Communication, Renegotiation and Coordination with Private Values

Yuval Heller and Christoph Kuzmics

July 29, 2021

Abstract

An equilibrium is communication-proof if it is unaffected by new opportunities to communicate and renegotiate. We characterize the set of equilibria of coordination games with pre-play communication in which players have private preferences over the feasible coordinated outcomes. Communication-proof equilibria provide a narrow selection from the large set of qualitatively diverse Bayesian Nash equilibria in such games. Under a communication-proof equilibrium, players never miscoordinate, play their jointly preferred outcome whenever there is one, and communicate only the ordinal part of their preferences. Moreover, such equilibria are robust to changes in players’ beliefs, interim Pareto efficient, and evolutionarily stable.

Keywords: cheap talk, communication-proofness, renegotiation-proofness, secret handshake, incomplete information, evolutionary robustness. JEL codes: C72, C73, D82

1 Introduction

We characterize communication-proof equilibria for a class of coordination games with pre-play cheap-talk communication in which all agents have private information about what...
action they would prefer to coordinate on. A Bayesian Nash equilibrium is communication-proof if, after the pre-play cheap talk and given the information that this reveals, a new opportunity for additional communication does not allow the players to jointly deviate to a Pareto-improving equilibrium.¹

We are interested in two typical kinds of situations for which communication-proofness is an appropriate solution concept, albeit for different reasons in the two situations. The first kind of situation is one in which agents are sophisticated and keep (strategically) communicating until they reach a mutually beneficial solution. Communication-proofness is defined to capture this idea, similarly to the notions of renegotiation-proofness in contract theory (Hart and Tirole, 1988) or in the repeated games literature (Farrell and Maskin, 1989). As an example, consider a situation of two firms trying to collude by implementing a market-sharing agreement by which one firm sells in certain regions whereas the rival sells in other regions, and each firm has private information about which regions they prefer to serve.

The second kind of situation is one in which communication is feasible and in which behavior is governed by a long-run learning (or evolutionary) process. Communication-proofness here corresponds to a requirement of evolutionary stability at an interim level, when agents can experiment with new behavior that is contingent on the use of additional communication (Robson, 1990). As an example consider the problem of two pedestrians suddenly finding themselves face-to-face and trying to get past each other, when they have private information about the direction they want to take after the encounter.

The standard solution concept of Bayesian Nash equilibrium is not helpful in predicting whether players can achieve coordination in such incomplete-information settings, how efficient it is if they do, and how communication is used to achieve it. Coordination games with pre-play communication have a wide range of qualitatively very different equilibria. Among these are babbling equilibria with a high likelihood of miscoordination that are evolutionarily stable in the absence of communication, and equilibria in which agents reveal some information about the intensity of their preferences and yet often miscoordinate.

Casual observation suggests that players manage to coordinate in at least some such situations. Considering an instance of our first example, we note that firms “competing” in the 1997 series of regional FCC auctions allocating licenses for slices of the electromagnetic spectrum were able to use the very limited public communication possibilities of the trailing digits of their bids to reveal information on their preferred regions in order to successfully

¹The notion of communication-proofness was introduced by Blume and Sobel (1995) in their study of sender-receiver games with one-sided private information.
coordinate to collude (Cramton and Schwartz, 2000).

Players also typically coordinate effectively in our second example: Pedestrians typically are able to avoid bumping into each other, even though there is no uniform social norm such as “always stay on the right” as there is for cars (Young, 1998). Moreover, pedestrians often use brief nonverbal communication to signal their preferred direction (e.g., a slight movement to the left or right, a tilt of the head, a glance in a certain direction), and the (coordinated) direction in which they pass each other depends on this communication.

We show that communication-proof equilibria have a specific structure that is consistent with these casual observations. We show that a strategy is a communication-proof equilibrium if and only if it satisfies the following three independent and easy to verify properties: players never miscoordinate, they play their jointly preferred outcome whenever there is one, and they communicate only the ordinal part of their preferences (i.e., communicate their preferred outcome without revealing any information on the intensity of their preferences).

The equilibria that satisfy these properties have a simple structure. In all these equilibria communication induces the agents to endogenously face games in which their ordinal preferences are common knowledge. In those cases in which agents agree about the optimal joint action, they coordinate efficiently, i.e., on the action that both prefer. In cases in which they disagree, they still coordinate, but, as the coordinated outcome is not determined by the players’ cardinal preferences, this coordination is generally not ex-ante efficient.

Finally, we show that communication-proof equilibria do not depend on the distribution of private preferences and are thus robust to changes in players’ (first- or higher-order) beliefs. In particular, communication-proof equilibrium strategies remain communication-proof even in setups in which the players’ distributions of types are interdependent. Also, communication-proof equilibria do not depend on the exact timing of the renegotiation (relative to the communication).

**Relationship to the literature** Game theorists have long recognized that coordination is an important aspect of successful economic and social interaction, that it requires an explanation even in complete-information coordination games, and that it does not occur in all circumstances. One possible explanation for some, fairly simple, examples of coordination

---

2In fact, the result of this paper that communication that relies on each player simultaneously sending either 0 or 1 is all that is needed for successful coordination provides another argument against allowing even a brief form of explicit communication between oligopolistic competitors.

3This is motivated by Goffman (1971, p. 6): “Take, for example, techniques that pedestrians employ in order to avoid bumping into one another. [...] There are an appreciable number of such devices; they are constantly in use and they cast a pattern on street behavior. Street traffic would be a shambles without them.”
is the concept of a focal point, due to Schelling (1960), which is, loosely speaking, a strategy profile that jumps out at players as clearly the right way to play a game. Perhaps one of the situations in which we most plausibly expect coordination is when people play the same coordination game many times with different people and there is some evolutionary (or learning) process. This approach is already present in the “mass action” interpretation of equilibrium given by Nash (1950), and then taken up more formally in Maynard Smith and Price (1973) who define the notion of evolutionary stability. It is well known that all pure equilibria in coordination games are evolutionarily stable (whereas mixed equilibria are not stable). This literature thus supports the view that while play in the long run will be coordinated, it is not necessarily efficiently coordinated.4

Another explanation for coordination is that it is achieved through communication, even if it is simply cheap talk as in Crawford and Sobel (1982). Early seminal contributions in this direction are Farrell (1987) and Rabin (1994). Communication alone, however, only adds equilibria: the equilibria of the game without communication “survive” the introduction of communication as babbling equilibria. The problem, therefore, of how play focuses on the coordinated equilibria does not go away, and one can again appeal to one of the above-mentioned criteria to explain why this might happen.

There is a literature that studies the evolutionary outcome of coordination games with cheap talk, initiated by Robson (1990). If play is stuck in an inferior equilibrium, a small group of experimenting agents can recognize each other by means of a “secret handshake” and play Pareto-optimal strategy with each other and the inferior equilibrium strategy with agents who are not part of this group, thereby outperforming the agents outside the group.

The above-mentioned literature focuses on complete-information games. However, one of the main reasons why people communicate is that they have privately known preferences that they feel useful to share at least partially before finally choosing actions, as seen in the above examples. One of the main stumbling blocks of studying how communication helps achieve coordination in the presence of incomplete information is that it “requires overcoming formidable multiple-equilibrium problems” (Crawford and Haller, 1990, p. 592).

We identify Blume and Sobel’s (1995) notion of communication-proofness, adapted to our two-sided private information setting, as the appropriate extension of the secret-handshake argument to incomplete-information games. With our characterization result we then show that the plausible refinement of communication-proof equilibria removes this mul-

4Kandori et al. (1993) and Young (1993) show that in the long run and under persistent low-probability errors an evolutionary process leads to the risk-dominant, not necessarily Pareto-dominant, equilibrium.
tiplicity problem to a large extent, in the sense that all communication-proof equilibria, in contrast to Bayesian Nash equilibria, make very similar predictions.

**Structure** Section 2 presents our model. Section 3 defines Bayesian Nash equilibria and the three key properties that communication-proof equilibria have. Section 4 defines the concept of communication-proofness. Section 5 presents the main result and a sketch of its proof. Section 6 discusses the efficiency properties of communication-proof equilibria. Section 7 identifies the scope and limits of the main result (within the class of two player two action incomplete information coordination games). Section 8 concludes with an additional discussion of related literature and a brief discussion of extensions of this model that are described in full detail in the extended working paper version of this paper Heller and Kuzmics (2021). The formal proofs are presented in the appendix.

## 2 Model

We consider a setup in which two agents with private idiosyncratic preferences play a two-action coordination game that is preceded by pre-play cheap talk.

**Players and types** There are two players, each of which can choose one of two actions, \( L \) and \( R \). Each player has a privately known “value” or “type.” The two players’ values are independently drawn from a common atomless distribution with a continuous cumulative distribution function \( F \) with full support on the unit interval \( U = [0, 1] \) and with density \( f \) (i.e., \( f(u) > 0 \) for each \( u \in U \)).

**Payoff matrix** For any realized pair of types, \( u \) and \( v \), the players play a coordination game given by the following payoff matrix, where the first entry is the payoff of the player of type \( u \) (choosing row) and the second entry is the payoff of the player of type \( v \) (choosing column). We call this game the *coordination game without communication* and denote it by \( \Gamma \).

**Interpretation of the model and the motivating examples** In the example of two firms trying to collude by market-sharing, choosing the same action corresponds to dividing the

---

5Allowing distributions without full support induces a minor difference in our results: in this setup communication-proofness implies binary communication (as defined in Section 3) only of messages that are used with positive probability, whereas with full support it implies binary communication also of unused messages.
market such that each firm is a monopolist in one of the two regions, and choosing different actions corresponds to the firms competing in the same region, which yields a low profit normalized to zero. A firm’s type $u$ corresponds to how profitable it is for the firm to be a monopolist in one region relative to being one in the other region.

In the second motivating example of pedestrians suddenly finding themselves face to face and trying to get past each other, each action corresponds to the direction in which the pedestrians turn to avoid bumping into each other. When both pedestrians choose the same side (say, each pedestrian chooses her left), the pedestrians do not bump into each other, whereas when they choose different sides they do bump into each other, in which case they get a low payoff normalized to zero. A pedestrian’s type reflects her private preference for the direction in which she would like to turn to avoid a collision due to the direction she plans to take after the encounter. That is, a type $u > 1/2$ corresponds to a pedestrian who plans to head right after the encounter, and thus choosing $R$ is more convenient than choosing $L$ as it induces a shorter walking path.

### Pre-play communication

After learning their type, but before playing this coordination game, the two players each simultaneously send a publicly observable message from a finite set of messages $M$ (satisfying $4 \leq |M| < \infty$), where $\Delta(M)$ is the set of all probability distributions over messages in $M$.\(^6\)\(^7\) We assume that messages are costless. We call the game, so amended, the coordination game with communication and denote it by $(\Gamma,M)$.

### Strategies

A player’s (ex-ante) strategy in the coordination game with communication is then a pair $\sigma = (\mu, \xi)$, where $\mu : U \to \Delta(M)$ is a (Lebesgue measurable) message function.

\(^6\)Our results essentially remain the same if $M$ is countably infinite. The assumption of $|M| \geq 4$ implies that a single round of communication during the renegotiation stage can achieve a sufficient degree of communication for our main results to hold (see Section 4). Our results remain the same for $M = 2$ if one allows the players during the renegotiation stage to either have two stages of communication or to rely on a (binary) sunspot.

\(^7\)In Appendix C we show that, communication-proof equilibria in coordination games are unaffected by the length (number of rounds) of communication (in contrast to the results in other setups of incomplete-information games; see, e.g., Aumann and Hart, 2003).
that describes which (possibly random) message is sent for each possible realization of the
agent’s type, and \(\xi : M \times M \to U\) is an action function that describes the maximal type (cutoff
type) that chooses \(L\) as a function of the observed message profile; that is, when an agent who
follows strategy \((\mu, \xi)\) observes a message profile \((m, m')\) (message \(m\) sent by the agent, and
message \(m'\) sent by the opponent), then the agent plays \(L\) if her type \(u\) is at most \(\xi (m, m')\)
(i.e., if \(u \leq \xi (m, m')\)), and she plays \(R\) if \(u > \xi (m, m')\).

Let \(\mu_u (m)\) denote the probability, given message function \(\mu\), that a player sends message
\(m\) if she is of type \(u\). Let \(\bar{\mu} (m) = \mathbb{E}_u [\mu_u (m)]\) be the mean probability that a player of
a random type sends message \(m\) (where the expectation is taken with respect to \(F\)). Let
\(\text{supp} (\bar{\mu}) = \{m \in M \mid \bar{\mu} (m) > 0\}\) denote the support of \(\bar{\mu}\). We say that message \(m\) is in the
support of \(\sigma = (\mu, \xi)\), denoted by \(m \in \text{supp} (\sigma)\), if \(m \in \text{supp} (\bar{\mu})\).

With a slight abuse of notation we write \(\xi (m, m') = L\) when all types (who send message
\(m\) with positive probability) play \(L\) (i.e., when \(\xi (m, m') \geq \sup (u \in U \mid \mu_u (m) > 0))\), and we
write \(\xi (m, m') = R\) when all types play \(R\) (i.e., when \(\xi (m, m') \leq \inf (u \in U \mid \mu_u (m) > 0))\).

3 Equilibrium Strategies and Three Key Properties

We here define the standard notion of (Bayesian Nash) equilibrium strategies, present the
three key properties that communication-proof equilibria turn out to have, and present ex-
amples of equilibria in the coordination game with communication with and without these
properties. These equilibria are illustrated in Figure 1 at the end of this section.

Given a strategy profile \((\sigma, \sigma')\) and a type profile \(u, v \in U\), let \(\pi_{u,v} (\sigma, \sigma')\) denote the
payoff of a player of type \(u\) who follows strategy \(\sigma\) and faces an opponent of type \(v\) who
follows strategy \(\sigma'\). Formally, for \(\sigma = (\mu, \xi)\) and \(\sigma' = (\mu', \xi')\),

\[
\pi_{u,v} (\sigma, \sigma') = \sum_{m \in M} \sum_{m' \in M} \mu_u (m) \mu_v (m') \left( (1 - u) \mathbb{1}_{\{u \leq \xi (m, m')\}} \mathbb{1}_{\{v \leq \xi' (m', m)\}} + u \mathbb{1}_{\{u > \xi (m, m')\}} \mathbb{1}_{\{v > \xi' (m', m)\}} \right),
\]

where \(\mathbb{1}_{\{x\}}\) is the indicator function equal to 1 if statement \(x\) is true and zero otherwise. Let

\[
\pi_u (\sigma, \sigma') = \mathbb{E}_v [\pi_{u,v} (\sigma, \sigma')] = \int_{v=0}^{1} \pi_{u,v} (\sigma, \sigma') f (v) \, dv
\]

\(8\)In Appendix A.1 we show that the restriction to cut-off strategies is without loss of generality: any “generalized” strategy \(\xi : M \times M \to U\) is dominated by a strategy with a “threshold” action function.
denote the expected interim payoff of a player of type \( u \) who follows strategy \( \sigma \) and faces an opponent with a random type who follows strategy \( \sigma' \). Finally, let,
\[
\pi(\sigma, \sigma') = \mathbb{E}_u [\pi_u(\sigma, \sigma')] = \int_{u=0}^1 \pi_u(\sigma, \sigma') f(u) du
\]
denote the ex-ante expected payoff of an agent who uses strategy \( \sigma \) against strategy \( \sigma' \).

A strategy \( \sigma \) is a \textit{(symmetric Bayesian Nash) equilibrium strategy} if \( \pi_u(\sigma, \sigma) \geq \pi_u(\sigma', \sigma) \) for each \( u \in U \) and each strategy \( \sigma' \in \Sigma \). Let \( \mathcal{E} \subseteq \Sigma \) denote the set of all equilibrium strategies of \( \langle \Gamma, M \rangle \).

Three key properties We call a strategy \( \sigma = (\mu, \xi) \in \Sigma \) \textit{mutual-preference consistent} if whenever \( u, v < 1/2 \) then \( \xi(m, m') = \xi(m', m) = L \) for all \( m \in \text{supp}(\mu_u) \) and all \( m' \in \text{supp}(\mu_v) \) and if whenever \( u, v > 1/2 \) then \( \xi(m, m') = \xi(m', m) = R \) for all \( m \in \text{supp}(\mu_u) \) and all \( m' \in \text{supp}(\mu_v) \). That is, players with the same ordinal preference coordinate on their mutually preferred outcome.

We call a strategy \textit{coordinated} if \( \xi(m, m') = \xi(m', m) \in \{L, R\} \) for any messages \( m, m' \in \text{supp}(\bar{\mu}) \). A coordinated strategy leads to a coordinated outcome with probability one.

For any message \( m \in M \), define the expected probability of a player’s opponent playing \( L \) conditional on the player sending \( m \) and the opponent following \( \sigma = (\mu, \xi) \in \Sigma \), as
\[
\beta^\sigma(m) = \int_{u=0}^1 \sum_{m' \in \text{supp}(\mu_u)} \mu_u(m') 1_{\{u \leq \xi(m', m)\}} f(u) du.
\]

We say that strategy \( \sigma \) has \textit{binary communication} if there are two numbers \( 0 \leq \underline{\beta}^\sigma \leq \overline{\beta}^\sigma \leq 1 \) such that for all messages \( m \in M \) we have \( \beta^\sigma(m) \in [\underline{\beta}^\sigma, \overline{\beta}^\sigma] \), for all messages \( m \in M \) such that there is a type \( u < 1/2 \) with \( \mu_u(m) > 0 \) we have \( \beta^\sigma(m) = \overline{\beta}^\sigma \), and for all messages \( m \in M \) such that there is a type \( u > 1/2 \) with \( \mu_u(m) > 0 \) we have \( \beta^\sigma(m) = \underline{\beta}^\sigma \). That is, binary communication implies that players (essentially) use just two kinds of messages: any message sent by types \( u < 1/2 \) induces the same consequence of maximizing the probability of the opponent playing \( L \), and any message sent by types \( u > 1/2 \) induces the opposite consequence of maximizing the opponent’s probability of playing \( R \). Note that, as defined here, a strategy with no communication also has binary communication (in which the player’s message does not affect the probability of the partner playing \( L \)).

In Appendix B we show that none of these three properties is implied by the other two.
Left tendency $\alpha$  Consider a strategy that satisfies the above three properties. Coordination and mutual-preference consistency jointly determine the behavior of agents with the same ordinal preferences (i.e., when both types are below $1/2$, or both above $1/2$). The property of binary communication, then, implies that the probability with which the players coordinate on $L$, conditional on having different ordinal preferences (i.e., conditional on one player having type $u < 1/2$ and the other player having type $v > 1/2$), is independent on the message sent by the player. We denote this probability by $\alpha$, and refer to it as the left tendency of the strategy. We can express $\beta$ and $\bar{\beta}$ as follows:

$$\beta = F(1/2)\alpha$$ and $$\bar{\beta} = F(1/2) + (1 - F(1/2))\alpha.$$ 

The first equality ($\beta = F(1/2)\alpha$) is implied by the fact that when any type $u > 1/2$ sends a message expressing her preference for coordination on $R$, the players coordinate on $L$ only if the opponent’s preferred outcome is $L$ (which happens with a probability of $F(1/2)$), and they then coordinate on $L$ with a probability of $\alpha$. The second equality ($\bar{\beta} = F(1/2) + (1 - F(1/2))\alpha$) is implied by the fact that when any type $u < 1/2$ sends a message expressing her preference for coordination on $L$, the players coordinate on $L$ with probability one if the opponent’s preferred action is $L$, and they coordinate on $L$ with a probability of $\alpha$ if the opponent’s preferred action is $R$.

Examples of equilibria satisfying all properties  The following strategies, denoted by $\sigma_L$, $\sigma_R$, and $\sigma_C$, are prime examples (that play a special role in later sections) of strategies that are all mutual-preference consistent, coordinated, and have binary communication.

The strategies $\sigma_L$ and $\sigma_R$ are given by the pairs $(\mu^*, \xi_L)$ and $(\mu^*, \xi_R)$, respectively. The message function $\mu^*$ has the property that there are messages $m_L, m_R \in M$ such that message $m_L$ indicates a preference for $L$ and $m_R$ a preference for $R$, and the action functions $\xi_L$ and $\xi_R$ are defined as follows:

$$\mu^*(u) = \begin{cases} m_L & u \leq 1/2 \\ m_R & u > 1/2. \end{cases} \quad \xi_L(m, m') = \begin{cases} R & m = m' = m_R \\ L & \text{otherwise,} \end{cases} \quad \xi_R(m, m') = \begin{cases} L & m = m' = m_L \\ R & \text{otherwise.} \end{cases}$$

This means that the “fallback norm” of $\sigma_L$ (which is applied when the agents have different preferred outcomes) is to coordinate on $L$, while that of $\sigma_R$ is to coordinate on $R$. In other words the left tendency of $\sigma_L$ is one and the left tendency of $\sigma_R$ is zero.

Strategy $\sigma_C = (\mu_C, \xi_C)$ has the “fallback norm” of using a joint lottery to choose the
coordinated outcome. Each agent simultaneously sends a random bit and the coordinated outcome depends on whether the random bits are equal or not.

We denote four distinct messages by \( m_{L,0}, m_{L,1}, m_{R,0}, m_{R,1} \in M \), where we interpret the first subscript (R or L) as the agent’s preferred direction, and the second subscript (0 or 1) as a random binary number chosen with probability \( \frac{1}{2} \) each by the agent. Formally, the message function \( \mu_C \) is defined as follows:

\[
\mu_C(u) = \begin{cases} 
\frac{1}{2} m_{L,0} \oplus \frac{1}{2} m_{L,1} & u \leq \frac{1}{2} \\
\frac{1}{2} m_{R,0} \oplus \frac{1}{2} m_{R,1} & u > \frac{1}{2},
\end{cases}
\]

where \( \alpha m \oplus (1 - \alpha)m' \) is a lottery with a probability of \( \alpha \) on message \( m \) and \( 1 - \alpha \) on message \( m' \). In the second stage, if both agents share the same preferred outcome they play it. Otherwise, they coordinate on \( L \) if their random numbers differ, and coordinate on \( R \) otherwise. Formally:

\[
\xi_C(m, m') = \begin{cases} 
R & (m, m') \in \{(m_{R,0}, m_{R,0}), (m_{R,0}, m_{R,1}), (m_{R,0}, m_{L,0}), (m_{R,1}, m_{L,1}) \\
& (m_{R,1}, m_{R,1}), (m_{R,1}, m_{R,0}), (m_{L,0}, m_{R,0}), (m_{L,0}, m_{R,1})
\}

L & \text{otherwise}.
\]

The outcome of \( \sigma_C \) can also be implemented by a fair joint lottery that determines which of the two players determines the coordinated action used by both players. This alternative implementation yields exactly the same outcome: if both agents share the same preferred outcome they play it, and conditional on the agents disagreeing on the preferred outcome, they coordinate on each action with equal probability.

**One-dimensional set of strategies satisfying the properties**  The set of strategies with the above three properties (coordination, mutual-preference consistency, and binary communication) is essentially one-dimensional because the left tendency \( \alpha^\sigma \in [0, 1] \) of such a strategy \( \sigma \) describes all payoff-relevant aspects. Two strategies with the same left tendency can only differ in the way in which the players implement the joint lottery when they have different preferred outcomes, but these implementation differences are nonessential, as the probability of the joint lottery inducing the players to coordinate on \( L \) remains the same.

Any left-tendency \( \alpha^\sigma \in [0, 1] \cap \mathbb{Q} \) can be implemented by a jointly controlled lottery in which the players send random messages in such a way that they are indifferent between all
messages, and the joint distribution of messages induces $\alpha^\sigma$ (Aumann and Maschler, 1968). This is demonstrated for $\alpha^\sigma = 1/2$ in the strategy $\sigma^C$ presented above.

Note that of all the strategies that satisfy the three properties, strategies $\sigma_L$ and $\sigma_R$ are the simplest in terms of the number of “bits” needed to implement the message function. Strategy $\sigma_C$ is in a certain sense the fairest: conditional on a coordination conflict, i.e., conditional on one agent having a type between 0 and $1/2$ and the other agent having a type between $1/2$ and 1, both agents expect the same payoff. By contrast, strategy $\sigma_L$ favors types below $1/2$, and strategy $\sigma_R$ favors types above $1/2$.

Examples of equilibria not satisfying some of the properties The coordination game with communication $\langle \Gamma, M \rangle$ admits many more equilibria that satisfy only some or even none of the three properties defined above.

First, the game admits babbling equilibria, which do not satisfy mutual-preference consistency. Each babbling equilibrium can be identified with an $x \in [0, 1]$ that satisfies $F(x) = x$, where agents choose $L$ if and only if their type is below $x$. The case of $x = 1$ (resp., $x = 0$) corresponds to a uniform norm of always playing $L$ (resp., $R$). A case of $x \in (0, 1)$ corresponds to an inefficient babbling equilibria, in which agents sometimes miscoordinate.

The game also admits equilibria in which agents reveal some information about the intensity of their preferences (i.e., some information beyond only stating whether $u \leq 1/2$ or $u > 1/2$). One simple example of such an equilibrium is Example 1 in Section 6.

Illustration of equilibria and the first-best outcome Figure 1 illustrates five of the equilibria described above: the equilibria that satisfy the three key properties: $\sigma_L$, $\sigma_R$, and $\sigma_C$, the babbling equilibrium of always playing $R$, and the equilibrium $\sigma_{ex}$, which satisfies none of the three key properties. It also depicts (in the bottom right panel) the first-best outcome in which the players reveal their types and then coordinate on the action that maximizes the sum of payoffs (i.e., the players coordinate on $L$ if $u + v \leq 1$ and they coordinate on $R$ if $u + v > 1$). This is not an equilibrium: each player has an incentive to present a more extreme type than her real type (e.g., all types $u > 1/2$ would claim to have type 1).

4 Definition of Communication-Proofness

For any given strategy in $\Sigma$ employed by both players in the game $\langle \Gamma, M \rangle$, communication and knowledge of this strategy lead to updated and possibly, different and asymmetric infor-
Figure 1: **Six example strategies.** The axis represent the two players’ types $u$ and $v$. Letters $L$, $R$, $C$, and $M$ represent coordination on $L$, coordination on $R$, coordination on a random action, and miscoordination (both players playing their preferred action), respectively.

Information about the two agents’ types. Suppose that the updated distributions of types are given by some distribution functions $G$ and $H$. The two agents then face a (possibly asymmetric) game of coordination without communication, which we shall denote by $\Gamma(G,H)$. Note that the original game (without communication) $\Gamma$ is then given by $\Gamma(F,F)$.

Let $f_m$ be the type density conditional on the agent following a given strategy in the game $\langle \Gamma,M \rangle$ and sending a message $m \in \text{supp}(\bar{\mu})$, i.e., $f_m(u) = f(u)\mu_u(m)/\bar{\mu}(m)$, and let $F_m$ be the cumulative distribution function associated with density $f_m$.

We allow players to renegotiate after communication. Renegotiating players can use a new round of communication. Given a strategy of the game $\langle \Gamma,M \rangle$ employed by both players, we denote the induced renegotiation game after a positive probability message pair $m,m' \in M$ by $\langle \Gamma(F_m,F_m'),M \rangle$.

Let $\pi^H_u(\sigma,\sigma')$ denote the expected payoff for the player using strategy $\sigma$ of type $u$ given

---

9The density $f_m$ depends on the given strategy in the game $\langle \Gamma,M \rangle$. For aesthetic reasons we refrain from giving this strategy a name and from indicating this obvious dependence in our notation.
strategy profile \((\sigma, \sigma')\) in game \(\langle \Gamma(G, H), M \rangle\):

\[
\pi_u^H(\sigma, \sigma') = \mathbb{E}_{v \sim H} \left[ \pi_{u,v}(\sigma, \sigma') \right] = \int_{v=0}^{1} \pi_{u,v}(\sigma, \sigma') h(v) dv,
\]

and similarly let \(\pi_v^G(\sigma, \sigma')\) denote the expected payoff for the player using strategy \(\sigma'\) of type \(v\) in \(\langle \Gamma(G, H), M \rangle\):

\[
\pi_v^G(\sigma, \sigma') = \mathbb{E}_{u \sim G} \left[ \pi_{u,v}(\sigma, \sigma') \right] = \int_{u=0}^{1} \pi_{u,v}(\sigma, \sigma') g(u) du.
\]

Let \(\mathcal{E}(G, H)\) be the set of all (possibly asymmetric) equilibria of the coordination game with communication \(\langle \Gamma(G, H), M \rangle\). Let \(\pi_{u,m',H}(\sigma, \sigma')\) (resp., \(\pi_{v,m',G}(\sigma, \sigma')\)) denote the post-communication payoff for the player using strategy \(\sigma\) (resp., \(\sigma'\)) of a type \(u\) (resp., \(v\)) given \((\sigma = (\mu, \xi), \sigma' = (\mu', \xi'))\) in game \(\langle \Gamma(G, H), M \rangle\) conditional on message pair \(m \in \text{supp}(\sigma), m' \in \text{supp}(\sigma')\):

\[
\pi_{u,m',H}(\sigma, \sigma') = \begin{cases} 
(1-u)H_m(\xi'(m', m)) & \text{if } u \leq \xi(m, m') \\
u(1-H_m(\xi'(m', m))) & \text{if } u > \xi(m, m')
\end{cases}
\]

\[
\pi_{v,m',G}(\sigma, \sigma') = \begin{cases} 
(1-v)G_m(\xi(m, m')) & \text{if } v \leq \xi'(m', m) \\
v(1-G_m(\xi(m, m'))) & \text{if } v > \xi'(m', m).
\end{cases}
\]

Following Blume and Sobel (1995), we say that a strategy profile \((\tau, \tau')\) CP trumps another profile \((\sigma, \sigma')\) if there is a possible pair of messages such that, given the information induced by the message pair, the profile \((\tau, \tau')\), using another round of communication, yields a Pareto-improvement over the post-communication expected payoffs induced by \((\sigma, \sigma')\).

**Definition 1.** Strategy profile \((\tau, \tau') \in \Sigma^2\) CP trumps strategy profile \((\sigma, \sigma') \in \Sigma^2\) with respect to distribution profile \((G, H)\) and message profile \(m \in \text{supp}(\sigma), m' \in \text{supp}(\sigma')\) if\(^{10}\)

1. \((\tau, \tau') \in \mathcal{E}(G_m, H_{m'})\), and

2. \(\pi_u^{H_{m'}(\tau, \tau')} \geq \pi_{u,m',H}(\sigma, \sigma')\) and \(\pi_v^{G_m(\tau, \tau')} \geq \pi_{v,m',G}(\sigma, \sigma')\), for all \(u \in \text{supp}(G_m)\) and all \(v \in \text{supp}(H_{m'})\) with strict inequality for some \(u \in \text{supp}(G_m)\) or some \(v \in \text{supp}(H_{m'})\).

\(^{10}\)For conceptual consistency we could additionally require that a CP-trumping strategy profile be symmetric after a pair of identical messages. We refrain from imposing this, as it would make the notation cumbersome and would not change the set of (strongly or weakly) communication-proof strategies in our setting.
We say that a strategy $\sigma$ is strongly communication-proof if for any possible message profile, there does not exist a new equilibrium, which might require another round of communication, that Pareto-dominates the post-communication payoff of $\sigma$. The weaker notion of weak communication-proofness allows such a Pareto-improving equilibrium to exist as long as this latter equilibrium is not stable in the sense that it is CP trumped by another equilibrium. Formally:

**Definition 2.** An equilibrium strategy $\sigma \in \mathcal{E}$ is strongly communication-proof if it is not CP trumped with respect to $(F,F)$ by any strategy profile.

**Definition 3.** An equilibrium strategy $\sigma \in \mathcal{E}$ is weakly communication-proof if for any strategy profile $(\tau, \tau')$ that CP trumps $(\sigma, \sigma)$ with respect to $(F,F)$ and message profile $(m,m')$, there exists a strategy profile $(\rho, \rho')$ that CP trumps $(\tau, \tau')$ with respect to $(F_m,F_m')$.

Observe that in games with complete information, our two notions coincide, and they are both equivalent to the Pareto frontier of the set of Nash equilibria, i.e., to the subset of Nash equilibria that are not Pareto-dominated by other Nash equilibria. Blume and Sobel’s (1995) notion of communication-proofness lies in between our two notions. Blume and Sobel’s notion is defined in the spirit of von Neumann and Morgenstern’s (1944) set stability: the set of equilibria is divided into stable and unstable equilibria; a strategy profile is communication-proof à la Blume and Sobel if it is not CP trumped by a stable equilibrium; and the set of stable equilibria is defined consistently (any stable equilibrium is only CP trumped by unstable equilibria, and any unstable equilibrium is CP trumped by some stable equilibrium). Blume and Sobel show that any sender-receiver game (in which only one player has private information and her set of actions is a singleton) admits a communication-proof equilibrium.

### 4.1 Evolutionary Interpretation of Communication-Proofness

The notion of robustness to a secret handshake (Robson, 1990) has been applied to games with complete information (see, e.g., Matsui, 1991; Wärneryd, 1991; Kim and Sobel, 1995; Santos et al., 2011). The notion relies on the argument that if play is stuck in an inferior equilibrium $\sigma$, a small group of experimenting agents can recognize each other by means of a “secret handshake” and play a Pareto-improving equilibrium $\sigma'$ with each other and play the inferior equilibrium $\sigma$ with agents who are not part of this group, thereby outperforming the agents outside the group.

To the best of our knowledge the notion has not been applied to games with private types. Arguably, there are two main ways to adapt robustness to secret handshakes to a setup with
private types: ex-ante adaptation and interim adaptation. Ex-ante adaptation assumes that if there exists an alternative equilibrium $\sigma'$ with a higher ex-ante payoff than the current equilibrium $\sigma$, then agents would use a secret handshake to play $\sigma'$ among themselves. We think that ex-ante adaptation is problematic in a setup, which is common in applications, in which agents can only use the secret handshake after they know their type. It seems unlikely that a type would agree to use a secret handshake that decreases her payoff.

In such setups, it seems reasonable to use an interim adaptation, which allows only secret handshakes that benefit all types. This is exactly what is captured by the definition of strong communication-proofness. One could adapt this interim notion of secret handshake, by only allowing “stable” secret handshake that are robust to additional deviations (which is captured by the notion of weak communication-proofness), or allowing agents to use secret handshake also before communicating (See Remark 1). The results of Sections 5-6 show that all these variants of interim robustness to secret handshake lead to the same characterization of robust strategies (namely, to our characterization of coalition-proof equilibria.)

5 Main Result

Our main result shows that both of our notions of communication-proofness coincide in our setup, and they are characterized by satisfying the three key properties of Section 3.

**Theorem 1.** Let $\sigma$ be a strategy of the game with communication $\langle \Gamma, M \rangle$. The following three statements are equivalent:

1. $\sigma$ is mutual-preference consistent, coordinated, and has binary communication.

2. $\sigma$ is a strongly communication-proof equilibrium strategy.

3. $\sigma$ is a weakly communication-proof equilibrium strategy.

**Sketch of proof.** The proof that “1” implies “2” is fairly straightforward (and is proven in Appendix A.2). The proof implies, in particular, that $\sigma_L$, $\sigma_R$, and $\sigma_C$ are not CP trumped by any other strategy profiles. It is immediate that “2” implies “3.” We here provide a sketch of the proof that “3” implies “1.” The proof in Appendix A.2 is split into three lemmas, each showing that one of the three properties must hold.

Lemma 2 proves that a weakly communication-proof equilibrium strategy must be coordinated: if play after any message pair is not coordinated then it is CP trumped in the
renegotiation game by either $\sigma_L$, $\sigma_R$, or $\sigma_C$. To see this, suppose first that both players use thresholds below $1/2$. Then this strategy is Pareto-dominated by $\sigma_R$ as types above $1/2$ gain because $\sigma_R$ induces their first-best outcome, and types below $1/2$ gain because $\sigma_R$ yields a higher coordination probability and a higher probability of the opponent playing this type’s preferred action $L$. Analogously, an equilibrium in which both players use thresholds above $1/2$ is Pareto-dominated by $\sigma_L$. Suppose, finally, that player one uses threshold $x < 1/2$, while player two uses threshold $x' > 1/2$. Observe that $x < 1/2$ (resp., $x' > 1/2$) can be an equilibrium threshold only if player two (resp., player one) plays $L$ with an average probability of less (resp., more) than $1/2$. This, implies that players in these equilibria coordinate with a probability of at most $1/2$, and one can show that such a low coordination probability implies that these equilibria are Pareto-dominated by $\sigma_C$.

Next, we show in Lemma 3 that a weakly communication-proof equilibrium strategy must have binary communication. The reason for this is that if a strategy is coordinated, then different messages can only lead to different ex-ante probabilities of coordination on $L$ (and $R$). Thus, any type who favors $L$, i.e., any type $u < 1/2$, will choose a message to maximize this probability, while any type $u > 1/2$ will choose a message to minimize this probability. Thus, essentially only two kinds of messages are used in a coordinated equilibrium strategy.

Finally, we show in Lemma 4 that a weakly communication-proof equilibrium strategy must be mutual-preference consistent. Given that it is coordinated, we know that any message pair will lead to either coordination on $L$ or on $R$. If it is not mutual-preference consistent then, without loss of generality, there are two types $u, u' < 1/2$ that, with positive probability, send a message pair $(m, m')$ that leads them to coordinate on $R$. But then all types who send this message pair would be weakly better off (and some strictly better off) if instead of coordinating on $R$ they use strategy $\sigma_R$, which would allow them to coordinate on $L$ if and only if both types are below $1/2$.

As the two notions of communication-proofness coincide in our setup, we henceforth omit the word “weakly”/“strongly” and write communication-proof equilibrium strategy to describe either of our (equivalent) solution concepts. Note that the set of communication-proof equilibria is completely independent of the distribution $F$ (i.e., for any two distributions of types $F$ and $F'$, strategy $\sigma$ is a communication-proof equilibrium in $\Gamma(F)$ if and only if it is communication-proof $\Gamma(F')$.) It is not difficult to show that this implies that any communication-proof equilibrium strategy remains communication-proof even in setups in which the distributions of types are correlated, and in setups in which different types have different beliefs about the opponent’s type.

16
\section*{6 On Efficiency}

In this section we investigate the efficiency properties of communication-proof equilibria. We first provide an example of an equilibrium with high ex-ante payoffs that is not, however, communication-proof. We then show that all communication-proof equilibria, while not necessarily ex-ante payoff optimal, are at least interim Pareto efficient. Finally, we show that at least one of the equilibria, $\sigma_L$ and $\sigma_R$, provides the highest ex-ante payoff of all the coordinated equilibria, and that any equilibrium without communication is Pareto-dominated by either one of these extreme communication-proof equilibria or by the action-symmetric communication-proof equilibrium $\sigma_C$.

**High payoff of non-coordinated equilibria** Equilibria with miscoordination (which cannot be communication-proof due to Theorem 1) may induce agents to credibly reveal some cardinal information about their type. This can happen if there is a message that induces a higher probability of coordinating on the agent’s preferred outcome but also a higher probability of miscoordination compared with some other available message. Such a message can then be chosen by extreme types with $u$ far from $1/2$, while moderate types with $u$ closer to $1/2$ choose the other message. Such equilibria with miscoordination may induce a higher ex-ante payoff, if the benefit from signaling the extremeness of the type outweighs the loss due to miscoordination. Consider the following example.

**Example 1.** For simplicity we let the distribution of types $F$ be discrete with four atoms $1/10 + \varepsilon$, $1/2 - \varepsilon$, $1/2 + \varepsilon$, $9/10 - \varepsilon$, with a probability of $1/4$ for each atom and $\varepsilon > 0$ sufficiently small.\footnote{One can easily adapt the example to an atomless distribution of types, in which each atom is replaced with a continuum of nearby types.} The game admits three babbling equilibria: always coordinating on $L$, always coordinating on $R$, both with an ex-ante payoff of $1/2$, and playing $L$ if and only if the type is less than $1/2$ with an ex-ante payoff of $7/20 < 1/2$ for all $\varepsilon$ sufficiently small. Theorem 1 (together with the symmetry of the distribution $F$) implies that with communication, any communication-proof equilibrium strategy (in particular $\sigma_L$ or $\sigma_R$) induces the same expected ex-ante payoff of $3/5 > 1/2$ for all $\varepsilon$ sufficiently small.

This game also has a (non-communication-proof) equilibrium strategy with miscoordination that yields a higher ex-ante payoff than the communication-proof payoff of $3/5$, provided that the message set $M$ has sufficiently many elements. To simplify the presentation we here allow the players to use public correlation devices to determine their joint play after sending
messages, which can be approximately implemented by a sufficiently large message set (à la Aumann and Maschler, 1968). Let $m_L, m_l, m_r, m_R \in M$ and consider strategy $\sigma = (\mu, \xi)$ as follows. Let $\mu^{(1/10 + \varepsilon)} = m_L, \mu^{(1/2 - \varepsilon)} = m_l, \mu^{(1/2 + \varepsilon)} = m_r,$ and $\mu^{(9/10 - \varepsilon)} = m_R,$ and let $\xi(m_a, m_b) = L$ if $a, b \in \{L, l\}, \xi(m_a, m_b) = R$ if $a, b \in \{r, R\}, \xi(m_L, m_r) = \xi(m_l, m_l) = L, \xi(m_l, m_R) = \xi(m_R, m_l) = R, \xi(m_l, m_r) = \xi(m_r, m_l)$ be a joint lottery to coordinate on $L$ or $R$ with probability $1/2$ each, and, finally, let $\xi(m_L, m_R) = \xi(m_l, m_R)$ be a joint lottery to coordinate on $L$ or $R$ with probability $3/10$ each, and to play the inefficient mixed equilibrium (in which each type plays her preferred outcome with probability $9/10 - \varepsilon$) with probability $4/10$. It is straightforward to verify that for, say $\varepsilon = 1/100$, this strategy is indeed an equilibrium strategy with an ex-ante payoff of around 0.627, which is higher than the ex-ante payoff of $3/5$ of all the communication-proof equilibria. This equilibrium strategy is not coordinated (nor does it satisfy the other two properties of mutual-preference consistency and binary communication) and hence, by Theorem 1, it is not communication-proof.

**Interim (pre-communication) Pareto optimality** An (ex-ante symmetric) social choice function is a function $\phi : [0, 1]^2 \rightarrow \Delta \left(\{L, R\}^2\right)$ assigning to each pair of types a possibly correlated profile with the condition that $\phi_{u,v}(a, b) = \phi_{v,u}(b, a)$ for any $a, b \in \{L, R\}$, where\(^ {12}\) $\phi_{u,v} \equiv \phi(u, v)$. We interpret $\phi_{u,v}$ as the correlated action profile played by the two players when a player of type $u$ interacts with a player of type $v$. Let $\Phi$ be the set of all such functions.

Any strategy of any coordination game with communication induces a social choice function in $\Phi$, but not all social choice functions in $\Phi$ can be generated by a strategy of a given coordination game with communication. One can interpret $\Phi$ as the set of outcomes that can be implemented by a designer who perfectly observes the types of both players and, can force the players to play arbitrarily.

For each type $u \in [0, 1]$, let $\pi_u(\phi)$ denote the expected payoff of a player of type $u$ under social choice function $\phi$, i.e., $\pi_u(\phi) = \mathbb{E}_v \left[ (1 - u) \phi_{u,v}(L, L) + u \phi_{u,v}(R, R) \right]$.

A strategy is interim Pareto-dominated if there is a social choice function that is weakly better for all types, and strictly better for some types.

**Definition 4.** A strategy $\sigma \in \Sigma$ is interim Pareto-dominated by function $\phi \in \Phi$ if $\pi_u(\sigma, \sigma) \leq \pi_u(\phi)$ for each type $u \in [0, 1]$, with a strict inequality for a positive measure set of types.

\(^ {12}\) We restrict attention to symmetric social choice functions in order to maintain our focus on symmetric equilibria, and in order to allow us to use a simpler notation without player subscripts. Proposition 1 below, however, also holds even if we allow asymmetric social choice functions.
A strategy $\sigma \in \Sigma$ is interim Pareto optimal if it is not interim Pareto-dominated by any $\phi \in \Phi$. Note that our requirement of Pareto optimality is strong because we allow the designer to perfectly observe the players’ types, and to enforce non-Nash play on the players.

Our next result shows that all communication-proof equilibria satisfy our strong requirement of interim Pareto optimality. That is, even a designer with perfect ability to observe the players’ types and to enforce any behavior cannot achieve a Pareto improvement with respect to any communication-proof equilibrium strategy.\textsuperscript{13}

**Proposition 1.** Every communication-proof equilibrium strategy of a coordination game with communication is interim Pareto optimal.

*Sketch of proof; see Appendix A.3 for the formal proof.* Recall that by Theorem 1 and the discussion on the one-dimensional set of strategies in Section 3, any communication-proof equilibrium strategy $\sigma$ is characterized by its left tendency $\alpha^\sigma$. In order for a social choice function $\phi$ to improve the payoff of any type $u < 1/2$ (resp., $u > 1/2$) relative to the payoff induced by $\sigma$, it must be that $\phi$ induces any $u < 1/2$ (resp., $u > 1/2$) to coordinate on $L$ with probability larger (resp., smaller) than $\alpha^\sigma$. This implies that the probability of two players coordinating on $L$, conditional on the players having different preferred outcomes, must be larger (resp., smaller) than $\alpha^\sigma$. However, these two requirements contradict each other. \hfill $\square$

Earlier we have given an example of an equilibrium strategy that provides a higher ex-ante payoff than any communication-proof equilibrium. This strategy involved a certain degree of miscoordination. In the following proposition we show that any equilibrium without miscoordination, i.e., any coordinated equilibrium, must provide an ex-ante expected payoff that is less than or equal to the maximal ex-ante payoff of the two “extreme” communication-proof strategies $\sigma_L$ and $\sigma_R$.

**Proposition 2.** Let $\sigma \in \mathcal{E}$ be a coordinated equilibrium strategy. Then

$$\pi(\sigma,\sigma) \leq \max \{ \pi(\sigma_L,\sigma_L), \pi(\sigma_R,\sigma_R) \}.$$  

*Sketch of proof; see Appendix A.3 for the formal proof.* Let $\alpha^\sigma$ be the probability of two players who each follow $\sigma$ to coordinate on $L$, conditional on the players having different preferred outcomes, having different preferred outcomes, and in the case of $\sigma_L$, the probability of two players coordinating on $L$, conditional on the players having different preferred outcomes.

\textsuperscript{13}As discussed in the extended working paper version, Heller and Kuzmics (2021), the result that any communication-proof equilibrium is interim Pareto-optimal holds also for asymmetric equilibria. Moreover, two of these asymmetric communication-proof equilibria are also ex-ante Pareto efficient: the equilibrium that always chooses the action preferred by Player 1, and the analogous equilibrium that always chooses the action preferred by Player 2.
ferred outcomes. It is easy to see that \( \sigma \) is dominated by the communication-proof equilibrium strategy with the same left tendency \( \alpha^\sigma \), and that the payoff of the latter strategy is a convex combination of the payoffs of \( \sigma_L \) and \( \sigma_R \), which implies that \( \pi(\sigma, \sigma) \leq \max \{ \pi(\sigma_L, \sigma_L), \pi(\sigma_R, \sigma_R) \} \).

Remark 1. One could refine the notion of communication-proofness to allow agents to renegotiate to a Pareto-improving equilibrium also in earlier stages (à la Benoit and Krishna, 1993): in the interim stage before observing the realized messages induced by the original equilibrium, and in the ex-ante stage before each agent observes her own type. Proposition 1 implies that allowing agents to renegotiate also in the interim stage does not change the set of communication-proof equilibria. Proposition 2 implies that if \( \pi(\sigma_L, \sigma_L) \neq \pi(\sigma_R, \sigma_R) \) then allowing agents to renegotiate also in the ex-ante stage yields a unique “all-stage” communication-proof equilibrium, which is either \( \sigma_L \) or \( \sigma_R \) (while the set of communication-proof equilibria is not affected by introducing ex-ante renegotiation if \( \pi(\sigma_L, \sigma_L) = \pi(\sigma_R, \sigma_R) \)).

Next, we show that \( \sigma_L \) or \( \sigma_R \) provides a strictly higher ex-ante expected payoff than any equilibrium of the game without communication (and therefore than any babbling equilibrium of the game with communication). Recall that in the coordination game without communication any equilibrium is characterized by a cutoff value \( x \in [0, 1] \) such that \( x = F(x) \) with the interpretation that types \( u \leq x \) play \( L \) and types \( u > x \) play \( R \).

Let \( \pi_u(x, x') \) denote the payoff of an agent with type \( u \) who follows a strategy with cutoff \( x \) and faces a partner of unknown type who follows a strategy with cutoff \( x' \):

\[
\pi_u(x, x') = 1_{\{u \leq x\}} F(x') (1 - u) + 1_{\{u > x\}} (1 - F(x')) u,
\]

and let \( \pi(x, x') = \mathbb{E}_u[\pi_u(x, x')] \) be the ex-ante expected payoff of an agent who follows \( x \) and faces a partner who follows \( x' \). Next we show that any (possibly asymmetric) equilibrium in the game without communication is Pareto-dominated by either \( \sigma_L \), \( \sigma_R \), or \( \sigma_C \).

**Corollary 1.** Let \((x, x')\) be a (possibly asymmetric) equilibrium in the coordination game without communication. Then \( \pi_u(x, x') \leq \pi_u(\sigma_L, \sigma_L) \) for all types \( u \in U \), or \( \pi_u(x, x') \leq \pi_u(\sigma_R, \sigma_R) \) for all types \( u \in U \), or \( \pi_u(x, x') \leq \pi_u(\sigma_C, \sigma_C) \) for all types \( u \in U \). Moreover, all the inequalities are strict for almost all types.

Corollary 1 is immediately implied by Lemma 2 in Appendix A.2, and the sketch of proof of the lemma is presented as part of the sketch of the proof of Theorem 1.
Multidimensional Sets of Types

In our model we made the simplifying assumption that miscoordination provides the same payoff (normalized to zero) to both players. This is not completely innocuous. In this section we explore which results are still true in this more general setting. Consider the following multidimensional set of types. Let \( \hat{U} \), a subset of \( \mathbb{R}^4 \), be the set of payoff matrices of binary coordination games, with \( u_{ab} \) being the payoff if a player chooses action \( a \in \{L,R\} \) while her opponent chooses action \( b \in \{L,R\} \):

\[
\hat{U} = \{(u_{LL}, u_{LR}, u_{RL}, u_{RR}) \mid u_{LL} > u_{RL} \text{ and } u_{RR} > u_{LR}\}.
\]

Thus, all types strictly prefer coordination on the same action as the partner to miscoordination. Let \( \hat{\Gamma} = \hat{\Gamma}(G) \) denote the coordination game with the type space \( \hat{U} \), endowed with an atomless CDF \( G \) over \( \hat{U} \) with a density \( g \). Similarly, let \( (\hat{\Gamma}, M) \) be the corresponding game with communication.

Given a type \( u = (u_{LL}, u_{LR}, u_{RL}, u_{RR}) \), let \( \varphi_u \in [0, 1] \) denote type \( u \)'s indifference threshold, which is the probability of the opponent playing \( L \) that induces an agent of type \( u \) to be indifferent: \( \varphi_u = \frac{(u_{RR} - u_{LR})}{(u_{LL} - u_{RL} + u_{RR} - u_{LR})} \).

Observe that an agent with indifference threshold \( \varphi_u \), where \( \varphi_u \) is a number always between 0 and 1, prefers to play \( L \) (\( R \)) if her partner plays \( L \) with probability larger (smaller) than \( \varphi_u \). In other words, for a given probability of her partner playing \( L \), a type \( u \) prefers to play \( L \) if and only if \( \varphi_u \) is less than that probability. Thus, the indifference threshold \( \varphi_u \) replaces what we denoted by \( u \) in the main model. In particular, in this setting we can also restrict attention to cutoff action functions. These are now applied to \( \varphi_u \) instead of to \( u \).

Thus, under a strategy \( \sigma = (\mu, \xi) \) a player plays action \( L \) after observing a message pair \((m, m')\) if and only if \( \varphi_u \leq \xi(m, m') \).

Recall, that action \( L \) is risk-dominant (Harsanyi and Selten, 1988) if it is a best reply against the opponent randomizing equally over the two actions, i.e., if \( \varphi_u \leq 1/2 \Leftrightarrow u_{LL} - u_{LR} \geq u_{RR} - u_{RL} \). The crucial assumption that we implicitly made in our main model is that for any type of player the action that she prefers to coordinate on is also risk-dominant.

**Definition 5.** An atomless distribution over the payoff space \( U \) with density function \( g : U \to \mathbb{R} \) satisfies unambiguous coordination preferences if for any \( u \in U \) with \( g(u) > 0 \) we have \( u_{LL} \geq u_{RR} \Leftrightarrow \varphi_u \leq 1/2 \).

Under a distribution over types with unambiguous coordination preferences, every type
in its support prefers coordinating on action L if and only if that type finds action L risk dominant. Under the assumption that the distribution satisfies unambiguous coordination preferences, Thm. 1 goes through unchanged if we set

\[ F(\varphi) = \int_{\{u \in U : \varphi_u \leq \varphi\}} g(u) du \]

to be the implied distribution over the players’ indifference threshold induced by density \( g \). As in the baseline model, we assume that \( F(\varphi) \) has full support on the interval \([0, 1]\).

**Theorem 2** (Theorem 1 adapted to a multidimensional set of types). Let \( \sigma \) be a strategy of a game \( \langle \hat{\Gamma}, M \rangle \) that satisfies unambiguous coordination preferences. Then the following three statements are equivalent:

1. \( \sigma \) is mutual-preference consistent, coordinated, and has binary communication.
2. \( \sigma \) is a strongly communication-proof equilibrium strategy.
3. \( \sigma \) is a weakly communication-proof equilibrium strategy.

The proof is presented in Appendix A.4. The intuition is the same as in Theorem 1. The adaptation of Lemma 2, to the current setup relies on having unambiguous coordination preferences. Example 2 demonstrates that the restriction of unambiguous coordination preferences is necessary for the “3 \( \Rightarrow \) 1” part of the result.

**Example 2.** There are four possible preference types as follows:

|   | \( u_{L1} \) |   | \( u_{L2} \) |   | \( u_{R1} \) |   | \( u_{R2} \) |
|---|---|---|---|---|---|---|---|
|   | L | R | L | R | L | R | L | R |
| L | 2 | 0 | L | 2 | -15 | L | 1 | 0 |
| R | 0 | 1 | R | 0 | 1 | R | 0 | 2 |

The distribution \( F \) is such that\(^{14} \) \( P(u_{L1}) = P(u_{R1}) = 1/18 \) and \( P(u_{L2}) = P(u_{R2}) = 8/18 \). Let \( M = \{m_L, m_R\} \) and let \( \sigma = (\mu, \xi) \) be such that \( \mu(u_{L1}) = \mu(u_{L2}) = m_L \) and \( \mu(u_{R1}) = \mu(u_{R2}) = m_R \) (making \( \sigma \) mutual-preference consistent), and \( \xi(m_L, m_L) = L, \xi(m_R, m_R) = R, \xi(u_{L1}, m_L, m_R) = L, \xi(u_{L2}, m_L, m_R) = L, \) and \( \xi(u_{R1}, m_R, m_L) = R \) as well as \( \xi(u_{R1}, m_R, m_R) = R \).

It is straightforward to verify that \( \sigma \) is an equilibrium with expected payoffs \( 1/2 \cdot 2 + 1/2 \cdot (8/9 \cdot 2 + 1/9 \cdot 0) = 1 + 8/9 \) for types \( u_{L1} \) and \( u_{R1} \), and \( 1/2 \cdot 2 + 1/2 \cdot 1/9 \cdot 1 = 1 + 1/18 \) for types \( u_{L2} \) and \( u_{R2} \). Observe that no type wants to misreport her preferred outcome in round one. In particular, a misreporting type \( u_{L2} \) will get a payoff of \( 1/2 \cdot 1 + 1/2 \cdot 8/9 \cdot 1 = 17/18 < 1 + 1/9 \).

\(^{14}\)This distribution is discrete, but could be modified to an atomless distribution without changing the result.
**Proposition 3.** *The equilibrium given in Example 2 is not strongly communication proof, but it is weakly communication proof.*

*Sketch of proof; see Appendix A.5 for the formal proof.* After message pair $(m_L, m_R)$ there can be no Pareto-improving equilibrium (with further communication) in which type $L_1$ (also $R_1$) send messages that lead to coordinated play, as these messages could be imitated by type $L_2$ ($R_2$, resp.) and then total payoffs to all types would need to exceed the total available payoff. However, if type $L_1$ ($R_1$) sends a message that would reveal this type (is not sent by the other type) continuation play must be coordinated. Thus, any message type $L_1$ ($R_1$) sends must also be sent by type $L_2$ ($R_2$) and must lead to miscoordination at least for some opponent messages. This miscoordinated equilibrium can then be shown to be CP trumped. \(\square\)

## 8 Discussion

Our notion of communication-proofness adapts Blume and Sobel’s (1995) notion from sender-receiver games to games in which all players have incomplete information, all can communicate, and all can choose actions. Our notion is also related to notions of renegotiation-proofness that have been applied to repeated games (e.g., Farrell and Maskin, 1989; Benoit and Krishna, 1993), and to mechanisms and contracts in the presence of asymmetric information (e.g., Forges, 1994; Neeman and Pavlov, 2013; Maestri, 2017; Strulovici, 2017).

Starting with the secret handshake argument provided in Robson (1990) (see also the earlier related notion of “green beard effect” in Hamilton, 1964; Dawkins, 1976), there is a sizable literature on the evolutionary analysis of costless pre-play communication before players engage in a complete information coordination game. This includes, e.g., Sobel (1993), Blume et al. (1993), Wärneryd (1993), Kim and Sobel (1995), Bhaskar (1998), and Hurkens and Schlag (2003). Suppose that a complete information coordination game has two Pareto-rankable equilibria. Then the Pareto-inferior equilibrium is not evolutionarily stable as it can be invaded by mutants who use a previously unused message as a secret handshake: if their opponent does not use the same handshake they simply play the Pareto-inferior equilibrium (as do all incumbents), but if their opponent also uses the secret handshake both sides play the Pareto-superior equilibrium. Our notion of communication-proofness extends the secret handshake argument to games with incomplete information by requiring that a communication-proof equilibrium should not to be Pareto-dominated by another equilibrium.
after any observed message profile.\textsuperscript{15}

One argument that can be presented against the notion of communication-proofness is that non-communication-proof equilibria can be sustained by the following off-the-equilibrium path behavior: if any player proposes a joint deviation, then the equilibrium specifies that the opponent rejects the offer and that both players shift their behavior to playing an equilibrium that is bad for the proposer. This kind of off-the-equilibrium path proposer punishment would indeed deter players from suggesting joint deviations.\textsuperscript{16}

Recall that we give the notion of communication-proofness two different interpretations: either we think of communication-proof equilibria as the plausible final outcomes of the deliberations of two rational and communicating agents, or we think of these equilibria as the stable outcomes of a long-run learning or evolutionary process.

Under each of these two interpretations one can counter the above proposer-punishing argument. Under the two rational deliberating agents interpretation, one can argue that agents may just have to be careful and subtle in the way they phrase their proposal. Suppose both agents face a situation (after initial messages are sent) in which they are about to play a Pareto-inferior action profile (relative to some possible available equilibrium in the induced game). They should then both realize that their proposer-punishing scheme, which prevents them from renegotiation, is not in their joint best interest and be able to overcome this.

Moreover, under the evolutionary interpretation, there is a more formal counterargument against proposer-punishing schemes. Any off-the-equilibrium path behavior is subject to evolutionary drift; see, e.g., Binmore and Samuelson (1999) for a general treatment of such drift. Eventually the contingent behavior of agents off-the-equilibrium path will start to drift to alternative behavior such as simply ignoring such proposals for joint deviation, or to a willingness to consider them (without applying a punishment). After sufficient drift in this direction, it will be in the agents’ interest to offer Pareto-improving joint deviations. This implies that non-communication-proof equilibria with proposer-punishment mechanisms can only be neutrally stable (which is commonly interpreted as medium-run stability), but they cannot be evolutionarily stable. By contrast, communication-proof equilibria are evolutionarily stable and, as such drift cannot take behavior away from a communication-proof equilibrium.

\textsuperscript{15}Another closely related solution concept is Swinkels’s (1992) notion of robustness to equilibrium entrants. In a recent paper, Newton (2017) provides an evolutionary foundation for players developing the ability to renegotiate into a Pareto-better outcome (“collaboration” in the terminology of Newton).

\textsuperscript{16}These kinds of proposer-punishing mechanisms are explored in solution concepts of renegotiation-proofness that explicitly specify a structured renegotiation protocol, such as Busch and Wen (1995), Santos (2000), and Safronov and Strulovici (2019).
Another related literature deals with stable equilibria in coordination games with private values, but without pre-play communication. Sandholm (2007) (extending earlier results of Fudenberg and Kreps, 1993; Ellison and Fudenberg, 2000) shows that mixed Nash equilibria of the game with complete information can be purified in the sense of Harsanyi (1973) in an evolutionarily stable way. \(^{17}\) Finally, two related papers analyze stag-hunt games with private values. Baliga and Sjöström (2004) show that introducing pre-play communication induces a new equilibrium in which the Pareto-dominant action profile is played with high probability. Jelnov et al. (2018) show that in some cases a small probability of another interaction can substantially affect the set of equilibrium outcomes in stag-hunt games with private values.

Finally, in the extended working paper version of this paper, Heller and Kuzmics (2021), we study the robustness of the main result in various extensions of the model. There we show that communication-proof equilibria also satisfy a variety of standard evolutionary stability notions. We show that nothing changes if we allow for asymmetric play and asymmetric coordination games. Introducing dominant action types refines the set of communication-proof equilibria to essentially a single one, where the left tendency of this strategy is equal to the share of extreme types whose dominant action is L. We there also discuss extensions of our model to multiple players and multiple actions.

A Formal Proofs

A.1 Undominated Action Strategies

In what follows we show that our restriction to threshold action functions is without loss of generality, in the sense that each generalized strategy is dominated by a threshold strategy.

Let \(\Gamma(F, G)\) be a coordination game without communication (possibly played after a pair of messages is observed in the original game \(\langle \Gamma, M \rangle\)). A generalized strategy is a measurable function \(\eta : U \to \Delta(\{L, R\})\) that describes a mixed action as a function of the player’s type. A generalized strategy in \(\Gamma(F, G)\) corresponds to a generalized action function \(\xi : U \times M \times M \to \triangle\{L, R\}\), given a specific pair of observed messages \((m, m')\), i.e., \(\eta(u) \equiv \xi(u, m, m')\).

A pair of generalized strategies \(\eta, \tilde{\eta}\) are almost surely realization equivalent (abbr., equiv-
alent), denoted by $\eta \approx \tilde{\eta}$, if they induce the same behavior with probability one, i.e., if

$$
\mathbb{E}_{u \sim F} [\eta(u) \neq \tilde{\eta}(u)] \equiv \int_{u \in U} f(u) 1_{\{\eta(u) \neq \tilde{\eta}(u)\}} du = 0.
$$

It is immediate that two equivalent generalized strategies always induce the same (ex-ante) payoff, i.e., that $\pi(\eta, \eta') = \pi(\tilde{\eta}, \eta')$ for each generalized strategy $\eta'$.

A generalized strategy is a cutoff strategy if there exists a type $x \in [0, 1]$ such that $\eta(u) = L$ for each $u < x$ and $\eta(u) = R$ for each $u > x$. A generalized strategy $\eta$ is strictly dominated by generalized strategy $\tilde{\eta}$ if $\pi(\eta, \eta') < \pi(\tilde{\eta}, \eta')$ for any opponent’s generalized strategy $\eta'$.

The following result shows that any generalized strategy is either equivalent to a cutoff strategy, or it is strictly dominated by a cutoff strategy.

**Lemma 1.** Let $\eta$ be a generalized strategy. Then there exists a cutoff strategy $\tilde{\eta}$, such that either $\eta$ is equivalent to $\tilde{\eta}$, or $\eta$ is strictly dominated by $\tilde{\eta}$.

**Proof.** If $\mathbb{E}_{u \sim F} [\eta_a(L)] = 1$ (resp., $\mathbb{E}_{u \sim F} [\eta_a(L)] = 0$), then $\eta$ is equivalent to the cutoff strategy of always playing $L$ (resp., $R$). Thus, suppose that $\mathbb{E}_{u \sim F} [\eta_a(L)] \in (0, 1)$. Let $x \in (0, 1)$ be such that $F(x) = \mathbb{E}_{u \sim F} [\eta_a(L)] = \int_{u} \eta_a(L) f(u) du$. Let $\tilde{\eta}$ then be the cutoff strategy with cutoff $x$, i.e., $\tilde{\eta}_a(L) = 1$ if $u \leq x$ and $\tilde{\eta}_a(L) = 0$ if $u > x$. Assume that $\eta$ and $\tilde{\eta}$ are not equivalent, i.e., $\eta \not\approx \tilde{\eta}$. Let $\eta'$ be an arbitrary generalized strategy of the opponent. By construction, strategies $\eta$ and $\tilde{\eta}$ induce the same average probability of choosing $L$. Strategies $\tilde{\eta}$ and $\eta$ differ in that $\tilde{\eta}$ induces lower types to choose $L$ with higher probability, and higher types to choose $L$ with lower probability, i.e., $\eta_a(L) \leq \tilde{\eta}_a(L)$ for any type $u \leq x$ and $\eta_a(L) \geq \tilde{\eta}_a(L)$ for any type $u > x$. Since $\eta \not\approx \tilde{\eta}$ and $\mathbb{E}_{u \sim F} [\eta_a(L)] \in (0, 1)$, it follows that the inequalities are strict for a positive measure of types, i.e.,

$$0 < \int_{u < x} f(u) 1_{\{\eta(u) < \tilde{\eta}(u)\}} du \quad \text{and} \quad 0 < \int_{u > x} f(u) 1_{\{\eta(u) > \tilde{\eta}(u)\}} du.
$$

The fact that lower types always gain more (less) from choosing $L$ ($R$) relative to higher types, with a strict inequality unless the opponent always plays $R$ ($L$), implies that $\pi(\eta, \eta') < \pi(\tilde{\eta}, \eta')$.

**A.2 Proof of Theorem 1**

We first prove the “1 $\Rightarrow$ 2” part. Suppose that $\sigma = (\mu, \xi) \in \Sigma$ is mutual-preference consistent, coordinated, and has binary communication. As $\sigma$ is mutual-preference consistent it must
satisfy \( \text{supp}(F_m) \subseteq [0, 1/2] \) or \( \text{supp}(F_m) \subseteq [1/2, 1] \) for any message \( m \in \text{supp}(\mu) \). Consider any \( m, m' \in \text{supp}(\mu) \). There are three cases to consider. Suppose first that \( \text{supp}(F_m), \text{supp}(F_{m'}) \subseteq [0, 1/2] \). Then as \( \sigma \) is mutual-preference consistent we have that \( \xi(m, m') = \xi(m', m) = L \). Thus \( \xi \) describes best-reply behavior after this message pair. Moreover this behavior is the best possible outcome for any type in \([0, 1/2]\) and thus for any type in \( \text{supp}(F_m) \) and \( \text{supp}(F_{m'}) \).

The second case of \( \text{supp}(F_m), \text{supp}(F_{m'}) \subseteq [1/2, 1] \) is analogous.

Suppose, finally, that, w.l.o.g., \( \text{supp}(F_m) \subseteq [0, 1/2] \) and \( \text{supp}(F_{m'}) \subseteq [1/2, 1] \). As \( \sigma \) is coordinated we have that \( \xi(m, m') = \xi(m' , m) = L \) or \( \xi(m, m') = \xi(m', m) = R \). Action function \( \xi \), therefore, again describes best-reply behavior. Moreover, one player always obtains her most preferred outcome. In order for a new strategy profile to improve the opponent’s outcome, this new profile must require the former player to deviate from her most preferred outcome. Thus, no equilibrium \( \sigma' \) in the game \( \langle \Gamma(F_m, F_{m'}), M \rangle \) Pareto dominates \( \sigma \) after this message pair. This shows that action function \( \xi \) is a best response to \( \mu \) and to itself given \( \mu \) and that, moreover, it cannot be CP trumped. It remains to show that the message function \( \mu \) is optimal when the opponent chooses \( \sigma = (\mu, \xi) \).

Consider type \( u \in [0, 1/2] \) and consider this type’s choice of message. As \( \sigma \) has binary communication and is coordinated, different messages \( m \in M \) can only trigger different probabilities of coordinating on \( L \) with a highest likelihood of such coordination for any message \( m \in \text{supp}(\mu_u) \). Therefore, type \( u \) is indifferent between any message \( m \in \text{supp}(\mu_u) \) and weakly prefers sending any message \( m \in \text{supp}(\mu_u) \) to sending any message \( m' \not\in \text{supp}(\mu_u) \). An analogous statement holds for types \( u \in [1/2, 1] \). This concludes the proof of the “1 \( \Rightarrow 2 \)” part of the theorem.

We prove the “3 \( \Rightarrow 1 \)” part in three lemmas, one for each of the three properties.

**Lemma 2.** Every weakly communication-proof equilibrium strategy \( \sigma = (\mu, \xi) \) is coordinated.

**Proof.** We need to show that for any message pair \( m, m' \in \text{supp}(\bar{\mu}) \),

\[
\text{either } \xi(m, m') \geq \sup \{ u \mid \mu_u(m) > 0 \} \text{ or } \xi(m, m') \leq \inf \{ u \mid \mu_u(m) > 0 \}.
\]

Let \( m, m' \in \text{supp}(\bar{\mu}) \) and assume to the contrary that

\[
\inf \{ u \mid \mu_u(m) > 0 \} < \xi(m, m') < \sup \{ u \mid \mu_u(m) > 0 \}.
\]

As \( \sigma \) is an equilibrium, we have \( \inf \{ u \mid \mu_u(m') > 0 \} < \xi(m', m) < \sup \{ u \mid \mu_u(m') > 0 \} \).
cause otherwise the sender of $m'$ would play $L$ with probability one or $R$ with probability one, in which case the best reply of the sender of message $m$ would be to play $L$ (or $R$) regardless of her type.

Let $x = \xi(m,m')$ and $x' = \xi(m',m)$. We now show that the equilibrium $(x,x')$ of the game without coordination $\Gamma(F_m,F_{m'})$ is CP-trumped by either $\sigma_L$, $\sigma_R$, or $\sigma_C$. There are three cases to be considered. Case 1: Suppose that $x,x' \leq 1/2$. We now show that in this case the equilibrium $(x,x')$ is CP-trumped by $\sigma_R$. Consider the player who sent message $m$.

Case 1a: Consider a type $u \leq x$. Then we have

$$
(1-u)F_{m'}(1/2) + u(1 - F_{m'}(1/2)) \geq (1-u)F_{m'}(x'),
$$

where the left-hand side is type a $u$ agent’s payoff under strategy profile $\sigma_R$ and the right-hand side the payoff under strategy profile $(x,x')$. The inequality follows from the fact that $u(1 - F_{m'}(1/2)) \geq 0$, and $F_{m'}(1/2) \geq F_{m'}(x')$ follows from the fact that $F_{m'}$ is nondecreasing (as it is a cumulative distribution function). This inequality is strict for all $u$ except for $u = 0$ in the case where $x' = 1/2$.

Case 1b: Now consider a type $u$ with $x < u \leq 1/2$. Then we have

$$
(1-u)F_{m'}(1/2) + u(1 - F_{m'}(1/2)) > u(1 - F_{m'}(x')),
$$

where the left-hand side is a type $u$ agent’s payoff under strategy profile $\sigma_R$ and the right-hand side is the payoff under strategy profile $(x,x')$. The inequality follows from the fact that by $u \leq 1/2$ we have that $1-u \geq u$, and therefore $(1-u)F_{m'}(1/2) + u(1 - F_{m'}(1/2)) \geq u$.

Case 1c: Finally, consider a type $u > 1/2$. Then we have $u > u(1 - F_{m'}(x'))$, where the left-hand side is a type $u$ agent’s payoff under strategy profile $\sigma_R$ and the right-hand side is the payoff under strategy profile $(x,x')$.

The analysis for the player who sent message $m'$ is analogous.

Case 2: Suppose that $x,x' \geq 1/2$. The analysis is analogous to Case 1 if we replace $\sigma_R$ with $\sigma_L$.

Case 3: Suppose, w.l.o.g. for the remaining cases, that $x \leq 1/2 \leq x'$. The equilibrium $(x,x')$ in this case is Pareto-dominated by $\sigma_C$. To see this, consider the player who sent message $m$.

Case 3a: Consider a type $u \leq x$. Then we have

$$
(1-u)[F_{m'}(1/2) + 1/2(1 - F_{m'}(1/2))] + u^{1/2}(1 - F_{m'}(1/2)) > (1-u)F_{m'}(x'),
$$

28
where the left-hand side is a type $u$ agent’s payoff under strategy profile $\sigma_C$ and the right-hand side the payoff under strategy profile $(x, x')$. The inequality follows from the fact that we have $F_{m'}(x') = x \leq 1/2$ due to $(x, x')$ being an equilibrium.

Case 3b: Now consider a type $u$ with $x < u \leq 1/2$. Then we have

$$\left(1 - u\right) \left[F_{m'}(1/2) + 1/2 \left(1 - F_{m'}(1/2)\right)\right] + u^{1/2} \left(1 - F_{m'}(1/2)\right) > u \left(1 - F_{m'}(x')\right),$$

where the left-hand side is a type $u$ agent’s payoff under strategy profile $\sigma_C$ and the right-hand side the payoff under strategy profile $(x, x')$. The inequality follows from the fact that by $u \leq 1/2$ we have $1 - u \geq u$ and thus $(1 - u) \left[F_{m'}(1/2) + 1/2 \left(1 - F_{m'}(1/2)\right)\right] + u^{1/2} \left(1 - F_{m'}(1/2)\right) \geq u$.

Case 3c: Finally, consider a type $u > 1/2$. Then we have

$$u \left[\left(1 - F_{m'}(1/2)\right) + 1/2 F_{m'}(1/2)\right] + (1 - u)^{1/2} F_{m'}(1/2) > u \left(1 - F_{m'}(x')\right),$$

where the left-hand side is a type $u$ agent’s payoff under strategy profile $\sigma_C$ and the right-hand side is the payoff under strategy profile $(x, x')$. The inequality follows from the fact that we have $F_{m'}(1/2) > 0$ and $F_{m'}(1/2) \leq F_{m'}(x')$.

The analysis for the player who sent message $m'$ is analogous. \qed

**Lemma 3.** Every weakly communication-proof equilibrium strategy $\sigma$ has binary communication.

**Proof.** Let $\sigma$ be a weakly communication-proof equilibrium strategy. Recall that

$$\beta^\sigma(m) = \int_{u=0}^{1} \sum_{m' \in M} \sigma_u(m') \mathbb{1}_{\{u \leq \xi_{(m,m')}\}} f(u) du.$$

As $\sigma$ is coordinated by Lemma 2, the payoff to a type $u$ from sending $m \in \text{supp}(\mu)$ is

$$(1 - u) \beta^\sigma(m) + u \left(1 - \beta^\sigma(m)\right).$$

For a type $u < 1/2$ the problem of choosing a message to maximize her payoffs is thus equivalent to choosing a message that maximizes $\beta^\sigma(m)$. We thus must have that there is a $\beta^\sigma \in [0, 1]$ such that for all $u < 1/2$ and all $m \in \text{supp}(\mu_u)$, we have $\beta^\sigma(m) = \beta^\sigma$. Analogously, we must have a $\beta^\sigma \in [0, 1]$ such that for all $u > 1/2$ and all $m \in \text{supp}(\mu_u)$, we have $\beta^\sigma = \beta^\sigma$. Clearly also $\beta^\sigma \leq \beta^\sigma$. To extend the argument to unused messages $m \not\in \text{supp}(\mu)$ we rely on the full support assumption. Assume to the contrary that there is a
message $m \not\in \text{supp}(\bar{\mu})$ with $\beta^\sigma(m) > \overline{\beta}^\sigma$ (resp., $\beta^\sigma(m) < \overline{\beta}^\sigma$). Then any sufficiently high (resp., low) type $u$ would strictly earn by deviating to sending message $m$ and playing $L$ (resp., $R$), which contradicts the supposition that $\sigma$ is an equilibrium strategy. \qed

**Lemma 4.** Every weakly communication-proof equilibrium strategy $\sigma$ is mutual-preference consistent.

**Proof.** By Lemma 2 a weakly communication-proof equilibrium strategy $\sigma = (\mu, \xi)$ is coordinated. Suppose that it is not mutual-preference consistent. Then there is either a pair $(m, m')$ such that there are types $u, v < 1/2$ with $m \in \text{supp}(\mu_u)$ and $m' \in \text{supp}(\mu_v)$ such that play after $(m, m')$ is coordinated on $R$, or a pair $(m, m')$ such that there are types $u, v > 1/2$ with $m \in \text{supp}(\mu_u)$ and $m' \in \text{supp}(\mu_v)$ such that play after $(m, m')$ is coordinated on $L$. In the former (resp., latter) case strategy $\sigma$ is CP-trumped by $\sigma_R$ (resp., $\sigma_L$) in the game $(\Gamma(F_m, F_{m'}), \{m_L, m_R\})$ because $\sigma_R$ (resp., $\sigma_L$) does not affect the payoff of all types $u \geq 1/2$ (resp., $u \leq 1/2$), and it strictly improves the payoff to all types $u < 1/2$ (resp., $u > 1/2$). \qed

### A.3 Proofs of Section 6 (On Efficiency)

**Proof of Proposition 1.** By Theorem 1 and the discussion of the one-dimensional set of strategies satisfying the key properties in Section 3 a communication-proof equilibrium strategy $\sigma$’s payoff is determined by its left tendency $\alpha \equiv \alpha^\sigma \in [0, 1]$. This payoff is given by

$$\pi_u(\sigma, \sigma) = (1-u)[F(1/2) + \alpha (1-F(1/2))] + u(1-\alpha)[1-F(1/2)],$$

for each $u \in (0, 1/2]$, and it is given by

$$\pi_u(\sigma, \sigma) = (1-u)\alpha F(1/2) + u[(1-F(1/2)) + F(1/2)(1-\alpha)].$$

for each type $u \in (1/2, 1]$. The payoff to a type $u$ from given social choice function $\phi$ is given by

$$\pi_u(\phi) = (1-u)\mathbb{E}_v \phi_{u,v}(L, L) + u\mathbb{E}_v \phi_{u,v}(R, R).$$

Now suppose that $\phi$ interim Pareto dominates $\sigma$. Then $\pi_u(\phi) \geq \pi_u(\sigma, \sigma)$ for all $u \in [0, 1]$ with a strict inequality for a positive measure of $u$. As $\pi_u(\sigma, \sigma)$ is a convex combination of two payoffs, this implies that:

$$\mathbb{E}_v \phi_{u,v}(L, L) \geq F(1/2) + \alpha (1-F(1/2)) \text{ for any } u \leq 1/2,$$

and

$$\mathbb{E}_v \phi_{u,v}(R, R) \geq 1 - F(1/2) - \alpha (1-F(1/2)) \text{ for any } u \geq 1/2.$$
with at least one of the inequalities holding strictly for a positive measure of types. Thus,

\[ \mathbb{E}_v \phi_{u,v}(R,R) \geq (1 - F(1/2)) + F(1/2)(1 - \alpha) \]  

for any \( u > 1/2 \), 

(2)

where, for instance, \( \mathbb{E}_{\{v \geq 1/2\}} \) denotes the expectation conditional on \( v > 1/2 \). Substituting this last equality in Eq. (1) yields the following inequality

\[ F(1/2) \mathbb{E}_{\{v \leq 1/2\}} \phi_{u,v}(L,L) + (1 - F(1/2)) \mathbb{E}_{\{v > 1/2\}} \phi_{u,v}(L,L) \geq F(1/2) + \alpha (1 - F(1/2)) \]

for any \( u \leq 1/2 \). The fact that \( \mathbb{E}_{\{v \leq 1/2\}} \phi_{u,v}(L,L) \leq 1 \) implies that \( \mathbb{E}_{\{v > 1/2\}} \phi_{u,v}(L,L) \geq \alpha \) for any \( u \leq 1/2 \). An analogous argument (applied to Equation (2)) implies that \( \mathbb{E}_{\{v < 1/2\}} \phi_{u,v}(R,R) \geq 1 - \alpha \), for any \( u > 1/2 \), with at least one of these inequalities holding strictly for a positive measure of types. This implies that

\[ \mathbb{E}_{\{u \leq 1/2\}} \mathbb{E}_{\{v > 1/2\}} \phi_{u,v}(L,L) \geq \alpha \]  

and

\[ \mathbb{E}_{\{u > 1/2\}} \mathbb{E}_{\{v < 1/2\}} \phi_{u,v}(R,R) \geq 1 - \alpha, \]

with at least one of the two inequalities holding strictly. By the symmetry of \( \phi \) we have \( \phi_{u,v}(R,R) = \phi_{v,u}(R,R) \) and thus

\[ \mathbb{E}_{\{u \leq 1/2\}} \mathbb{E}_{\{v > 1/2\}} \phi_{u,v}(L,L) + \mathbb{E}_{\{u > 1/2\}} \mathbb{E}_{\{v < 1/2\}} \phi_{u,v}(R,R) \geq 1, \]

which contradicts \( \phi_{u,v} \) being a social choice function. \( \square \)

The proof of Proposition 2 uses the following lemma (which is of independent interest).

Lemma 5. Let \( \sigma \in \mathcal{E} \) be a coordinated equilibrium strategy. Then there is a communication-proof strategy \( \sigma' \) such that either \( \sigma \) and \( \sigma' \) are interim payoff equivalent or \( \sigma' \) interim Pareto dominates \( \sigma \).

Proof. Let \( \sigma = (\mu, \xi) \in \mathcal{E} \) be coordinated. For each message \( m \in M \), let \( p_m \in [0,1] \) be the probability that the players coordinate on \( L \), conditional on the agent sending message \( m \):

\[ p_m = \sum_{m' \in M} \mu(m') \mathbb{1}_{\xi(m,m') = L}. \]

As \( \sigma \) is coordinated, it follows that \( 1 - p_m \) is the probability that the players coordinate on \( R \), conditional on the agent sending message \( m \).
Let \( \bar{p} = \max_{m \in M} p_m \) be the maximal probability, and let \( \underline{p} = \min_{m \in M} p_m \) be the minimal probability. By definition, \( \underline{p} \leq \bar{p} \). As \( \sigma \) is an equilibrium strategy, \( \underline{p} < \bar{p} \) implies that all types \( u < 1/2 \) send a message inducing probability \( \bar{p} \) and all types \( u > 1/2 \) send a message inducing probability \( \underline{p} \). Therefore, the expected payoff of a type \( u \leq 1/2 \) is given by \( \pi_u (\sigma, \sigma) = \bar{p} (1-u) + (1-\bar{p}) u \), and the expected payoff of any type \( u > 1/2 \) is equal to \( \pi_u (\sigma, \sigma) = \underline{p} (1-u) + (1-\underline{p}) u \). This is also true if \( \underline{p} = \bar{p} \). Note that for types \( u < 1/2 \), the expected payoff strictly increases in \( \bar{p} \) and for types \( u > 1/2 \) the type’s expected payoff strictly decreases in \( \underline{p} \).

We consider three cases. Suppose first that \( \underline{p} \leq \bar{p} \leq F(1/2) \). Then let \( \sigma' = \sigma_R \). This strategy is also coordinated and its induced payoffs can be written in the same form as those for strategy \( \sigma \) with \( \underline{p}' = 0 \) and \( \bar{p}' = F(1/2) \). Thus, we get that \( \pi_u (\sigma', \sigma') \geq \pi_u (\sigma, \sigma) \) for every \( u \in [0,1] \). This implies that \( \sigma \) is either interim (pre-communication) payoff equivalent to or Pareto-dominated by \( \sigma' = \sigma_R \).

The second case where \( F(1/2) \leq \underline{p} \leq \bar{p} \) is analogous to the first one, with \( \sigma' = \sigma_L \). In the final case \( \underline{p} < F(1/2) < \bar{p} \). Let \( \alpha \in [0,1] \) be such that \( F(1/2) + (1 - F(1/2)) \alpha = \bar{p} \) and let \( \sigma' \) be a communication-proof strategy with left tendency \( \alpha \). Then \( \underline{p} \geq \alpha F(1/2) \) and by construction \( \sigma \) is either interim (pre-communication) payoff equivalent to or Pareto dominated by \( \sigma' \).

**Proof of Proposition 2.** By Lemma 5 we have that every coordinated equilibrium strategy \( \sigma \) is interim (pre-communication) Pareto-dominated by some communication-proof strategy with some left tendency \( \alpha \in [0,1] \) denoted by \( \sigma_\alpha \). We thus have that \( \pi (\sigma, \sigma) \leq \pi (\sigma_\alpha, \sigma_\alpha) \).

The ex-ante expected payoff of to a \( u \) type under strategy \( \sigma_\alpha \) is given by

\[
\pi_u (\sigma_\alpha, \sigma_\alpha) = (1-u) [F(1/2) + \alpha (1-F(1/2))] + u (1-\alpha) (1-F(1/2)) \quad \text{for } u \leq 1/2
\]

\[
\pi_u (\sigma_\alpha, \sigma_\alpha) = (1-u) \alpha F(1/2) + u [1-F(1/2) + (1-\alpha)F(1/2)] \quad \text{for } u > 1/2.
\]

It is straightforward to verify that \( \pi_u (\sigma_\alpha, \sigma_\alpha) = \alpha \pi_u (\sigma_1, \sigma_1) + (1-\alpha) \pi_u (\sigma_0, \sigma_0) \). for every \( u \).

As \( \sigma_1 = \sigma_L \) and \( \sigma_0 = \sigma_R \) and as for all \( u \in [0,1] \) \( \pi_u (\sigma_\alpha, \sigma_\alpha) \) is the same convex combination of \( \pi_u (\sigma_L, \sigma_L) \) and \( \pi_u (\sigma_R, \sigma_R) \), we have \( \pi (\sigma_\alpha, \sigma_\alpha) = \alpha \pi (\sigma_1, \sigma_1) + (1-\alpha) \pi (\sigma_0, \sigma_0) \), which implies that \( \pi (\sigma, \sigma) \leq \pi (\sigma_\alpha, \sigma_\alpha) \leq \max \{ \pi (\sigma_L, \sigma_L), \pi (\sigma_R, \sigma_R) \} \).
A.4 Proof of Theorem 2

The proof of Theorem 2 mimics the proof of Theorem 1 except that Lemma 2 has to be adapted somewhat as follows (this is the only place where one uses the assumption of unambiguous coordination preferences).

**Lemma 6.** Assume that the atomless distribution of types have unambiguous coordination preferences. Let \( \sigma = (\mu, \xi) \) be a weakly communication-proof equilibrium strategy. Then it is coordinated.

**Proof.** We need to show that for any message pair \( m, m' \in \text{supp}(\bar{\mu}) \),

\[
\text{either } \xi(m, m') \geq \sup \{ \phi_u | \mu_u(m) > 0 \} \text{ or } \xi(m, m') \leq \inf \{ \phi_u | \mu_u(m) > 0 \} .
\]

Let \( m, m' \in \text{supp}(\bar{\mu}) \) and assume to the contrary that

\[
\inf \{ \phi_u | \mu_u(m) > 0 \} < \xi(m, m') < \sup \{ \phi_u | \mu_u(m) > 0 \} .
\]

As \( \sigma \) is an equilibrium, we must have \( \inf \{ \phi_u | \mu_u(m') > 0 \} < \xi(m', m) < \sup \{ \phi_u | \mu_u(m') > 0 \} \). (Otherwise the \( m' \) message sender would play \( L \) with probability one or \( R \) with probability one, in which case the \( m \) message sender’s best response would be to play \( L \) (or \( R \)) regardless of her type). Let \( x = \xi(m, m') \) and \( x' = \xi(m', m) \). In what follows we will show that the equilibrium \( (x, x') \) of the game without communication \( \Gamma(F_m, F_{m'}) \) is Pareto-dominated by either \( \sigma_L \), \( \sigma_R \), or \( \sigma_C \) (all based on \( \phi_u \) instead of \( u \)).

There are three cases to be considered. Case 1: Suppose that \( x, x' \leq 1/2 \). We now show that in this case the equilibrium \( (x, x') \) is Pareto-dominated by \( \sigma_R \). Consider the player who sent message \( m \).

Case 1a: Consider a type \( u \) with \( \phi_u \leq x \). Then we have

\[
u_{LL}F_m(x') + \left(1 - F_m'(x')\right)u_{LR} \leq u_{LL}F_{m'}(1/2) + u_{LR}\left(1 - F_{m'}(1/2)\right) \leq u_{LL}F_{m'}(1/2) + u_{RR}\left(1 - F_{m'}(1/2)\right),
\]

where the first expression is the type \( u \) agent’s payoff under strategy profile \( (x, x') \) and the last expression is her payoff under strategy profile \( \sigma_R \). The first inequality follows from \( u_{LL} \geq u_{LR} \) and \( F_m'(1/2) \geq F_m'(x') \) by the fact that \( F_m' \) is nondecreasing (as it is a CDF), and the second inequality follows from \( u_{RR} \geq u_{LR} \). This inequality is strict when \( u_{LL} > u_{LR} \) and \( F_{m'}(1/2) > F_{m'}(x') \) or when \( u_{RR} > u_{LR} \).
Case 1b: Now consider a type $u$ with $x < \varphi_u \leq 1/2$. Then we have

$$u_{RL}F_m'(x') + u_{RR}(1 - F_m'(x')) \leq u_{LL}F_m'(x') + u_{RR}(1 - F_m'(x')) \leq u_{LL}F_m'(1/2) + u_{RR}(1 - F_m'(1/2)),$$

where the first expression is the type $u$ agent’s payoff under strategy profile $(x, x')$ and the last expression is her payoff under strategy profile $\sigma_R$. The first inequality follows from $u_{LL} \geq u_{RL}$ and the second one from $F_m'(1/2) \geq F_m'(x')$ and $u_{LL} \geq u_{RR}$. Note also that the second inequality follows from the assumption of unambiguous coordination preferences and $\varphi_u \leq 1/2$. This inequality is strict when $u_{LL} > u_{RL}$ or when $F_m'(1/2) > F_m'(x')$ and $u_{LL} > u_{RR}$. The inequality follows from the observation that $u_{RR} > u_{RL}$ because $u_{RR} > u_{LL}$ by the assumption of unambiguous coordination preferences, and $u_{LL} \geq u_{RL}$ by the fact that it is a coordination game.

The analysis for the player who sent message $m'$ is analogous.

Case 2: Suppose that $x, x' \geq 1/2$. The analysis is analogous to Case 1 if we replace $\sigma_R$ with $\sigma_L$.

Case 3: Suppose, without loss of generality for the remaining cases, that $x \leq 1/2 \leq x'$. We show that the equilibrium $(x, x')$ in this case is Pareto-dominated by $\sigma_C$. Consider the player who sent message $m$.

Case 3a: Consider a type $u$ such that $\varphi_u \leq x$. Then we have

$$u_{LL}[F_m'(1/2) + 1/2(1 - F_m'(1/2))] + u_{RR}1/2(1 - F_m'(1/2)) > u_{LL}F_m'(x') + u_{LR}(1 - F_m'(x')),$$

where the right-hand side is the type $u$ agent’s payoff under strategy profile $(x, x')$ and the left-hand side is her payoff under strategy profile $\sigma_C$. The inequality follows from the observation that $u_{RR} \geq u_{LR}$ and $F_m'(x') \leq 1/2$ by the fact that $F_m'(x') = x$ when $(x, x')$ is an equilibrium.

Case 3b: Now consider a type $u$ with $x < \varphi_u \leq 1/2$. Then we have

$$u_{RL}F_m'(x') + u_{RR}(1 - F_m'(x')) \leq u_{LL}F_m'(x') + u_{RR}(1 - F_m'(x')) \leq$$

$$u_{LL}[1/2 + 1/2F_m'(x')] + u_{RR}1/2(1 - F_m'(1/2)),$$

where the first expression is the type $u$ agent’s payoff under strategy profile $(x, x')$ and the last expression is her payoff under strategy profile $\sigma_C$. The first inequality follows from $u_{LL} \geq u_{RL}$ and the second one from $u_{LL} \geq u_{RR}$ by the assumption of unambiguous coordination.
preferences given $\varphi_u \leq \frac{1}{2}$ and $F_m(x') = x$ by $(x,x')$ being an equilibrium and $x < \frac{1}{2}$. The inequality is strict if $u_{LL} > u_{RL}$ or $u_{LL} > u_{RR}$.

Case 3c: Finally, consider a type $u$ with $\varphi_u > \frac{1}{2}$. Then we have

$$u_{RL}F_m(x') + u_{RR}\left(1 - F_m(x')\right) < u_{RL}\frac{1}{2}F_m'(1/2) + u_{RR}\left[(1 - F_m'(1/2)) + \frac{1}{2}F_m'(1/2)\right]$$

$$\leq u_{LL}\frac{1}{2}F_m'(1/2) + u_{RR}\left[(1 - F_m'(1/2)) + \frac{1}{2}F_m'(1/2)\right],$$

where the first expression is a $u$ type’s payoff under strategy profile $(x,x')$ and the last expression is her payoff under strategy profile $\sigma_C$. The first inequality follows from $u_{RR} > u_{LL} \geq u_{RL}$ by the assumption of unambiguous coordination preferences and from $(1 - F_m'(1/2)) \geq (1 - F_m'(x'))$ as $F_m'$ is nondecreasing.

The analysis for the player who sent message $m'$ is analogous. $\square$

### A.5 Proof of Proposition 3

We prove Proposition 3 through a series of claims. Note that, given the equilibrium in question, after message pairs $(m_L,m_L)$ and $(m_R,m_R)$ no Pareto improvement is possible. It remains to be shown that, while there is a Pareto-improving equilibrium (with new communication) after message pair $(m_L,m_R)$, all Pareto-improving equilibria after message pair $(m_L,m_R)$ are themselves CP-trumped. The following claims refer to the situation after observed message pair $(m_L,m_R)$.

**Claim 1.** Suppose a further message pair leads to updated beliefs of $\alpha, 1 - \alpha$ of the $L$ type being $L_1$ or $L_2$, respectively, and $\beta, 1 - \beta$ of the $R$ type being $R_1$ or $R_2$, respectively. The following table provides the full list of Bayes Nash equilibria in the updated coordination game without (further) communication:

| $L_1, L_2$ | $R_1, R_2$ | payoffs $L_1, L_2; R_1, R_2$ | $\alpha$ | $\beta$ |
|------------|------------|-------------------------------|----------|----------|
| $L, L$     | $L, L$     | $2, 2; 1, 1$                  | $[0, 1]$ | $[0, 1]$ |
| $R, R$     | $R, R$     | $1, 1; 2, 2$                  | $[0, 1]$ | $[0, 1]$ |
| $mix, R$   | $mix, L$   | $\frac{2}{3}, \frac{2}{3}; 2/3, 2/3$ | $\geq \frac{2}{3}$ | $\geq \frac{2}{3}$ |
| $mix, R$   | $R, mix$   | $\frac{2}{3}, \frac{2}{3}; \frac{16}{9}, \frac{1}{9}$ | $\geq \frac{1}{9}$ | $\leq \frac{2}{3}$ |
| $L, mix$   | $mix, L$   | $\frac{16}{9}, \frac{1}{9}; 2/3, 2/3$ | $\leq \frac{2}{3}$ | $\geq \frac{1}{9}$ |
| $L, mix$   | $R, mix$   | $\frac{16}{9}, \frac{1}{9}; \frac{16}{9}, \frac{1}{9}$ | $\leq \frac{1}{9}$ | $\leq \frac{1}{9}$ |
| $L, R$     | $R, L$     | $2(1 - \beta), \beta; 2(1 - \alpha), \alpha$ | $[\frac{1}{9}, \frac{2}{3}]$ | $[\frac{1}{9}, \frac{2}{3}]$ |
The last two columns provide the range of $\alpha$ and $\beta$ under which the various strategy profiles are equilibria.

**Proof.** The proof follows straightforwardly from the observations that a probability of opponent playing action L (R) of $\frac{1}{3}$ makes type $L_1$ ($R_1$) indifferent between actions L and R, while a probability of opponent playing action L (R) of $\frac{8}{9}$ makes type $L_2$ ($R_2$) indifferent between actions L and R.

**Claim 2.** Of the equilibria provided in Claim 1 equilibria that are not given in the following table are CP trumped by other equilibria.

| $L_1$, $L_2$ | $R_1$, $R_2$ | payoffs $L_1$, $L_2$; $R_1$, $R_2$ | $\alpha$ | $\beta$ | $\alpha + \beta$ |
|-------------|-------------|-----------------------------|--------|--------|------------------|
| $L$, $L$    | $L$, $L$    | 2, 2; 1, 1                 | $\in [0, 1]$ | $\in [0, 1]$ | |
| $R$, $R$    | $R$, $R$    | 1, 1; 2, 2                 | $\in [0, 1]$ | $\in [0, 1]$ | |
| $L$, mix    | $R$, mix    | $\frac{16}{9}$, $\frac{1}{9}$; $\frac{16}{9}$, $\frac{1}{9}$ | $\leq \frac{1}{9}$ | $\leq \frac{1}{9}$ | |
| $L$, $R$    | $R$, $L$    | $2(1 - \beta), \beta$; $2(1 - \alpha), \alpha$ | $\in \left[\frac{1}{9}, \frac{7}{18}\right]$ | $\in \left[\frac{1}{9}, \frac{7}{18}\right]$ | $\leq \frac{1}{2}$ |

**Proof.** All mixed equilibria except $((L,\text{mix}),(R,\text{mix}))$ are Pareto dominated by either (L,L) or (R,R), see Claim 1. Equilibrium $((L,R),(R,L))$ (when $\alpha, \beta$ are outside the domain given in the table above) is dominated by a convex combination of (L,L) and (R,R) (it is dominated by (L,L) if $\alpha \geq \frac{1}{2}$, dominated by (R,R) if $\beta \geq \frac{1}{2}$, and dominated by a joint lottery that yields (L,L) with probability $1 - 2\beta$ and (R,R) with the remaining probability $2\beta$ if $\alpha, \beta < \frac{1}{2} < \alpha + \beta$).

**Claim 3.** In any equilibrium of this game with or without additional communication, type $L_1$ ($R_1$) receives a payoff that is at least as high as that of type $L_2$ ($R_2$).

**Proof.** The stage game payoff matrix for $L_1$ weakly exceeds that of $L_2$. Suppose there is an equilibrium in which $L_2$ expects a strictly higher payoff that $L_1$. Then $L_1$ can imitate $L_2$ and get at least the same payoff, a contradiction.

**Claim 4.** Consider an equilibrium of this game with communication that CP trumps the considered equilibrium $((L,R),(R,L))$ after $(m_R, m_L)$ and that is not itself CP trumped by another strategy. If there is a message $m$ that only type $L_1$ ($R_1$) sends, then play after this message must be fully coordinated (against all opponent messages).

**Proof.** A message $m$ that only $L_1$ sends reveals $L_1$ (leads to an updated belief that $m$ sender is of type $L_1$ with probability $\alpha = 1$). Then only coordinated equilibria are possible as
undominated equilibria - see Claim 2 above for cases with $\alpha = 1$. An analogous argument can be made for type $R_1$.

**Claim 5.** In any equilibrium of this game with communication that CP trumps the considered equilibrium $((L,R),(R,L))$ and that is not itself CP trumped by another strategy, there can be no message sent with positive probability by type $L_1$ ($R_1$) that leads to coordinated play against all opponent messages.

**Proof.** Suppose there is a message $m$ that type $L_1$ sends with positive probability that leads to coordinated play (for all opponent messages). Then to Pareto-dominate the original equilibrium $L_1$ must expect a payoff of at least $16/9$. Then type $L_2$ could imitate $L_1$ and also obtain the same payoff that $L_1$ obtains (because when play is coordinated both types receive the same payoff). Any other message $m'$ that type $L_2$ sends must also provide the same payoff. Suppose play after message $m'$ is not fully coordinated. Then type $L_1$ can send message $m'$ and imitate $L_2$’s behavior and receive a strictly higher payoff than $L_2$ does. Thus, message $m'$ must also lead to fully coordinated play against all opponent messages. Any message $m''$ sent by $L_1$ and not $L_2$ must lead to fully coordinated play as well by Claim 4. Thus, all messages sent with positive probability must lead to fully coordinated play. Given this, both types $L_1$ and $L_2$ receive a payoff greater or equal to $16/9$. But then $R_1$ can only obtain a payoff of at most $3 - 16/9 = 11/9$ (with 3 being the maximal total payoff in any encounter), and the new equilibrium is no Pareto-improvement, a contradiction. 

**Claim 6.** In any equilibrium of this game with communication that CP trumps the considered equilibrium $((L,R),(R,L))$ and that is not itself CP trumped by another strategy, any message sent with positive probability by type $L_1$ ($R_1$) must also be sent by $L_2$ ($R_2$).

**Proof.** Suppose not and there is a message $m$ that reveals $L_1$. Then by Lemma 4 this message must lead to coordinated play which contradicts Lemma 5.

**Claim 7.** In any equilibrium of this game with communication that Pareto-dominates the considered equilibrium $(L,R),(R,L)$ and that is not itself CP trumped by another strategy, there must be a message-pair $(m,m')$ sent with positive probability (by both types, respectively) that leads to an $(L,R),(R,L)$ equilibrium with updated beliefs of $\alpha = P(L_1|m) \in [1/9, 1/2]$ and $\beta = P(R_1|m) \in [1/9, 1/2]$ that also satisfy $\alpha + \beta \leq 1/2$.

**Proof.** By Claim 5 every message sent must induce miscoordination against at least some opponent message. By Claim 6 every message $L_1$ sends $L_2$ also sends. Thus, there must
be a message $m$ that leads to an updated belief that $L_1$ sent this message of weakly more than $1/9$ (analogously, $m'$ for $R$ types). The only possible miscoordinated (and undominated) equilibrium given $(m,m')$ is given by $(L,R),(R,L)$ - see Table above. We must have $\alpha = P(L_1|m) \in [1/9,1/2]$ and $\beta = P(R_1|m) \in [1/9,1/2]$ that also satisfy $\alpha + \beta \leq 1/2$. Otherwise $(L,R),(R,L)$ is Pareto dominated, a contradiction. 

Claim 8. Consider the stage game with $\mu = P(L_1|m) \in [1/9,1/2]$ and $\nu = P(R_1|m) \in [1/9,1/2]$ that also satisfy $\mu + \nu \leq 1/2$. Then there is a strategy that CP trumps the equilibrium $((L,R),(R,L))$.

Proof. Consider the following strategy. Types $L_1$ and $R_1$ send message $m_1$ with probability 1. Types $L_2$ and $R_2$ send message $m_1$ with probability $1/3$ and $m_2$ with probability $2/3$. Continuation play is given by the following table:

|       | $m_1$  | $m_2$  |
|-------|--------|--------|
| $m_1$ | $(L,R),(R,L)$ | $(L,L),(L,L)$ |
| $m_2$ | $(R,R),(R,R)$ | $1/2((L,L),(L,L)) + 1/2((R,R),(R,R))$ |

with strategies given for $((L_1,L_2),(R_1,R_2))$ in this sequence. Importantly $((L,R),(R,L))$ is an equilibrium after $m_1,m_1$ because $1/9 \leq \alpha = P(L_1|m) = \nu/\nu + (1-\nu)/3 \leq 2/3$ (which is true for $\nu \leq 1/2 - 1/9 = 7/18 < 2/5$ and analogously $1/9 \beta \leq 2/3$. In this equilibrium $L_1$ and $R_1$ types have a payoff of $2(1-\nu)$ and $2(1-\mu)$, respectively, which is the same payoff they get in the original $(L,R),(R,L)$ equilibrium. Types $L_2$ and $R_2$ receive a payoff of more than 1, which is more than they receive in the original $((L,R),(R,L))$ equilibrium. Thus the new strategy CP trumps the original one. \qed

Claims 7 and 8 combine to prove that any strategy that CP trumps $((L,R),(R,L))$ in the given game is itself CP trumped. This proves that $((L,R),(R,L))$ is weakly communication proof. By Lemma 8 we also have that $((L,R),(R,L))$ is not strongly communication proof. This proves Proposition 3.

Note, however, that any strategy that is coordinated and mutual-preference consistent and has binary communication is strongly communication-proof also in the general setting, and that only the “$3 \Rightarrow 1$” part of the main result fails without the assumption of unambiguous coordination preferences. One can show that any communication-proof equilibrium strategy must satisfy mutual-preference consistency, but, possibly, need not satisfy the other two properties (namely, coordination and binary communication).
B More on Properties of Strategies

In this appendix we demonstrate that no single one of the three properties (mutual-preference consistency, coordination, and binary communication) is implied by the other two. Clearly a strategy that has binary communication and is coordinated must be an equilibrium. No other combination of two of the three properties implies that a strategy is an equilibrium. Finally, we also define what it means for a strategy to be ordinal preference-revealing and show that this is implied by it being mutual-preference consistent.

Consider the following strategy $\sigma = (\mu, \xi)$ in the game with communication with a message set $M$ that contains at least three elements. Let $m_L^1, m_L^2, m_R \in M$, let

$$\mu(u) = \begin{cases} 
  m_L^1 & \text{if } u \leq \frac{1}{4} \\
  m_L^2 & \text{if } \frac{1}{4} < u \leq \frac{1}{2} \\
  m_R & \text{if } u > \frac{1}{2}
\end{cases}$$

and let $\xi$ be such that $\xi(m_L^i, m_L^j) = L$ for all $i, j \in \{1, 2\}$, $\xi(m_R, m_R) = R$, $\xi(m_L^1, m_R) = \xi(m_R, m_L^1) = R$, and $\xi(m_L^2, m_R) = \xi(m_R, m_L^2) = L$. This strategy is mutual-preference consistent and coordinated but does not have binary communication. It is not an equilibrium as types $u \leq \frac{1}{4}$ would strictly prefer to send message $m_L^2$.

Consider the following strategy $\sigma = (\mu, \xi)$ in the game with communication with a message set $M$ that contains at least two elements. Let $m_L, m_R \in M$, let

$$\mu(u) = \begin{cases} 
  m_L & \text{if } u \leq \frac{1}{2} \\
  m_R & \text{if } u > \frac{1}{2}
\end{cases}$$

and let $\xi$ be such that $\xi(m_L, m_L) = L$, $\xi(m_R, m_R) = R$, $\xi(m_L, m_R) = \frac{1}{4}$, and $\xi(m_R, m_L) = \frac{3}{4}$. This strategy is mutual-preference consistent, has binary communication, but is not coordinated. For almost all type distributions $F$ this is not an equilibrium: it is only an equilibrium if $F$ satisfies

$$\frac{F(\frac{3}{4}) - F(\frac{1}{2})}{1 - F(\frac{1}{2})} = \frac{1}{4} \quad \text{and} \quad \frac{F(\frac{1}{4})}{F(\frac{1}{2})} = \frac{3}{4}.$$ 

Finally, for a strategy that has binary communication and is coordinated but not mutual-preference consistent, consider the equilibrium strategy that always leads to coordination on $L$ for any pair of messages.

Note also that an equilibrium does not necessarily satisfy any of the three properties. The
interior cutoff babbling equilibria mentioned in Section 3 are not coordinated and not mutual-preference consistent. The equilibrium of Example 1 does not have binary communication.

Call a strategy $\sigma = (\mu, \xi) \in \Sigma$ ordinal preference-revealing if there exist two nonempty, disjoint, and exhaustive subsets of $\text{supp}(\bar{\mu})$ denoted by $M_L$ and $M_R$ (i.e., $\text{supp}(\bar{\mu}) = M_L \cup M_R$) such that if $u < 1/2$, then $\mu_u(m) = 0$ for each $m \in M_R$, and if $u > 1/2$, then $\mu_u(m) = 0$ for each $m \in M_L$. With an ordinal preference-revealing strategy a player indicates her ordinal preferences. A strategy $\sigma$ that is mutual-preference consistent is also ordinal preference-revealing (but not vice versa). Suppose not. Then there is a message $m$ and two types $u < 1/2$ and $v > 1/2$ such that $\mu_u(m), \mu_v(m) > 0$. But then no matter how we specify $\xi(m,m)$ we get either that if two types $u$ meet they do not coordinate on $L$ with probability one or if two types $v$ meet they do not coordinate on $R$ with probability one.

C Multiple Rounds of Communication

Consider a variant of the coordination game with communication in which players have $T \geq 1$ of rounds of communication. In each such round players simultaneously send messages from the set $M$. Players observe messages after each round and can, thus, condition their message choice and then their final action choice on the history of observed message pairs up to the point in time where they take their message or action decision. Renegotiation then possibly takes place once at the end of this communication phase but before the final action choices are made. Let $\mathcal{M} = \bigcup_{t=0}^{T-1} (M \times M)^t$, where $(M \times M)^0 = \emptyset$.

A (pure) message protocol is a function $m : \mathcal{M} \rightarrow M$ that describes the message sent by an agent as a deterministic function of the message profiles observed in the previous rounds of communication. Let $\mathcal{M}$ be the set of all message protocols. A strategy $\sigma = (\mu, \xi)$ is a pair where $\mu : U \rightarrow \Delta(\mathcal{M})$ denotes the message function, prescribing a (possibly random) message protocol for each type, and $\xi : (M \times M)^T \rightarrow U$ denotes the action function by means of describing the cutoff (the highest possible value of $u$) for the two players to choose action $L$ after observing the final message history. Renegotiation is modeled, as in the main text, as a possibility for the two players to play an equilibrium of a new game with another round of communication after all messages are sent, possibly using a different message set.

Next, we adapt the notion of binary communication to fit multiple rounds of communication. For any message protocol $m \in \mathcal{M}$, let $\beta^\sigma(m)$ denote the expected probability of a player’s opponent playing $L$ conditional on the player following message protocol $m \in \mathcal{M}$ and the opponent following strategy $\sigma = (\mu, \xi) \in \Sigma$. We say that strategy $\sigma$ has binary com-
if there are two numbers $0 \leq \beta^\sigma \leq \bar{\beta}^\sigma \leq 1$ such that for all message protocols $m \in \mathcal{M}$ we have $\beta^\sigma(m) \in [\beta^\sigma, \bar{\beta}^\sigma]$, for all message protocols $m \in \mathcal{M}$ such that there is a type $u < 1/2$ with $\mu_u(m) > 0$ we have $\beta^\sigma(m) = \beta^\sigma$, and for all message protocols $m \in \mathcal{M}$ such that there is a type $u > 1/2$ with $\mu_u(m) > 0$ we have $\beta^\sigma(m) = \bar{\beta}^\sigma$. That is, binary communication implies that players use just two kinds of message protocols: any message protocol used by types $u < 1/2$ induces the consequence of maximizing the probability of the opponent to play $L$, and any message protocol used by types $u > 1/2$ induces the opposite consequence of maximizing the probability of the opponent to play $R$.

Theorem 1, together with Propositions 1 and 2, holds in this setting with minor adaptations to the proof (omitted for brevity). Thus, regardless of the length of the pre-play communication, agents can reveal only their preferred outcome (but not the strength of their preference), and, regardless of having access to additional rounds of communication, they cannot improve the ex-ante expected payoff relative to the payoff induced by a single round of communication with a binary message.

References

Aumann, R. J. and S. Hart (2003). Long cheap talk. *Econometrica* 71(6), 1619–1660.

Aumann, R. J. and M. B. Maschler (1968). *Repeated Games of Incomplete Information: An Approach to the Non-zero-sum Case*. Report to the U.S. Arms Control and Disarmament Agency. Mathematica Inc., Princeton (reprinted in Chapter 5 of *Repeated Games with Incomplete Information*, 1995, MIT Press).

Baliga, S. and T. Sjöström (2004). Arms races and negotiations. *The Review of Economic Studies* 71(2), 351–369.

Benoit, J.-P. and V. Krishna (1993). Renegotiation in finitely repeated games. *Econometrica* 61(2), 303–323.

Bhaskar, V. (1998). Noisy communication and the evolution of cooperation. *Journal of Economic Theory* 82(1), 110–131.

Binmore, K. and L. Samuelson (1999). Evolutionary drift and equilibrium selection. *The Review of Economic Studies* 66(2), 363–393.
Blume, A., Y.-G. Kim, and J. Sobel (1993). Evolutionary stability in games of communication. Games and Economic Behavior 5(4), 547–575.

Blume, A. and J. Sobel (1995). Communication-proof equilibria in cheap-talk games. Journal of Economic Theory 65(2), 359–382.

Busch, L.-A. and Q. Wen (1995). Perfect equilibria in a negotiation model. Econometrica 63(3), 545–565.

Cramton, P. and J. A. Schwartz (2000). Collusive bidding: Lessons from the FCC spectrum auctions. Journal of Regulatory Economics 17(3), 229–252.

Crawford, V. and H. Haller (1990). Learning how to cooperate: Optimal play in repeated coordination games. Econometrica 58, 571–596.

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. Econometrica 50(6), 1431–1451.

Dawkins, R. (1976). The Selfish Gene. New York: Oxford University Press.

Ellison, G. and D. Fudenberg (2000). Learning purified mixed equilibria. Journal of Economic Theory 90(1), 84–115.

Farrell, J. (1987). Cheap talk, coordination and entry. Rand Journal of Economics 18, 34–39.

Farrell, J. and E. Maskin (1989). Renegotiation in repeated games. Games and Economic Behavior 1(4), 327–360.

Forges, F. (1994). Posterior efficiency. Games and Economic Behavior 6(2), 238–261.

Fudenberg, D. and D. M. Kreps (1993). Learning mixed equilibria. Games and Economic Behavior 5(3), 320–367.

Goffman, E. (1971). Relations in Public: Microstudies of the Social Order. New York, USA: Basic Books.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. Journal of Theoretical Biology 7(1), 1–16.

Harsanyi, J. C. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. International Journal of Game Theory 2(1), 1–23.
Harsanyi, J. C. and R. Selten (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, Massachusetts: MIT Press.

Hart, O. D. and J. Tirole (1988). Contract renegotiation and coasian dynamics. *The Review of Economic Studies* 55(4), 509–540.

Heller, Y. and C. Kuzmics (2021). Renegotiation and coordination with private values (extended version). Available at https://ssrn.com/abstract=3448035.

Hurkens, S. and K. H. Schlag (2003). Evolutionary insights on the willingness to communicate. *International Journal of Game Theory* 31(4), 511–526.

Jelnov, A., Y. Tauman, and C. Zhao (2018). Stag hunt with unknown outside options. mimeo.

Kandori, M., G. J. Mailath, and R. Rob (1993). Learning, mutation, and long run equilibria in games. *Econometrica* 61(1), 29–56.

Kim, Y.-G. and J. Sobel (1995). An evolutionary approach to pre-play communication. *Econometrica* 63(5), 1181–1193.

Maestri, L. (2017). Dynamic contracting under adverse selection and renegotiation. *Journal of Economic Theory* 171, 136–173.

Matsui, A. (1991). Cheap-talk and cooperation in a society. *Journal of Economic Theory* 54(2), 245–258.

Maynard Smith, J. and G. R. Price (1973). The logic of animal conflict. *Nature* 246, 15–18.

Nash, J. (1950). *Non-Cooperative Games*. Princeton University. Ph. D. dissertation.

Neary, P. R. and J. Newton (2017). Heterogeneity in preferences and behavior in threshold models. *Journal of Mechanism and Institution Design* 2, 1.

Neeman, Z. and G. Pavlov (2013). Ex post renegotiation-proof mechanism design. *Journal of Economic Theory* 148(2), 473–501.

Newton, J. (2017). Shared intentions: The evolution of collaboration. *Games and Economic Behavior* 104, 517–534.

Rabin, M. (1994). A model of pre-game communication. *Journal of Economic Theory* 63(2), 370–391.
Robson, A. J. (1990). Efficiency in evolutionary games: Darwin, Nash, and the secret handshake. Journal of Theoretical Biology 144(3), 379–396.

Safronov, M. and B. Strulovici (2019). Contestable norms. mimeo.

Sandholm, W. H. (2007). Evolution in bayesian games II: Stability of purified equilibria. Journal of Economic Theory 136(1), 641–667.

Santos, F. C., J. M. Pacheco, and B. Skyrms (2011). Co-evolution of pre-play signaling and cooperation. Journal of Theoretical Biology 274(1), 30–35.

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

Schelling, T. C. (1960). The Strategy of Conflict. Cambridge, Massachusetts, USA: Harvard University Press.

Sobel, J. (1993). Evolutionary stability and efficiency. Economics Letters 42(2-3), 301–312.

Strulovici, B. (2017). Contract negotiation and the Coase conjecture: A strategic foundation for renegotiation-proof contracts. Econometrica 85(2), 585–616.

Swinkels, J. M. (1992). Evolutionary stability with equilibrium entrants. Journal of Economic Theory 57(2), 306–332.

von Neumann, J. and O. Morgenstern (1944). Theory of Games and Economic Behavior. Princeton, New Jersey: Princeton University Press.

Wärneryd, K. (1991). Evolutionary stability in unanimity games with cheap talk. Economics Letters 36(4), 375–378.

Wärneryd, K. (1993). Cheap talk, coordination, and evolutionary stability. Games and Economic Behavior 5(4), 532–546.

Young, H. P. (1993). The evolution of conventions. Econometrica 61(1), 57–84.

Young, H. P. (1998). Individual Strategy and Social Structure: An Evolutionary Theory of Institutions. Princeton, New Jersey: Princeton University Press.