The Power and Limits of Sequential Communication in Coordination Games

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

October 31, 2017

Abstract

We study theoretically and experimentally the extent to which communication can solve coordination problems. We do this in the context of coordination games in which there is some conflict of interest. In contrast to existing studies, we allow players to chat sequentially and free-format. The main behavioral assumption that we make, which we dub the 'feigned-ignorance principle', is that players will ignore any communication unless they reach an agreement in which both players are better off. 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.

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 1996a), existing theories of communication are focussed 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.[2]

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

[2] 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

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

To study how people will communicate, we need to make some behavioral assumptions about what will count as an agreement. The main assumption we introduce is the '*feigned-ignorance principle*'.[3] According to this principle, any agreement reached in the communication stage is only effective if sticking to 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 understand that no agreement has been reached, and will play according to the equilibrium 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 (cf. 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., Gneezy 2005). 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.

We put our theoretical predictions to an experimental test. The results support the

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).

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

main predictions. In a Battle-of-the-Sexes game, communication is very effective and 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) 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.

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] We think that the feigned-ignorance principle improves on the alternative approach based on the concepts of *self-committing* and *self-signalling* messages. That approach states that

[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."

[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 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).

cheap talk messages are only credible when the sender's message is both self-committing and self-signaling (Aumann 1990; Farrell and Rabin 1996b). 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. The 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. In Stag-Hunt games where the message to cooperate is not self-signaling, Charness (2000) finds a very strong effect of one-sided messages on subjects' willingness to choose the risky cooperative action while Clark et al. (2001) find a lesser but still substantial effect of two-sided communication.[6] 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 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.

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

[6]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.

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 make an announcement. In most actual cases , it is ex ante not commonly known 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 equilibrium play dramatically from 0.48 to 0.95. Two-sided communication is much less effective though. One round of two-way communication raises equilibrium play only to 0.55, and three rounds yields an equilibrium 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 equilibrium play 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. Section 4 discusses the experimental results and Section 5 concludes.

2 Theoretical Background

2.1 Preliminaries

We consider a two-player simultaneous-move normal-form game G. Each player chooses some action $A_i \in A = \{H, L\}$, with payoffs $u_i(A_i, A_j)$. The payoffs are given in Table 1. We assume that $a > b > 0$ and $a > c$. For $c = 0$, it reduces to a "Battle-of-the-Sexes" game. For $c > 0$, it has the structure of a "Chicken" game. Let S^* be the set of pure-strategy Nash equilibria of the game G. Note that the game has two Nash-equilibrium outcomes in pure strategies (H, L) and (L, H), and a mixed strategy Nash-equilibrium in which each player randomizes between H and L, playing H with probability $p = (a-c)/(a+b-c)$ and expected payoff $\mu = ab/(a+b-c)$.

Before choosing their actions, there is a pre-play communication stage C. In the 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. Players can send messages $m_i \in M = \{\hat{A}_i, \{\hat{A}_i, \hat{A}_j\}, \text{OK}, \emptyset\}$, where $\hat{A}_i, \hat{A}_j \in \{H, L\}$. A player can thus send a message indicating her own intended action $\hat{A}_i$, or a combination of her own intended action and the expected action of the other player $\{\hat{A}_i, \hat{A}_j\}$. A player can send "OK" to signal agreement with the other player's message. 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 with one of the players sending $m = \emptyset$. The message $m = \emptyset$ is costless, sending any other message costs $\gamma > 0$ to *each* player.

We refer to game G including the communication stage as the extended game $G^*(G, C)$. The communication stage consists of multiple periods. In each period, one of the players can send one message. Strategies in the game $G^*(G, C)$ are messages in the communication stage (possibly mixed and contingent on time and the opponent's messages) followed by probability distributions (possibly degenerate) over elements of A, where the probabilities can depend on the messages sent in the communication stage. Payoffs in G^* are $u_i(A_i, A_j) - \gamma T$, where T is the total number of non-empty messages sent by both players.

Table 1: Payoff matrix of Game G

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

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

Some terminology:

Definition 1. *A **conversation** is the sequence of messages $(m_1, m_2, \ldots, m_T)$ until one of the players terminates the communication stage by sending $m_{T+1} = \emptyset$ at time $T + 1$.*

Definition 2. *Messages m_h and m_k are **conflicting** if they stipulate actions (A_i or A_j) or outcomes (A_i, A_j) that are incompatible, i.e., $A_{ih} \neq A_{ik}$ and/or $A_{jh} \neq A_{jk}$, where A_{ik} is the action* *for player i specified in message k. A conversation ends with non-conflicting messages if (i) it only contains a single message, or (ii) it has length $T \geq 2$ and messages m_{T-1} and m_T are non-conflicting.*

Definition 3. *An **agreement** is reached when a conversation ends with non-conflicting messages.*

Definition 4. *A **credible agreement** is reached when a conversation ends with non-conflicting messages, and the last message(s) specifies actions (A_i, A_j) that constitute an equilibrium set of pure strategies of the game G, i.e., $(A_i, A_j) \in S^*$.*

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

We call an agreement **demanding** if player i proposed the outcome and it is player i's most preferred outcome. This is the case for conversations such as $\{..., \{\hat{H}, \hat{L}\}, \text{OK}\}$. If it is in the players' best interest to live up to the agreement in the decision stage, we call this a **demanding equilibrium**. We call an agreement **conceding** if player i proposed the outcome and it is player i's least preferred outcome. This is the case for conversations such as $\{..., \{\hat{L}, \hat{H}\}\}$ and $\{..., \{\hat{L}\}\}$. If it is in the players' best interest to live up to the agreement, we call this a **conceding equilibrium**. An agreement to play (L, L) is a **compromise**. Finally, we say that an **agreement is immediate** if it is not preceded by any other messages in the conversation.

2.2 Equilibrium

In this section we characterize the set of equilibrium strategies if players have standard preferences. In principle, the set of equilibrium strategies can be very large. To narrow down the set of possible equilibria, we assume a natural language interpretation of messages, such that $\hat{A}$ corresponds to the intention to play A, and make several behavioral assumptions.

Our first behavioral assumption is that players will play the focal strategy if no messages are sent or no agreement is reached. In general, it depends on the specifics of the game what the focal strategy is. In our context, 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 (cf. Farrell 1987). This assumption is largely supported by the data.

Our second behavioral assumption is that if players reach a credible agreement, they will only act in accordance with the agreement if it gives a higher utility to both players than if they disregard the conversation and play the mixed-strategy equilibrium. 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.'[7]

We now make the above more precise and summarize the main elements in Assumption 1. Suppose that the conversation ended with a credible agreement that specifies $(\hat{A}_i, \hat{A}_j)$ as the outcome.[8] Let $\tilde{u}_i^*$ represent player i's expected payoff in the outcome that results from the credible agreement, and μ_i^f the expected payoff of the focal strategy in the absence of communication. Recall that μ is the expected payoff from playing the mixed strategy in game G, which we take to be the focal strategy in the absence of communication, i.e., $\mu_i^f = \mu$.

Assumption 1 (Feigned-Ignorance Principle)**.** *If a conversation ends with a credible agreement, and $\tilde{u}_i^* \geq \mu$ for both players, then each player believes that the other player will act in accordance with the agreement. Otherwise, each player ignores the conversation and believes the other player will play according to the mixed-strategy equilibrium of the game G.*

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.[9] Assumption 1 also restricts admissable 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 $\{\hat{H}, \hat{L}\}$ and the other player responds with "OK," an agreement is reached, and players will not interpret this sequence of messages as mere babbling or disagreement.

The following proposition presents the equilibria in which the players use pure strate-

[7]We thus assume that players may ignore an agreement even if it is credible and one of the players explicitly agreed with the proposal (by sending the message "OK"). It is easy to verify that the set of equilibrium outcomes is unaltered if we instead assume that players act in accordance with the agreement in such a case, as the message "OK" would not occur in equilibrium.

[8]For games such as ours, in which there is a unique best-response to the action of the other player, it is enough for a player to specify an action to reach mutual understanding. When this is not the case, players need to specify the outcome to avoid ambiguity about the intended outcome.

[9]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).

gies in the communication stage (see Appendix A for proofs).

Let $\gamma_1 = \frac{b(b-c)}{a+b-c}$ and $\gamma_2 = \frac{a(a-c)}{a+b-c}$.

Proposition 1 (Pure strategy equilibria). *Under Assumption 1, there exist only two possible subgame perfect Nash equilibrium outcomes of conversations in pure strategies (in the communication stage) when $b > c$: (i) immediate concession is an equilibrium conversation outcome if and only if $\gamma \leq \gamma_1$, and it is followed by the Nash-equilibrium outcome (L,H) of game G; (ii) immediate demanding is an equilibrium conversation outcome if and only if $\gamma \leq \gamma_2$, and it is followed by the Nash-equilibrium outcome (H,L) of game G. When $b < c$, communication is ineffective and players refrain from sending costly messages.*

The advantage 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 (getting the high payoff a but at the cost of sending more messages). In the following proposition, we characterize this equilibrium.

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 unique equilibrium in mixed strategies in C in which 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.*[10] *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 be tough

[10]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 $\{\hat{L}, \hat{H}\}$. We assume that player 1 mixes with probability q_1 in the first period, though any probability is supported in equilibrium.

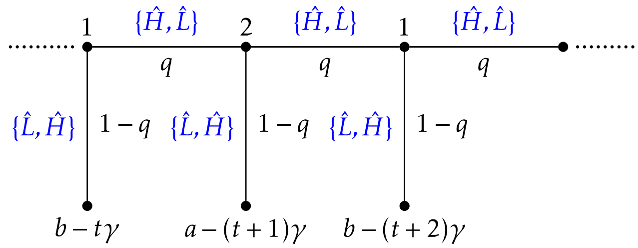

Figure 2: Part of a possible communication stage tree when period ≥ 4.

negotiators, 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. 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 are tough bargainers.

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

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 ($\mu > b$) for any $c > b$.

Given that communication can help for $b > c$ 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 equilibrium expected payoffs for different values of c. For relatively low values of c, there exist equilibria in which players communicate, increasing average 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. Even though without communication, payoffs are higher for higher values of c (compare Z to X), communication is ineffective for large c and average payoffs are lower than for lower values of c when players can communicate (compare Z to Y).

2.3 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 or guilt aversion (e.g., Gneezy 2005; Charness and Dufwenberg

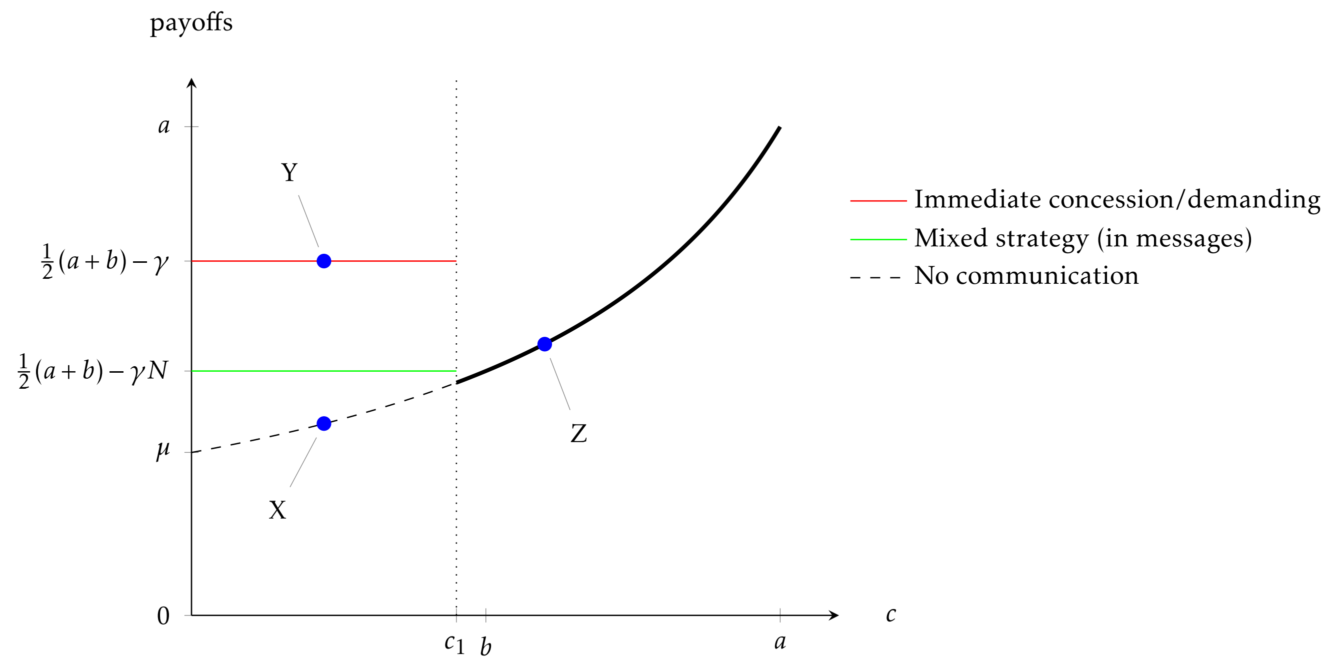

Figure 3: Average payoffs with and without communication for different equilibria. Notes: c_1 is the level that makes γ_1 equal γ, that is $\gamma = \frac{b(b-c_1)}{(a+b-c_1)}$. For $c > c_1$ there are no equilibria in which players send messages. The figure is drawn for (a, b, γ)=$(200, 100, 5)$.

2006; Vanberg 2008; Gneezy et al. 2013). Introducing a cost of lying does not affect any of the previously characterized equilibria, as in those equilibria players act in accordance with the agreement reached.[11] However, one may speculate that costs of lying can make communication more effective, and this turns out to be the case.

In this section we analyze the effects of lying aversion. We assume that people differ in the lying costs that they suffer whenever they deviate from what was intended in an agreement. In such a case, player i experiences lying cost k_i, which means that an amount of k_i is subtracted from her monetary outcome. Each player knows that players' lying costs are independently drawn from a continuous and strictly increasing cumulative distribution function $F(\cdot)$ that has full support on $[0, \bar{k}]$, and that each player i is only privately informed of her own k_i at the start of the game. We refer to this Bayesian game as $G^*(BG, C)$.

We now modify our main assumption to account for the presence of lying costs and incomplete information. First, we modify the expected utility to $\tilde{u}_i(k_i)$, which is player i's utility including the lying cost of deviating from an agreement. Second, with the presence of incomplete information, we apply Perfect Bayesian Equilibrium (PBE) as the solution concept. Let $\tilde{u}_i^*(k_i)$ represent the expected payoff of type k_i of player i in the PBE after reaching a credible agreement. In the game $G^*(BG, C)$, a credible agreement is reached when the conversation ends with non-conflicting messages, and the last messages specify actions for each type such that they constitute a PBE of the game $G^*(BG, C)$. Without a credible agreement, players do not experience costs of lying and each type of each player mixes with the same probability as in the game G with perfect information. In that case, the expected payoff from playing the mixed strategy equilibrium in game BG is still μ.

Assumption 1′. *If a conversation ends with a credible agreement, and $\tilde{u}_i^*(k_i) \geq \mu$ for each type of each player, then each type of each player believes that the other player will act in accordance with the equilibrium strategies in the PBE of the associated agreement. Otherwise, each type of each player believes that all the types of the other player will play according to the mixed-strategy equilibrium of the game BG.*

This assumption is in the same spirit as the definition and assumption for the case with complete information. The approach for complete information is a special case of

[11] 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 analysis in this section. We assume that if players reach an agreement to play (L,L), they understand that play will be in agreement with the PBE of that agreement, which means that players with a relatively high cost of lying will conform to the agreement while players with a relatively low cost of lying will deviate. If the fraction of the deviators is sufficiently low, a Bayesian Nash equilibrium exists in which any type of player finds it profitable to agree on (L,L).

We look for pure strategy equilibria in which players' strategies are functions from their type to the decision to reach an agreement or not in the conversation, and further to the decision to conform or deviate from the agreement if they make one. As is often the case, there is a multiplicity of equilibria, depending on players' beliefs. For instance, if all players believe that a proposal to play (L,L) only comes from the type with $k=0$, then no player wants to reach an agreement on (L,L). We focus on the equilibrium in which all types of players agree on (L,L). The equilibrium strategy 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). This implies that if players reach an agreement they expect the other player to deviate from the agreement with probability $F(k^*)$. In equilibrium, they reach such an agreement after one message. A player i who has a lying cost k_i is then indifferent between H and L when:

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

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

Another 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:

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

Moreover, in the role of second sender, the player with $k_i = k^*$ should not be willing to accept his or her least preferred equilibrium payoff b, otherwise the first sender should propose (H, L) instead of (L, L). This requires that $c > b$.

Proposition 3. *Under Assumption 1′, an agreement to play* (L, L) *can be supported as a Perfect Bayesian Equilibrium conversation outcome if and only 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.*

2.4 Summary

In the result section we will investigate which equilibrium is played by our subjects (if any). In addition, we will consider the following testable insights of the model.

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 explicit disagreement.

3. When players do not reach a credible agreement, or when the credible agreement is worse than the mixed strategy equilibrium 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 first or second mover is always favored, or it may be long and then the potential gains of communication are 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.

3 Experimental Design and Procedures

3.1 Treatment design

In the experiment, we implemented the payoff matrix of Table 1. Table 2 summarizes the different treatments. 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.[12]

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.

[12] 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.

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

The experiment 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 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. Average 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

4.1 Effects of Communication

Without communication, the average 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*.[13] 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-

[13]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.

sided Mann-Whitney test).[14] 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.

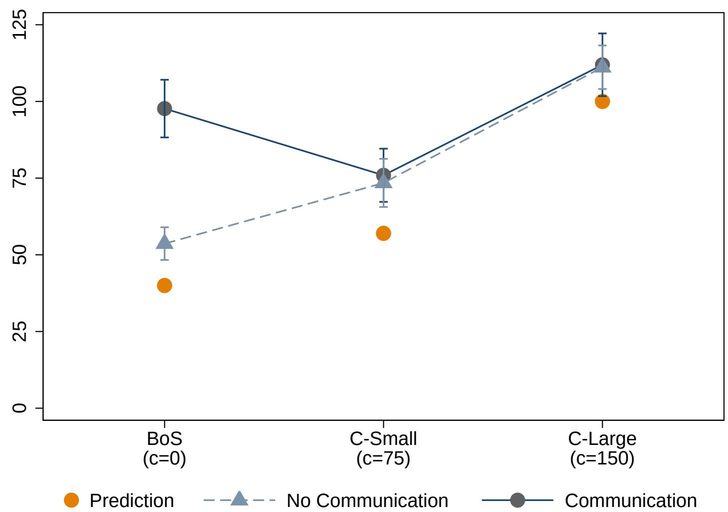

Figure 4: Average payoffs by treatment and communication. "Prediction" is the expected payoff if subjects play the mixed-strategy Nash equilibrium of game G. Bars are the 95% confidence intervals.

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

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: 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 average 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* ($p = 0.529$).[15]

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).

[15]Table 7 in Appendix C presents details about the frequencies of outcomes in the different treatments.

Table 4: Average payoffs of first and second sender

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

Notes: 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)**.** *The option to communicate substantially increases average earnings in the Battle-of-the-Sexes 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 communication?

What explains the differences in effectiveness of 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.[16] 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 support our assumption that subjects play the mixed-strategy equilibrium when they immediately end the conversation without talking.

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)**.** *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.*

[16] 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).

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.

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. 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*.[17] 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 small and insignificant increase of 2 percentage points in the likelihood of choosing H

[17] 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.

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)**.** *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, average earnings are not higher than without communication. This is illustrated in Figure 5, that shows average earnings by the length of the conversation. In *BoS*, average earnings are highest when only one message is sent. Sending more messages is associated with lower average earnings, probably because longer conversations are a consequence of disagreement. In the Chicken games, average earnings are close to the average earnings without communication when two or fewer messages are sent. For three or more messages, average 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).

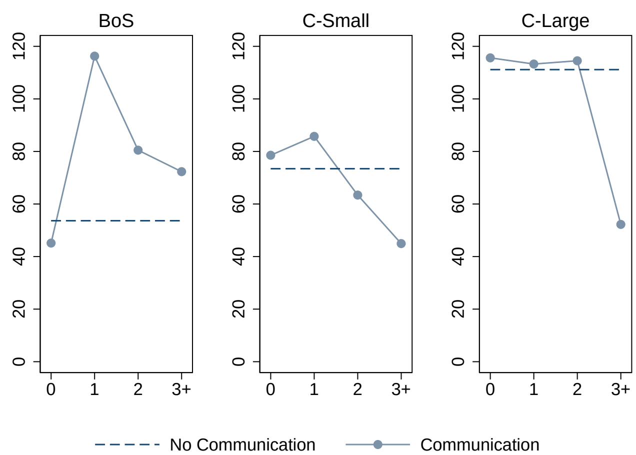

Figure 5: Average earnings by number of messages sent. The horizontal (dashed) lines are reference lines showing the average earnings without communication.

Result 5 (Communication length)**.** *In BoS, subjects use the option to communicate. 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 2.4: (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 average 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.

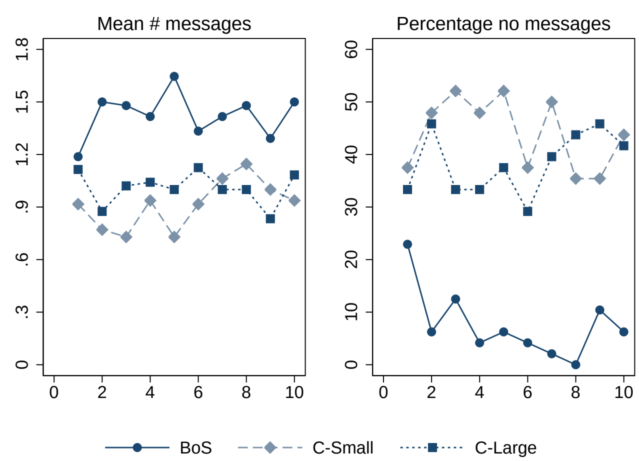

Figure 6: Average number of messages over rounds.

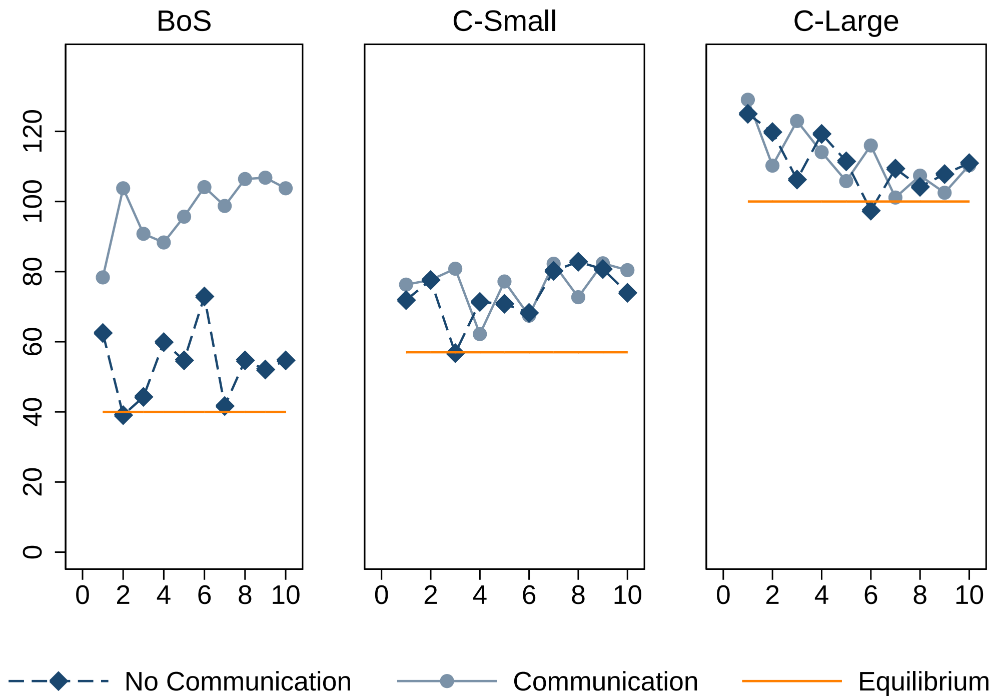

Figure 7: Average earnings over rounds. "Equilibrium" is the expected payoff if subjects play the mixed-strategy Nash-equilibrium of game G.

5 Conclusion

In this paper, we investigated how people play coordination games with conflicting interests when they have the possibility to sequentially send non-binding messages. Theoretically, we found that the effectiveness of sequential 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 $c < b$, as in the Battle-of-the-Sexes, then sequential communication is predicted to solve the coordination problem. Theoretically, 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.

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 and no conversation will be started. 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, 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. Lying averse players will then conform to the agreement, while lying neutral players deviate. The higher c, the more conforming behavior is to be expected.

In the experiment, we find that communication works like a charm in the Battle-of-the-Sexes. 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 in the Chicken games. In the aggregate, it does not allow subjects to benefit and subjects often simply forgo the possibility to talk. 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.

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. 2010; 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.

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.

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.

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.

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

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. (1996a). Cheap talk. *Journal of Economic Perspectives*, 10:103–118.

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

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.

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

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 of all, we show that any equilibrium conversation can be at most 2 messages long if players use pure strategies for sending messages. Clearly, it cannot be optimal to never stop sending messages, as this is costly and players can deviate to $m = \emptyset$ (in their first message) making them strictly better off. Thus, in any equilibrium the number of messages is finite. There also has to be a credible agreement that yield higher payoffs than mixed strategy equilibrium without communication, otherwise players can deviate to $m = \emptyset$ earlier in the game. Note that the only possible credible agreement that yields better payoff than mixed strategy equilibrium for both players result in playing (H, L) or (L, H) and requires $b > c$. Now suppose there would be $n \geq 2$ messages, then the conversation must ends with $\{\hat{L}, \hat{H}\}$; otherwise, if $\{\hat{L}, \hat{H}\}$ occurs earlier in the conversation, the other player best-responds by ending the communication stage. But then the player that receives L could deviate to sending message $\{\hat{L}, \hat{H}\}$ at $n-2$ messages or send $m = \emptyset$ at the second message if $n = 2$, after which the other player ends the communication stage. This way both players receive the same payoffs in G and save on message costs. This implies that in equilibrium $n \leq 1$. Suppose that $n = 0$, then the first sender can deviate to $\{\hat{L}, \hat{H}\}$, after which the second sender best-responds by ending the communication stage; both players receive higher payoffs than the mixed strategy equilibrium without communication. In equilibrium we must have $n = 1$, in which the first sender demands or concedes, and the second sender ends the communication stage immediately; in the game stage they both play according to the first message. $\{\{\hat{L}, \hat{H}\}\}$ can, for instance, be supported as an equilibrium outcome by the strategies "send $\{\hat{L}, \hat{H}\}$" for player 1 after conflicting messages and "send $m = \emptyset$" after a credible agreement, and "send $\hat{H}$ until player 1 sends $\{\hat{L}, \hat{H}\}$ and then send $m = \emptyset$" for player 2. Player 2 clearly does not want to deviate to sending another non-empty message at any point. Player 1 should also not wish to deviate to $m = \emptyset$, yielding the threshold level $\gamma_1 = \frac{b(b-c)}{a+b-c}$. The conversation $\{\{\hat{H}, \hat{L}\}\}$ can, for instance, be supported as an equilibrium outcome by the strategies "send $\hat{H}$ until player 2 sends $\{\hat{L}, \hat{H}\}$ and then send $\emptyset$" for player 1, and "send $\emptyset$" if there is a credible agreement and "send $\{\hat{L}, \hat{H}\}$ if there is conflicting messages" for player 2. Player 2 clearly does not want to deviate to sending another non-empty message at any point. Player 1 should not wish to deviate to $m = \emptyset$, yielding $\gamma_2 = \frac{a(a-c)}{a+b-c}$ and $b > c$.

Proof of Proposition 2. In the second period, player 2's probability q_1 of demanding must make player 1 indifferent between demanding (by sending $\{\hat{H},\hat{L}\}$) and conceding (by sending $\{\hat{L},\hat{H}\}$) in the first period, thus we have: $(1-q_1)(a-\gamma)+q_1(b-3\gamma)=b-\gamma$, which gives $q_1=\frac{a-b}{a-b+2\gamma}$. In the third period, player 1's probability q_2 of demanding must make player 2 indifferent between demanding (by sending $\{\hat{H},\hat{L}\}$) and conceding (by sending $\emptyset$) in the second period, thus we have: $(1-q_2)(a-3\gamma)+q_2(b-4\gamma)=b-\gamma$, which gives $q_2=\frac{a-b-2\gamma}{a-b+\gamma}$. If the conversation continues after three periods, there must be conflicting messages. A player must be indifferent between inducing agreement (by sending $\{\hat{L},\hat{H}\}$) or let the communication stage continue (by sending $\{\hat{H},\hat{L}\}$). thus we have: $(1-q)(a-2\gamma)+q(b-3\gamma)=b-\gamma$, which gives $q=\frac{a-b-\gamma}{a-b+\gamma}$.

The communication stage becomes static after t ($t\geq 3$) messages. Sending $\{\hat{L},\hat{H}\}$ yields a payoff of $b-t\gamma$ to player 1. After sending $\{\hat{H},\hat{L}\}$, player 2 sends $\{\hat{L},\hat{H}\}$ with probability $1-q$ (yielding a payoff of $a-(t+1)\gamma$ to player 1) and sends $\{\hat{H},\hat{L}\}$ with probability q. Denote the continuation payoff of reaching period $t+2$ by V. Player 1 must be indifferent between $b-t\gamma$ and $(1-q)(a-(t+1)\gamma+qV$. At period $t+2$, if reached, player 2 should again be indifferent between sending $\{\hat{H},\hat{L}\}$ and $\{\hat{L},\hat{H}\}$. Sending $\{\hat{L},\hat{H}\}$ 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 is true at any other period in the communication stage, except for the first three messages. Player 1 is indifferent between inducing agreement or continuing, and any probability of inducing agreement will support the equilibrium. Player 2 does not need to send $\{\hat{L},\hat{H}\}$ or $\{\hat{L},\hat{H}\}$ to signal agreement (leaving the conversation signals agreement), since there is no conflicting messages yet. However, for the sake of simplicity, we assume that player 1 also mixes with probability q_1 and $1-q_1$ in the first message, and that player 2 sends $\{\hat{L},\hat{H}\}$ or $\{\hat{L},\hat{H}\}$ in the second message as if there had been conflicting messages. Finally note that no player should wish to deviate to sending $m=\emptyset$, which is the case if $\gamma\leq\gamma_1$.

The expected number of messages is the sum of pt at each period. The probability that the conversation ends with one message is $(1-q_1)+q_1(1-q_1)$; the probability that the conversation ends with two message is 0; the probability that the conversation ends with three messages is $q_1^2(1-q_2)$; the probability that the conversation ends with four messages is $q_1^2q_2(1-q)$; the probability that the conversation ends with five messages is $q_1^2q_2q(1-q)$; the probability that the conversation ends with six messages is $q_1^2q_2q^2(1-q)$;...and the probability that the conversation ends with t (any $t\geq 4$) messages is $q_1^2q_2q^{t-4}(1-q)$. We therefore derive the expected number of messages $N=1+2q_1^2+\frac{q_1^2q_2}{1-q}$.

Proof of Proposition 3. We first show that, if *some* types have beliefs such that it is attractive for them to reach an agreement, then all other types find it attractive too. To see this, note that in any equilibrium in which they agree on (L, L), some types must prefer to conform to the agreement (if no type would conform, no type with positive k would be willing to enter into an agreement). Since the payoff of conforming to an agreement does not depend on k, and neither does the payoff after no agreement, this implies that all others types also prefer to conform to the agreement compared to no agreement, and possibly can do even better by deviating from the agreement. This implies that in equilibrium, either everyone's strategy is to reach an agreement after one message, or no one reaches an agreement at all.

We next show that when there exists a threshold k^* such that players can find it attractive to reach an agreement on (L, L), then players with a cost of lying lower than k^* deviate from the agreement and players with a cost of lying higher than k^* conform to the agreement. Consider that in equilibrium a player with a cost of lying k_0 deviate from the agreement, then a player with a cost of lying lower than k_0 will also deviate from the agreement as he or she suffer a lower cost from deviating. Similarly, if in equilibrium a player with a cost of lying k_1 conform to the agreement, then a player with a cost of lying higher than k_1 will also conform to the agreement as he or she suffer a higher cost from deviating.

Finally, we show that in this equilibrium, all players reach an agreement in the first message. Suppose they send a number of $n \geq 3$ messages to reach an agreement on (L, L), then the conversation must ends with $\{\{\hat{L}, \hat{L}\}\}$ or "OK". Then the last sender can deviate to sending message $\{\{\hat{L}, \hat{L}\}\}$ at $n - 2$ message to reach an agreement and save communication cost. If $n = 2$, then either the first sender can deviate to sending $\{\{\hat{L}, \hat{L}\}\}$ at $n = 1$, or the second sender can deviate to leaving the chat if first sender sent message $\{\{\hat{L}, \hat{L}\}\}$ at $n = 1$.

Appendix B: Experimental and Coding instructions

Experimental Instructions (for subjects)

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. 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.

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.

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 7 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 7: 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*.