Not agreeing can at best result in the same payoff a but with more costly messages. Thus, as part of any subgame perfect equilibrium profile, players must have beliefs specifying that a conceding message hl of sender 2 will be copied by sender 1. If players' strategies are such that they agree on a (H,L) after 4 or more messages, then sender 2 would strictly gain by sending the conceding message earlier on in the conversation: this would be copied and result in the same payoff in game G and a lower cost of messages.

This leaves the following four candidate SPE outcomes:

(i) Immediate termination ($m_1 = \varnothing$). This can be sustained as a SPE outcome for $\mu \geq b-\gamma \Leftrightarrow \gamma \geq \gamma_1 \equiv b(b-c)/(a+b-c)$. Immediate termination yields a payoff μ to each player. If $\mu < b-\gamma$, then the first sender can deviate to sending lh which then results in an agreement to play (L,H) and a payoff of $b-\gamma$. If $\mu \geq b-\gamma$, immediate termination can be sustained as a SPE outcome by the following strategies in the communication stage: the second sender always sends the message lh for every possible history, the first sender starts with the empty message and sends message lh at every other node in which it is sender 1's turn.

(ii) Immediate concession ($m_1 = lh$, $m_2 = \varnothing$). This can be sustained for $\gamma \leq \gamma_1$ by the following strategies in the communication stage: the first sender always sends lh for every possible history, the second sender responds with $\varnothing$ to $m_1 = lh$ and after that sends lh for every possible history, and after any other message m_1 sender 2 sends lh forever for every possible history. If $\gamma > \gamma_1$, the first sender gains by deviating to $m_1 = \varnothing$.

(iii) Immediate demanding ($m_1 = hl$, $m_2 = \varnothing$). This can be sustained for $\gamma \leq \gamma_2 \equiv a(a-c)/(a+b-c)$ by the following strategies in the communication stage: the first sender always sends hl for every possible history, and the second sender responds with $\varnothing$ to $m_1 = hl$ and after that sends hl for every possible history, and after any other message m_1 sender 2 sends hl forever for every possible history. If $\gamma > \gamma_2$, the first sender gains by deviating to $m_1 = \varnothing$.

(iv) Delayed demanding ($m_1 \in \{hh, ll\}, m_2 = hl, m_3 = hl$). In this SPE, the first message can be hh or ll, and this is followed by the second sender conceding. This can be sustained for $\gamma \leq \min\{\frac{1}{2}\gamma_1, \frac{1}{3}\gamma_2, \frac{1}{2}(a-b)\}$ by the following strategies in the communication stage: the first sender starts with hh or ll, and sends hl in any subgame that starts after this message. In any subgame after a first message of hl or lh by sender 1, sender 1 always sends lh. Sender 2 sends hl in any subgame that starts with a first message by sender 1 of ll or hh. In any of the other subgames that start with a first message of lh or hl by sender 1, sender 2 is always demanding and sends lh, except in the subgame that immediately starts after $m_1 = lh$, in which sender 2 chooses $m_2 = \varnothing$. If $\gamma > \frac{1}{3}\gamma_2$, the first sender gains by deviating to $m_1 = \varnothing$. If $\gamma > \frac{1}{2}(a-b)$ then sender 1 gains by deviating to $m_1 = lh$. If $\gamma > \frac{1}{2}\gamma_1$, the second sender gains by deviating to $m_2 = \varnothing$. It is also easy to verify that the second sender does not want to deviate from his or her planned strategy after sender 1 deviates from $m_1 = hh$ or $m_1 = ll$.

Proof of Proposition 2. In this equilibrium, as long as each of the players sent the demanding message at every past node, each player randomizes between being demanding (sending a demanding message, i.e., hl for sender 1 and lh for sender 2) and conceding (sending a conceding message, i.e., lh for sender 1 and hl for sender 2). In the first period, a conceding message is followed by $m_2 = \varnothing$. If sender 1 is demanding in the first period, and sender 2 concedes in the second period, then sender 2 does so by sending $m_2 = \varnothing$. In any other period, a conceding message is always copied by the other player which terminates the communication stage.

The condition that a player must be indifferent between conceding and demanding dictates the mixing probabilities of the other player. We first derive these probabilities for the case in which the communication stage went on for more than three periods. (The first few periods are slightly different because after the first message an agreement can be reached by sending the costless empty message, whereas after that period an agreement can only be reached by copying the other player's message, which is costly).

After two or more messages, a player must copy the previous player's message to reach agreement. Sending a conceding message yields a payoff of $b - t\gamma$ to that player. After a demanding message, the other player sends a conceding message with probability $1-q$ (yielding a payoff of $a-(t+1)\gamma$) and sends a demanding message with probability q. Denote the continuation payoff of reaching period $t+2$ by V. A player must be indifferent between $b - t\gamma$ and $(1-q)(a-(t+1)\gamma + qV$. At period $t+2$, if reached, the other player should again be indifferent between being demanding and conceding. Conceding yields $b-(t+2)\gamma$, so this must be equal to her continuation payoff V. Hence, we must have:

$$b - t\gamma = (1-q)(a-(t+1)\gamma) + q(b-(t+2)\gamma).$$

It follows that $q = \frac{a-b-\gamma}{a-b+\gamma}$. The same logic applies to any other period $t \geq 4$ in the communication stage, yielding the same mixing probabilities q and $1-q$ in each period.

The payoffs of reaching an agreement after the first message are slightly different because an agreement can be reached with a costless empty message after m_1. In the first period, any positive probability of being demanding can sustain the mixed-strategy equilibrium. In the second period, player 2's probability q_1 of being demanding must make player 1 indifferent between being demanding (by sending hl) and conceding (by sending lh) in the first period. This requires that $(1-q_1)(a-\gamma) + q_1(b-3\gamma) = b-\gamma$, which gives $q_1 = \frac{a-b}{a-b+2\gamma}$. In the third period, player 1's probability q_2 of being demanding must make player 2 indifferent between being demanding (by sending lh) and conceding (by sending $\varnothing$) in the second period. This requires that $(1-q_2)(a-3\gamma) + q_2(b-4\gamma) = b-\gamma$, which gives $q_2 = \frac{a-b-2\gamma}{a-b+\gamma}$. For the sake of simplicity, we assume that player 1 also mixes with probabilities q_1 and $1-q_1$ in the first message.

The equilibrium can be sustained by the following strategies in the communication stage: after any other message than hl or lh, the next player always sends a demanding message and the other player always sends a conceding message until the communication ends once they reach an agreement. No player should wish to deviate to sending $m = \varnothing$, which is the case if $\gamma \leq \gamma_1$.

The probability that the conversation ends after a single non-empty message is $(1 - q_1) + q_1(1 - q_1)$; the probability that it ends after 2 non-empty messages is 0; the probability that it ends after 3 non-empty messages is $q_1^2(1 - q_2)$; and the probability that it ends after exactly $n \geq 4$ non-empty messages is $q_1^2 q_2 q^{n-4}(1 - q)$. The expected number of messages is then $1 + 2q_1^2 + \frac{q_1^2 q_2}{1-q}$.

Proof of Proposition 3. In the main text, we show why in equilibrium first senders may immediately propose ll and second senders may immediately accept it by sending $\varnothing$, because all types are better off compared to ignoring the agreement. Here we show how the proposal can be supported in a perfect Bayesian equilibrium. An agreement on (L, L) can be sustained by the following strategies in the communication stage. Any type of sender 1 starts with a first message of ll and chooses $\varnothing$ in any other subgame in which he or she can make a choice, and believes that any type of sender 2 will send the empty message for every possible history. Any type of sender 2 chooses $m_2 = \varnothing$ not only in response to $m_1 = ll$, but also in any other subgame in which he or she can make a choice. The following beliefs of sender 2 accommodate sender 2's strategy. After receiving a message other than $m_1 = ll$, sender 2 believes that it comes from a player of type $k_1 = 0$, and therefore sender 2's optimal strategy is to terminate the conversation by sending $m_2 = \varnothing$, as sender 2 has no incentive to reach any agreement with a type 0 opponent. After receiving $m_1 = ll$, sender 2 believes that the other player has a probability $F(k^*)$ to deviate from an agreement on (L, L), and therefore it is optimal for any type of player 2 to agree to $m_1 = ll$ by sending $m_2 = \varnothing$. The strategies in BG following an agreement and the conditions under which no type wants to deviate are specified in the main text. Without an agreement players choose according to the mixed strategy equilibrium in BG.

Proof of Proposition 4. Under Assumption 1 and when $b \geq c$, the receiver ignores messages hh and ll and follows messages hl and lh (recall that messages are assumed to be costless, i.e., $\gamma = 0$). By backward induction, the sender's best response is to send message hl as it yields a higher payoff than lh to the sender. In equilibrium, the sender proposes to play (H, L) (i.e., sends message hl) and plays H, and the sender plays L. When $b < c$, any of the four messages, if followed by the players, leads to payoffs which at least makes one of the players worse off compared to the mixed-strategy equilibrium of game G. Therefore, in equilibrium, no sender's messages would be followed by the receiver nor the sender. Any message will be ignored, and both players follow the mixed-strategy equilibrium of game G irrespective of the message.

Appendix B: Experimental and Coding instructions

Experimental Instructions (for subjects)
What follows are the instructions for $c = 0$ in which there is communication in rounds 6-10 and 16-20.

[common to all treatments:]

Welcome to this experiment on decision-making. Please read the following instructions carefully.

During the experiment, do not communicate with other participants unless we explicitly ask you to do so. If you have any question at any time, please raise your hand, and an experimenter will come and assist you privately.

Your earnings depend on your own choices and the choices of other participants. During the experiment, your earnings are denoted in points. At the start of the experiment you will receive a starting capital of 300 points. In addition you can earn points during the experiment. At the end of the experiment, your earnings will be converted to euros at the rate: 1 point = € 0.025. Hence, 40 points are equal to 1 euro. Your earnings will be paid to you privately.

You will be randomly matched with another person in the room. Each person will make a choice between H and L. If you and the other person both choose H, you will both receive nothing. If you choose H and the other person chooses L, then you receive 200 points and the other person receives 50 points. If you choose L and the other person chooses H, then you receive 50 points and the other person receives 200 points. If you and the other person both choose L, you will both receive 0 points.

The possible decisions and payoffs are also shown in the following matrix. In each cell of the matrix, the first number shows the amount of points for you, and the second number shows the amount of points for the other participant. In total, there will be 20 rounds. In each round, you are randomly rematched to another participant. At the end of each round, you will receive feedback about the decision of the other person and your payoffs.

Payoff matrix in Table 1 is displayed here.

In some rounds, you and the other person will have the opportunity to communicate before deciding between H and L. This happens in rounds 6-10 and 16-20.

[sequential communication:]
The communication works as follows. You and the other person can send messages to each other. There are four important rules for the communication:

- Only one person can send a message at a time. It will be randomly determined who can send the first message (you and the other person have an equal chance on being able to send the first message, independent of what happened in previous rounds). After that, you will take turns.

- Each of you have to pay 2 points for every message that is sent, no matter who sent the message. These points will be subtracted from your earnings. It is possible that your earnings in a round are negative. Any losses will be deducted from your starting capital.

- If it is your turn to send a message, you can also decide not to send any messages (by clicking on the "Leave chat" button). This will end the communication without affecting any of your earnings, and you and the other person will not be able to send any more messages in that round.

- You are not allowed to identify yourself in any way. If you identify yourself (for instance, by giving your name or describing what you look like or what you are wearing) you will be excluded from the experiment and lose all earnings including the starting capital.

[one-sided communication:]
The communication works as follows. Only one person can send a message (the sender). It will be randomly determined who will be the sender (you and the other person have an equal chance on being the sender, independent of what happened in previous rounds).

The sender has the choice between the following four messages:

- ”I intend to choose H and propose that you will choose H.”

- ”I intend to choose H and propose that you will choose L.”
- ”I intend to choose L and propose that you will choose H.”
- ”I intend to choose L and propose that you will choose L.”

The sender can also choose not to send a message. In that case, the other will receive a notification that the sender chose not to send a message.

[free-format chat:]
The communication works as follows. You and the other person can chat freely with each other. There are two important rules for the communication:

1. Communication ends once both of you decided to leave the chat. At the bottom of the communication page, there is a “Leave chat” button. If a person clicks this button, the chat pauses. The other person can then end the chat (by clicking ”OK”) or resume the chat (by clicking ”Cancel”). If the chat is resumed, you can both keep sending messages.
2. You are not allowed to identify yourself in any way. If you identify yourself (for instance, by giving your name or describing what you look like or what you are wearing) you will be excluded from the experiment and lose all earnings including the starting capital

[common to all treatments:]
In the rounds with communication, you will be paired with a different person in each round, so you will never chat with the same person twice. Likewise, in the rounds without communication, you will also be paired with a different person in each round, so you will never meet the same person twice in these rounds.

At the end of the experiment, 4 out of the 20 rounds will be randomly selected for payment. Your earnings equal the sum of the starting capital 300 points and your earnings in the 4 selected rounds. If your total earnings are negative, you will receive 0.

The instructions were followed by a few questions to test participants’ understanding.

Coding Instructions (for coders)

Thank you so much for helping us, your work is very valuable to us. Below are the instructions. Please read them carefully. If after reading you have any questions, please don't hesitate to ask any questions. Please work individually and do not discuss your choices with other people while you are working on this task. We will show you chat conversations between people that participated in an experiment. In the experiment, participants were paired and randomly assigned the role of "Sender 1" or "Sender 2." Each person had to make a choice between two options: "H" and "L." They made their choices at the same time, without knowing what the other person did. Before they made their decisions, they could send messages to each other. Sender 1 could start by sending a message, and after that they alternated. Sometimes only Sender 1 sent a message. Your task will be to classify messages. Always read the entire conversation before answering any questions. The first question is about the intended choice that Sender 1 expresses in his or her messages.

Question 1: Which intention does Sender 1 express?

Choose from: *Weak intention to choose H; strong intention to choose H; Weak intention to choose L; Strong intention to choose L; None of the above/I don't know.*

If Sender 1 writes "I will play H" or "I choose H, up to you" or "I will play H, you should play L" , then he or she expresses intentions to play H. If instead Sender 1 writes: "Let's both play L" or "Let's choose L", then he or she expresses intentions to play L. We ask you to make a distinction between weak and strong expressions of intentions. An expression is strong if the sender emphasizes that this is definitely what he or she will do. Examples of strong expressions are "I will definitely play H", "I play H no matter what", "I will choose H and that is final." If you cannot infer any intention based on Sender 1's messages, or if the intention doesn't fit with the above two categories, you can indicate this by selecting the bottom option ("None of the above/I don't know").

The second question is the same as the first question, but for the other sender:

Question 2: Which intention does Sender 2 express?

The third question is whether or not they made some agreement.

Question 3: Did the two senders reach an agreement?

Choose from: *No; Yes, on both choosing L; Yes, on one choosing L and the other choosing H; I don't know.*

By reaching an agreement we mean that the senders know about each other's intentions, and they show some approval or confirmation (such as "ok" or "I agree" or "yes let's do that"). For instance, if Sender 1 wrote: "Let's both play L" and Sender 2 wrote "Ok", then they reached an agreement. If Sender 1 wrote "I play H" and Sender 2 wrote "I play L" and Sender 1 responded by writing "Ok" then they also reached an agreement. You should only classify the chat as reaching an agreement if the intentions of both players are clear. For instance, if Sender 1 writes "I play H" and Sender 2 writes "okay", then it is not clear what Sender 2 will choose, and therefore this should not be classified as an agreement. Similarly, you should only classify the chat as reaching an agreement if at least one of the players shows approval or confirmation. If Sender 1 writes "I play H" and Sender 2 writes "me too", then they did not reach an agreement because none of the players shows any approval or confirmation.

Please also pay attention to the following: What matters is the written intention at the end of the conversation. It can happen that players change their mind. In such cases, please classify messages according to the most recent statement of a player. For instance, suppose Sender 1 writes: "I will play H no matter what," Sender 2 responds with "Let's both play L", after which Sender 1 writes "Ok." In this case, we would classify Sender 1's message as "Will choose L" and we would classify this as an agreement to both play L. Sometimes players will speak of "High" and "Low" instead of "H" and "L", but they mean the same thing. You always need to answer all three questions. If there is no message by Sender 2, then please select "I don't know." You will see many chats. Please try to stay focused and take a break if you need to. After you have finished coding all chats, we will ask you to recode 50 randomly chosen chats. We will use this to measure the consistency of coders.

Appendix C: Outcomes and choices

Table 9 presents more detailed information about the outcomes and choices. Without communication, subjects' choices respond to c in agreement with the mixed Nash equilibrium. The higher c, the lower the probability that they choose H. Overall, they choose H with a smaller probability than in the mixed-strategy equilibrium though. In the absence of communication, the frequencies of H choices straightforwardly translate to outcomes, because subjects have no means to correlate their choices.

Table 9: Distribution of outcomes and choices

	% of outcomes and choices					
	(H,H)	(H,L)	(L,H)	(L,L)	H	Predicted H
No Communication						
BoS	49	22	21	8	70	80
C-Small	35	25	25	15	60	71
C-Large	17	27	26	30	44	50
Communication						
BoS	19	70	10	1	59	–
C-Small	33	40	16	12	61	–
C-Large	16	28	21	35	40	–

Notes: Entries are frequencies of each outcome and choice of H in percentage points. (H,L) [(L,H)] presents the percentage of outcomes that favor first [second] sender. The last column shows the theoretically predicted percentage points of H choice.

A different picture emerges when communication is allowed. Communication diminishes the frequency of H choices in *BoS*. There, the major effect of communication is that it helps subjects coordinate on the outcome that favors the person who can first send a message. Interestingly, even though communication does not affect aggregate payoffs in the Chicken games, it does lead to an increase in coordination on the equilibrium preferred by the first sender at the expense of the second sender, in particular in *C-Small*. Furthermore, when subjects are allowed to communicate there is a slight increase in the relative frequency of (L,L) outcomes in *C-Large*, but there is no such increase in *C-Small*.