Communication, Renegotiation and Coordination with Private Values (Extended Version)

Heller, Yuval and Kuzmics, Christoph

Graz University, Bar-Ilan University

14 September 2020

Communication, Renegotiation and Coordination with Private Values (Extended Working Paper Version)*

Yuval Heller[†] and Christoph Kuzmics[‡]

July 27, 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

*We have benefited greatly from discussions with Srinivas Arigapudi, Tilman Borgers, Michael Greinecker, Jonathan Newton, Bill Sandholm, and Joel Sobel. We would like to express our gratitude to participants of various audiences for many useful comments: LEG2018 & LEG2019 conferences (in Lund and Bar-Ilan university, respectively), Bielefeld Game Theory 2018 workshop, Israeli Game Theory 2018 conference in IDC, and seminar audiences at Caltech, Tel Aviv University, University of Cyprus, Haifa University, and UC San Diego. Yuval Heller is grateful to the European Research Council for its financial support (#677057).

†Department of Economics, Bar Ilan University, Israel. Email: yuval.heller@biu.ac.il.

‡Department of Economics, University of Graz, Austria. Email: christoph.kuzmics@uni-graz.at.

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 coordinate to collude (Cramton and Schwartz, 2000).[2]

Players also typically coordinate effectively in our second example: Pedestrians typically are

[1] The notion of communication-proofness was introduced by Blume and Sobel (1995) in their study of sender-receiver games with one-sided private information.

[2] In 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.

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.[3]

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 or on the exact timing of the renegotiation (relative to the communication), and are thus robust to changes in players' (first- or higher-order) beliefs, interim Pareto efficient, and evolutionarily stable in the traditional sense. In particular, communication-proof equilibrium strategies remain communication-proof even in setups in which the players' distributions of types are interdependent.

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.[4] One possible explanation for some, fairly simple, examples of coordination 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

[3] The example is motivated by Goffman (1971, Chapter 1, 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."

[4] Evidence for miscoordination in lab experiments is reported by, e.g., Van Huyck et al. (1990), Mehta et al. (1994), and Blume and Gneezy (2010). According to Farrell and Klemperer (2007, Section 3.4) miscoordination also occurs regularly in real-life economic interactions.

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.[5] 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.[6]

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.

All the above-mentioned literature (on coordination) 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-proof equilibrium, 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 multiplicity problem to a large extent, in the sense that communication-proof equilibria, in contrast to Bayesian Nash equilibria, make very similar predictions.

Baseline model and extensions While we take into account incomplete information in the coordination problem, we try to keep the baseline model tractable by simplifying other aspects of the problem. We restrict attention in the baseline model to (incomplete information) two-player

[5]See, Weibull (1995) and Sandholm (2010) for a textbook treatment of evolutionary game theory.

[6]Kandori et al. (1993) and Young (1993) show that in the very long run and under persistent low-probability errors an evolutionary (learning) process leads to the risk-dominant, not necessarily Pareto-dominant, pure strategy equilibrium in two-by-two coordination games.

two-action (coded as left and right) coordination problems, in which every type has a coordination concern. Different types only differ in how much they prefer coordination on one action over the other one. Players can communicate only once before the game is played.

Sections 2.1 and 8 provide a series of extensions and robustness checks of our baseline model. First, we show that all our results hold for any length of pre-play communication. Second, we show that all our results hold for all two by two incomplete information coordination games, in which for all feasible types the preferred coordinated action is also the risk-dominant action (i.e., the best reply against an opponent who plays each action uniformly; see Harsanyi and Selten, 1988). Example 2 demonstrates that without this assumption there can be additionalcommunication-proof equilibria, in which players fail to coordinate after some messages. Next, we extend our key results to general (possibly asymmetric) coordination games, which might involve more than two players and more than two actions. Finally, we study a variant of our baseline model in which a few types have dominant actions. We show that in this setup, there is a unique communication-proof equilibrium strategy among the strategies satisfying our three key properties (the identity of this unique strategy depends on the distribution of types with dominant actions).

Structure Section 2 presents our baseline 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 appropriately adapted to our incomplete-information strategic setting. Section 5 presents the main result and a sketch of its proof. Section 6 discusses the efficiency properties of communication-proof equilibria. Section 7 discusses additional notions of evolutionary stability beyond the secret-handshake stability implied by communication-proofness. Section 8 provides a series of robustness checks and extensions. Section 9 concludes. The formal proofs are presented in the online appendices.

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 a full support on the unit interval $U=[0,1]$ and with density f (i.e., $f(u)>0$ for each $u\in U$).[7]

[7] Allowing distributions without full support induces a minor difference in our results: in this setup communication-proofness implies binary communication (as defined in Sec. 3) only of messages that are used with positive probability,

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).

Table 1: Payoff Matrix of the Coordination Game

		Type v	
		L	R
Type u	L	1-u, 1-v	0, 0
	R	0, 0	u, v

We call this game the *coordination game without communication* and denote it by Γ.

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 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. We assume that messages are costless. We call the game, so amended, the *coordination game with communication* and denote it by $\langle \Gamma, M \rangle$.

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* that describes

whereas with full support it implies binary communication also of unused messages.

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 (μ, ξ) 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 $u > \xi(m, m')$. (The choice that the threshold type plays L does not affect our analysis, given the assumption of F being atomless.) Let Σ be the set of all strategies in the game $\langle \Gamma, M \rangle$.

Let $\mu_u(m)$ denote the probability, given message function μ, 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 $\xi(m, m') \geq \sup(u \in U | \mu_u(m) > 0)$), and we write $\xi(m, m') = R$ when all types play R (i.e., when $\xi(m, m') \leq \inf(u \in U | \mu_u(m) > 0)$).

2.1 Simplifying Assumptions and Their Relaxation

For ease of exposition we present a simple model. Various simplifying assumptions can be relaxed without affecting any of the paper's results.

Number of available messages The assumption of M being finite is taken to simplify the notation; all of 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.

Multiple rounds of communication The baseline model assumes a single round of communication. In Appendix E.1 we show that all of our results hold in a setup in which cheap talk includes multiple rounds. That is, in our setup of communication-proof equilibria in coordination games, the length of communication does not matter (in contrast to the results in other setups of incomplete-information games; see, e.g., Aumann and Hart, 2003).

Multidimensional sets of types In Appendix E.2 we study general symmetric two-action two-player coordination games, where miscoordination may result in different payoffs to the L and R players. The "1 $\Rightarrow$ 2" part of Theorem 1 still holds in this general setup: any strategy that satisfies the three key properties is strongly communication-proof. Theorem 2 shows that the "$3 \Rightarrow 1$" part of the theorem holds as well under the restriction of "unambiguous coordination preferences," which requires that for all feasible types the preferred coordinated action also be the risk-dominant action (i.e., the best reply against an opponent who plays each action uniformly; see Harsanyi and Selten, 1988). Example 2 in Appendix E.2 demonstrates that without this restriction (i.e., with stag-hunt-like types for which the payoff-dominant action does not coincide with the risk-dominant action), equilibria with miscoordination can be communication-proof.

Asymmetric coordination game and asymmetric equilibria The baseline model assumes that both players' types have the same distribution F, and our solution concept focuses on symmetric equilibria. This is done to simplify the notation, and to ease the analysis of evolutionary stability in Section 7 (where some of the relevant solution concepts are defined only for symmetric games). Appendix E.4 shows that all our results hold in asymmetric coordination games in which the distributions of the types of the two players' positions may differ, or when allowing asymmetric equilibria of symmetric coordination games.

More than two players Our baseline model has only two players. In Appendix E.3 we show that our results hold for any number of $n \geq 2$ players, under the assumption that the payoff of each player of type u is equal to u if all players play R, it is equal to $1-u$ if all players play L, and it is equal to zero if all players do not play the same action. The interesting case in which players can get positive payoffs by coordination of a subset of the players on the same action is left for future research.

General action functions The baseline model restricts the action functions to the set of cutoff functions of the form $\xi : M \times M \to U$. In principle, we should allow more general action functions $\xi : U \times M \times M \to \triangle\{L, R\}$, which specify the probability that an agent chooses L as a function of the observed message profile and the agent's type. It is simple to see, however, and proven in Lemma 1 in Appendix A.1, that any "generalized" strategy is dominated by a strategy that uses a cutoff action function in the second stage. The intuition, is that following the observation of any pair of messages, lower types always gain more (less) than higher types from choosing L (R). Thus, the restriction to cutoff action functions is without loss of generality.

Other extensions affect some (but not all) of our results: allowing more than two actions, and allowing extreme types with dominant actions. We postpone the discussion of these extensions to Section 8

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 examples of equilibria in the coordination game with communication with and without these properties. These equilibria are illustrated in Figure 1 in the end of this section.

Given a strategy profile (σ, σ') 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 σ and faces an opponent of type v who follows strategy σ'. Formally, for $\sigma = (\mu, \xi)$ and $\sigma' = (\mu', \xi')$,

$$\pi_{u,v}\left(\sigma,\sigma'\right) = \sum_{m\in M}\sum_{m'\in M}\mu_u\left(m\right)\mu_v\left(m'\right)\left((1-u)1_{\{u\le\xi(m,m')\}}1_{\{v\le\xi'(m',m)\}} + u1_{\{u>\xi(m,m')\}}1_{\{v>\xi'(m',m)\}}\right),$$

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

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

denote the expected interim payoff of a player of type u who follows strategy σ and faces an opponent with a random type who follows strategy σ'. Finally, let,

$$\pi\left(\sigma,\sigma'\right) = \mathbb{E}_u\left[\pi_u\left(\sigma,\sigma'\right)\right] \equiv \int_{u=0}^{1}\pi_u\left(\sigma,\sigma'\right)f\left(u\right)du$$

denote the ex-ante expected payoff of an agent who uses strategy σ against strategy σ'.

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

Three key properties We call a strategy $\sigma = (\mu, \xi) \in \Sigma$ *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 *coordinated* if $\xi(m, m') = \xi(m', m) \in \{L, R\}$ for any pair of messages $m, m' \in \text{supp}(\bar{\mu})$. A coordinated strategy never leads to miscoordination after any (used) message pair.

For any message $m \in M$, define the expected probability of a player's opponent playing L conditional on the player sending message m and the opponent following strategy $\sigma = (\mu, \xi) \in \Sigma$, as

$$\beta^{\sigma}(m) = \int_{u=0}^{1}\sum_{m'\in\text{supp}(\mu_u)}\mu_u(m')1_{\{u\le\xi(m',m)\}}f(u)du.$$

We say that strategy σ has *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 $u < 0.5$ induces the same consequence of maximizing the probability of the opponent playing L, and any message sent by types $u > 0.5$ 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 *no single one of these three properties is implied by the other two*. Clearly, a strategy that has binary communication and is coordinated must be an equilibrium. In Appendix B we also show that no combination of any two of these three properties implies that a strategy is an equilibrium.

Left tendency α^{σ} 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 α^{σ}, and refer to it as the *left tendency* of the strategy. We can express $\underline{\beta}$ and $\overline{\beta}^{\sigma}$ as follows:

$$\underline{\beta}^{\sigma} = F(1/2)\alpha^{\sigma} \quad \text{and} \quad \overline{\beta}^{\sigma} = F(1/2) + (1 - F(1/2))\,\alpha^{\sigma}.$$

The first equality ($\underline{\beta}^{\sigma} = F(1/2)\alpha^{\sigma}$) 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 α^{σ}. The second equality ($\overline{\beta}^{\sigma} = F(1/2) + (1 - F(1/2))\,\alpha^{\sigma}$) 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 outcome is L, and they coordinate on L with a probability of α^{σ} if the opponent's preferred action is R.

Examples of equilibria satisfying all properties The following strategies, denoted by σ_L, σ_R, and σ_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 σ_L and σ_R are given by the pairs (μ^*, ξ_L) and (μ^*, ξ_R), respectively. The message function μ^* 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 ξ_L and ξ_R are defined as follows:

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

This means that the "fallback norm" of σ_L (which is applied when the agents have different preferred outcomes) is to coordinate on L, while that of σ_R is to coordinate on R. In other words the left tendency of σ_L is one and the left tendency of σ_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 $1/2$ each by the agent. Formally, the message function μ_C is defined as follows:

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

where $\alpha m \oplus (1-\alpha)m'$ is a lottery with a probability of α 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:

$$\xi_C(m, m') = \begin{cases} R & (m, m') \in \{(m_{R,0}, m_{R,0}), (m_{R,0}, m_{R,1}), (m_{R,0}, m_{L,0}), (m_{R,1}, m_{L,1}) \\ & \qquad (m_{R,1}, m_{R,1}), (m_{R,1}, m_{R,0}), (m_{L,0}, m_{R,0}), (m_{L,1}, m_{R,1})\} \\ L & \text{otherwise.} \end{cases}$$

The outcome of σ_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 σ 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 $\alpha^{\sigma} \in [0,1]$, which is a rational number, 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 random messages induces α^{σ} (Aumann and Maschler, 1968; see also Heller, 2010, for a recent implementation, which is robust to joint deviations). This is demonstrated for $\alpha^{\sigma} = 1/2$ in the strategy σ^{C} presented above.

Note that of all the strategies that satisfy the three properties, strategies σ_L and σ_R are the *simplest* in terms of the number of "bits" needed to implement the message function. Strategy σ_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 σ_L favors types below $1/2$, and strategy σ_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 iff 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 for a specific distribution F is Example 1 in Section 6. It is not completely straightforward to construct such examples for all possible distributions F if we consider only a single round of communication as we do in the main body of this paper. In Section 8, however, we show that our main results continue to hold if we allow multiple rounds of communication. In the case of multiple rounds of communication it is relatively straightforward to construct examples of equilibria that do not have binary communication and that, therefore, reveal some cardinal content of the players' preferences. For simplicity, assume that the distribution F is symmetric around $1/2$. That is $f(x) = f(1-x)$ for all $x \in [0,1]$ or, equivalently, $F(x) = 1 - F(1-x)$ for all $x \in [0,1]$. In particular, we have that $F(1/2) = 1/2$.

The equilibrium σ_{ex} is such that there is an x satisfying $0 < x < 1/2$ such that in the first round of communication players indicate whether their preferences are "extreme" (i.e., $u \leq x$ or $u > 1-x$)

or "moderately left" ($x < u \leq 1/2$) or "moderately right" ($1/2 < u \leq 1-x$). In the second round individuals only reveal additional information if in the first round one sent the extreme message and the other a moderate message, in which case the extreme type now reveals which side she prefers ($u \leq 1/2$ or $u > 1/2$). In this case, joint play is dictated by the extreme type's preferences. If both players sent the extreme message in the first round, then there is no more communication and both types follow their inclination (play L if $u \leq 1/2$ and R otherwise). This leads to miscoordination with a conditional probability of a half.

Any two moderate types essentially play the coordinated strategy σ_C; that is if they have the same preferred outcome, then they play it, and, otherwise, they use communication to induce a joint fair lottery over both playing L and both playing R. In Appendix C we formally present this strategy and show that for any distribution F there is an $x \in (0, 1/2)$ such that this strategy is a Nash equilibrium of the coordination game with two rounds of communication.

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: σ_L, σ_R, and σ_C, the babbling equilibrium of always playing R, and the equilibrium σ_{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 Σ 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 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) Γ 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 [8] $m \in \text{supp}(\bar{\mu})$, i.e.,

$$f_m(u) = \frac{f(u)\mu_u(m)}{\bar{\mu}(m)},$$

[8] The 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.

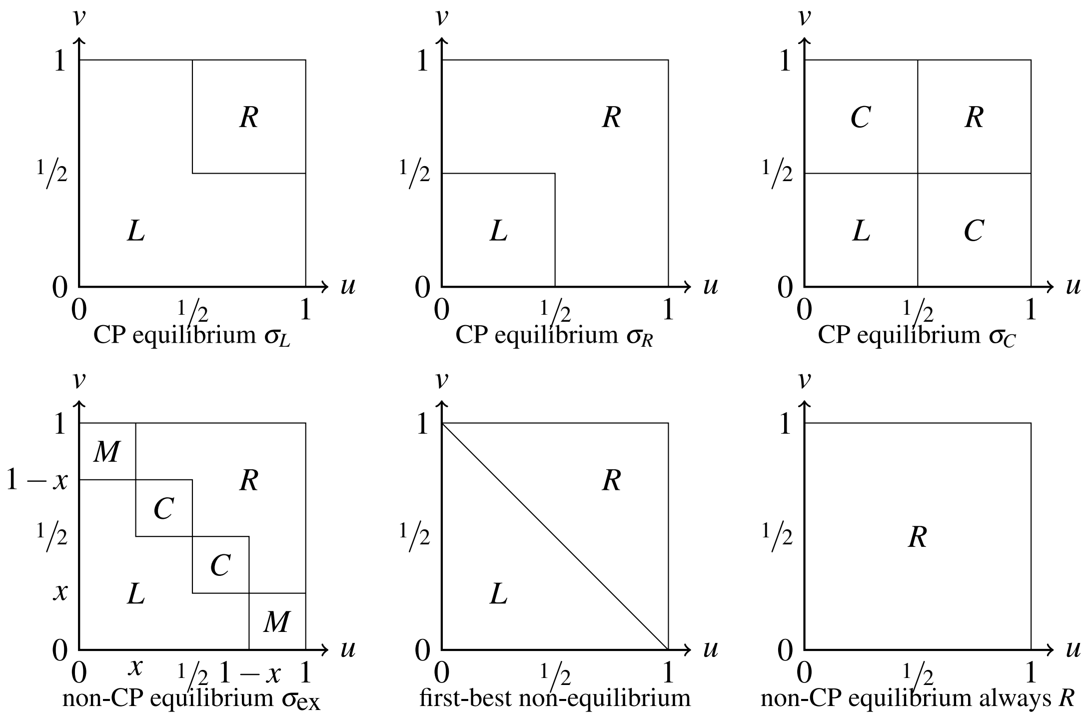

Figure 1: **Six example strategies.** The axis represent the two players' types u and v. Letters L, R, C, and M represent coordination on L, coordination on R, coordination on a random action, and miscoordination (both players playing their preferred action), respectively.

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 $\pi_u^H(\sigma, \sigma')$ denote the expected payoff for the player using strategy σ of type u given strategy profile (σ, σ') in game $\langle\Gamma(G, H), M\rangle$:

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

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

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

Let $\mathscr{E}(G, H)$ be the set of all (possibly asymmetric) equilibria of the coordination game with communication [9] $\langle\Gamma(G, H), M\rangle$. Let $\pi_u^{(m,m'),H_{m'}}(\sigma, \sigma')$ (resp., $\pi_v^{(m,m'),G_m}(\sigma, \sigma')$) denote the *post-communication*

[9] As the support of G and H may generally be a strict subset of $[0, 1]$, there may be many equivalent strategies (and,

payoff for the player using strategy σ (resp., σ') 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 $m \in \text{supp}(\sigma), m' \in \text{supp}(\sigma')$:

$$\pi_u^{(m,m'),H}(\sigma,\sigma') = \begin{cases} (1-u)H_{m'}(\xi'(m',m)) & \text{if } u \le \xi(m,m') \\ u(1-H_{m'}(\xi'(m',m))) & \text{if } u > \xi(m,m') \end{cases}$$

$$\pi_v^{(m,m'),G}(\sigma,\sigma') = \begin{cases} (1-v)G_m(\xi(m,m')) & \text{if } v \le \xi'(m',m) \\ v(1-G_m(\xi(m,m'))) & \text{if } v > \xi'(m',m). \end{cases}$$

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

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. $(\tau, \tau') \in \mathscr{E}(G_m, H_{m'})$, and
2. $\pi_u^{H_{m'}}(\tau, \tau') \ge \pi_u^{(m,m'),H}(\sigma, \sigma')$ and $\pi_v^{G_m}(\tau, \tau') \ge \pi_v^{(m,m'),G}(\sigma, \sigma')$, 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 $v \in \text{supp}(H_{m'})$.

We say that a strategy σ 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 σ. 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 $\sigma \in \mathscr{E}$ is *strongly communication-proof* if it is not CP trumped with respect to (F,F) by any strategy profile.

Definition 3. An equilibrium strategy $\sigma \in \mathscr{E}$ is *weakly communication-proof* if for any strategy profile (τ, τ') that CP trumps (σ, σ) with respect to (F,F) and message profile (m,m'), there exists a strategy profile (ρ, ρ') that CP trumps (τ, τ') 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 (nonempty) Pareto frontier of the set of Nash equilibria, i.e., to the subset of Nash

thus, many equivalent equilibria).

[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.

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 σ, a small group of experimenting agents can recognize each other by means of a "secret handshake" and play a Pareto-improving equilibrium σ' with each other and play the inferior equilibrium σ 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 σ' with a higher ex-ante payoff than the current equilibrium σ, then agents would use a secret handshake to play σ' 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 due to its ex-ante advantage.

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 (1) only allowing "stable" secret handshake that are robust to additional deviations (which is captured by the notion of weak communication-proofness), or (2) 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 σ be a strategy of the game with communication $\langle \Gamma, M \rangle$. The following three statements are equivalent:*

1. *σ is mutual-preference consistent, coordinated, and has binary communication.*
2. *σ is a strongly communication-proof equilibrium strategy.*
3. *σ 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 σ_L, σ_R, and σ_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 σ_L, σ_R, or σ_C. To see this, suppose first that both players use thresholds below $1/2$. Then this strategy is Pareto-dominated by σ_R as types above $1/2$ gain because σ_R induces their first-best outcome, and types below $1/2$ gain because σ_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 σ_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 σ_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 σ_R, which would allow them to coordinate on L if and only if both types are below $1/2$. $\square$

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 σ is a communication-proof equilibrium in $\Gamma(F)$ iff it is 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.

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, σ_L and σ_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 σ_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.[11] The

[11] One can easily adapt the example to an atomless distribution of types, in which each atom is replaced with a

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 ε sufficiently small. Theorem 1 (together with the symmetry of the distribution F) implies that with communication, any communication-proof equilibrium strategy (in particular σ_L or σ_R) induces the same expected ex-ante payoff of $3/5 > 1/2$ for all ε 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; see the similar construction in Case (II) of the proof of Proposition 9 in Appendix E.7.3). 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 $a, b \in \{r, R\}$, $\xi(m_L, m_r) = \xi(m_r, m_L) = L$, $\xi(m_l, m_R) = \xi(m_R, m_l) = R$, $\xi(m_l, m_r) = \xi(m_r, m_l)$ be a joint lottery to coordinate on L or R with probability $1/2$ each, and, finally, let $\xi(m_L, m_R) = \xi(m_R, m_L)$ 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 $\phi : [0,1]^2 \to \Delta\left(\{L,R\}^2\right)$ 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 Φ be the set of all such functions.

Any strategy of any coordination game with communication induces a social choice function in Φ, but not all social choice functions in Φ can be generated by a strategy of a given coordination game with communication. One can interpret Φ 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.

continuum of nearby 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.

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 ϕ, 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.

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.[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 σ is characterized by its left tendency α^{σ}. In order for a social choice function ϕ to improve the payoff of any type $u < 1/2$ (resp., $u > 1/2$) relative to the payoff induced by σ, it must be that ϕ induces any $u < 1/2$ (resp., $u > 1/2$) to coordinate on L with probability larger (resp., smaller) than α^{σ}. 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 α^{σ}. However, these two requirements contradict each other. □

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 σ_L and σ_R.

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

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

[13] As discussed at the end of Appendix E.4, 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.

Sketch of proof; see Appendix A.3 for the formal proof. Let α^{σ} be the probability of two players who each follow σ to coordinate on L, conditional on the players having different preferred outcomes. It is easy to see that σ is dominated by the communication-proof equilibrium strategy with the same left tendency α^{σ}, and that the payoff of the latter strategy is a convex combination of the payoffs of σ_L and σ_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 σ_L or σ_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 σ_L or σ_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 from Remark 2 and the text preceding it 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':

$$\pi_u\left(x,x'\right) = 1_{\{u\leq x\}}F\left(x'\right)(1-u) + 1_{\{u>x\}}\left(1-F\left(x'\right)\right)u,$$

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 σ_L, σ_R, or σ_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.

7 Evolutionary Stability

A common interpretation of a Nash equilibrium is a convention that is reached as a result of a process of social learning when similar games are repeatedly played within a large population. This interpretation seems very apt, for instance, if we think of our motivating example of how pedestrians avoid bumping into each other. Specifically, consider a population in which a pair of agents from a large population are occasionally randomly matched and play the coordination game with communication $\langle \Gamma, M \rangle$. The agents can observe past behavior of other agents who played similar games in the past. It seems plausible that the aggregate behavior of the population would gradually converge into a self-enforcing convention, which is a symmetric Nash equilibrium of $\langle \Gamma, M \rangle$ (see the "mass action" by Nash (1950); and see Weibull (1995) and Sandholm (2010) for a textbook introduction).

We have argued that communication-proofness is a necessary condition of evolutionary stability by capturing the idea of stability with respect to secret-handshake mutations as in Robson (1990). In this section we report results from Appendix D in which we investigate the evolutionary stability properties of both σ_L and σ_R (the results can be extended to all communication-proof strategies).

In Appendix D.1 we show that strategies σ_L and σ_R are neutrally stable strategies (NSS) in the sense of Maynard Smith and Price (1973), and evolutionarily stable strategies (ESS) if $|M| = 2$. This implies that σ_L and σ_R are robust to the presence of a small proportion of experimenting agents who behave differently than the rest of the population.

We are not quite satisfied with this result for three reasons. First, neutral stability is not the strongest form of evolutionary stability, although in games with cheap talk it is typically the strongest form of stability one can expect owing to the freedom that unused messages provide mutants; see, e.g., Banerjee and Weibull (2000).[14] Second, our game, owing to the incomplete information modeled here as a continuum variable, has a continuum of strategies, especially in the action phase after messages are observed. But with a continuum of strategies the notion of even an ESS is not sufficient to imply local convergence to the equilibrium from nearby states. The reason for this, see e.g., Oechssler and Riedel (2002), is that ESS for continuum models considers only the possibly large strategy deviation of a small proportion of individuals and not the small strategy deviation of possibly a large proportion of individuals. Lastly, our cheap-talk game is a two-stage game and, hence, an extensive-form game. It is well known that extensive-form games do not admit ESSs of the entire game (unless they are strict equilibria; see Selten, 1980), and hence it seems reasonable to explore the stability of equilibrium behavior in each stage separately.

To address these issues we investigate two additional evolutionary stability properties. We inves-

[14]We here do not consider set-valued concepts of evolutionarily stability such as *evolutionary stable sets* (Swinkels, 1992 and the related analysis in, e.g., Balkenborg and Schlag, 2001, 2007), nor the perturbation-based concept of *limit-ESS* (Selten, 1983; Heller, 2014), which lies in between ESS and NSS.

tigate the evolutionary stability in each induced game that is on the equilibrium path.[15] Evolution will not necessarily place huge restrictions on play in unreached induced games (see, e.g., Nachbar, 1990; Gale et al., 1995), but should do so for induced games reached with positive probability. To address this, we study the evolutionary stability properties of strategies σ_L and σ_R at the message level (taking as given the action functions in the second stage), and at the action level (after messages that are observed with positive probability). In Appendix D.2 we show that, when the action is fixed to be either ξ_L or ξ_R, the message function μ^* used by σ_L and σ_R is weakly dominant. This is a substantially stronger property than the message function being an NSS.[16]

In Appendix D.3 we investigate the evolutionary stability properties of the action choice induced by σ_L and σ_R after the players observe the message pair. As choosing an action as a function of a player's type is equivalent to choosing a cutoff from a continuum (the unit interval), we employ a stability concept designed for such cases. The issue is further complicated by the fact that, owing to the asymmetry after unequal messages, we need to employ a multidimensional stability concept. The literature provides one in the form of a neighborhood invader strategy developed for the double-population case by Cressman (2010), building on earlier work by Eshel and Motro (1981) and Apaloo (1997), among others. We show that the action choice induced by σ_L and σ_R indeed constitutes a neighborhood invader strategy after each pair of possible messages.

Remark 2 (Stability of inefficient equilibria without communication). Our analysis shows that in our setup with communication evolutionary stability lead to efficient outcome. This is not true without communication. A plausible equilibrium refinement in setups without communication is robustness to small perturbations in the behavior of the population (e.g., requiring Lyapunov stability of the best-reply dynamics, or continuous stability à la Eshel, 1983). Adapting the analysis of Sandholm (2007) to the current setup implies that an equilibrium is robust in this sense if and only if the density of the distribution of types at the relevant threshold x (with $x = F(x)$) is less than one. In particular, if the distribution of types satisfies $f(0), f(1) > 1$, then there exists $x \in (0,1)$ satisfying $x = F(x)$ and $f(x) < 1$. The corresponding equilibrium, which entails inefficient miscoordination, is then robust to small perturbations. Thus, coordination games without communication are likely to induce substantial miscoordination if the density of extreme types is high (i.e., if $f(0), f(1) > 1$).

[15]Our communication-proofness concept does not impose any restrictions on unreached induced games other than that the strategy in the whole game must be an equilibrium strategy: we do not require play in unreached induced games to be a Nash equilibrium nor do we require communication-proofness in unreached induced games. Therefore, we here do not demand evolutionary stability in unreached induced games.

[16]For dynamic evolutionary processes weakly dominated strategies are not always eliminated. See, e.g., Weibull (1995), Hart (2002), Kuzmics (2004), Kuzmics (2011), Laraki and Mertikopoulos (2013), Bernergård and Mohlin (2019) for a discussion of this issue. Note also that Kohlberg and Mertens (1986) made it a desideratum that a concept of *strategic* stability should not include weakly dominated strategies.

8 Extensions

Next, we present informally two extensions that affect some (but not all) of our results.

Coordination games with more than 2 actions In Appendix E.5 we analyze coordination games with more than two actions. In this setup we are able to prove somewhat weaker variants of our main result. First, we show that σ_C remains a strongly communication-proof equilibrium strategy in this more general setup (by contrast, strategies σ_L and σ_R might not be equilibria in this setup). While we do not have a full characterization of the set of communication-proof equilibrium strategies, we are able to show that strongly communication-proof equilibrium strategies must satisfy mutual-preference consistency and coordination whenever both players have sent the same message.

Extreme types with dominant actions Appendix E.6 extends our analysis to a setup in which some types find that one of the actions is a dominant action for them. We show that in the presence of these extreme types there exists an essentially unique strongly communication-proof equilibrium strategy that satisfies the three key properties, where the left tendency of this strategy is equal to the share of extreme types whose dominant action is L. In this setup moderate "leftists," i.e., types $u \in (0, 1/2)$, gain if there are more extreme "leftists" than extreme "rightists," in the sense that the above essentially unique communication-proof strategy induces a higher probability of coordination on action L when two agents with different preferred outcomes meet.

9 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.[17]

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. [18]

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

[17] 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).

[18] 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).

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.

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 (see also Remark 2).[19] 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.

References

Apaloo, J. (1997). Revisiting strategic models of evolution: The concept of neighborhood invader strategies. *Theoretical Population Biology 52*(1), 52–71.

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.

Balkenborg, D. and K. H. Schlag (2001). Evolutionarily stable sets. *International Journal of Game Theory 29*(4), 571–595.

Balkenborg, D. and K. H. Schlag (2007). On the evolutionary selection of sets of nash equilibria. *Journal of Economic Theory 133*(1), 295–315.

Banerjee, A. and J. W. Weibull (2000). Neutrally stable outcomes in cheap-talk coordination games. *Games and Economic Behavior 32*, 1–24.

[19] See also Neary and Newton (2017) who study coordination games without communication played on a graph, and provide sufficient conditions for heterogeneous equilibria with miscoordination to be stable.

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

Bernergård, A. and E. Mohlin (2019). Evolutionary selection against iteratively weakly dominated strategies. *Games and Economic Behavior 117*, 82–97.

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. and U. Gneezy (2010). Cognitive forward induction and coordination without common knowledge: An experimental study. *Games and Economic Behavior 68*(2), 488–511.

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: Journal of the Econometric Society*, 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.

Cressman, R. (2010). Css, nis and dynamic stability for two-species behavioral models with continuous trait spaces. *Journal of Theoretical Biology 262*(1), 80–89.

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.

Eshel, I. (1983). Evolutionary and continuous stability. *Journal of Theoretical Biology 103*(1), 99–111.

Eshel, I. and U. Motro (1981). Kin selection and strong evolutionary stability of mutual help. *Theoretical Population Biology 19*(3), 420–433.

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

Farrell, J. and P. Klemperer (2007). Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization 3*, 1967–2072.

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.

Gale, J., K. G. Binmore, and L. Samuelson (1995). Learning to be imperfect: The ultimatum game. *Games and Economic Behavior 8*(1), 56–90.

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.

Hart, S. (2002). Evolutionary dynamics and backward induction. *Games and Economic Behaviour 41*, 227–264.

Heller, Y. (2010). Minority-proof cheap-talk protocol. *Games and Economic Behavior 69*(2), 394–400.

Heller, Y. (2014). Stability and trembles in extensive-form games. *Games and Economic Behavior 84*, 132–136.

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: Journal of the Econometric Society*, 29–56.

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

Kohlberg, E. and J.-F. Mertens (1986). On the strategic stability of equilibria. *Econometrica 54*(5), 1003–1037.

Kuzmics, C. (2004). Stochastic evolutionary stability in extensive form games of perfect information. *Games and Economic Behavior 83*, 2, 321–36.

Kuzmics, C. (2011). On the elimination of dominated strategies in stochastic models of evolution with large populations. *Games and Economic Behavior 72*(2), 452–466.

Laraki, R. and P. Mertikopoulos (2013). Higher order game dynamics. *Journal of Economic Theory 148*(6), 2666–2695.

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.

Mehta, J., C. Starmer, and R. Sugden (1994). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review 84*(3), 658–673.

Nachbar, J. H. (1990). Evolutionary selection dynamics in games: Convergence and limit properties. *International Journal of Game Theory 19*(1), 59–89.

Nash, J. (1950). Non-cooperative games. *Ph. D. dissertation, Princeton University*.

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.

Oechssler, J. and F. Riedel (2002). On the dynamic foundation of evolutionary stability in continuous models. *Journal of Economic Theory 107*(2), 223–252.

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.

Sandholm, W. H. (2010). Local stability under evolutionary game dynamics. *Theoretical Economics 5*(1), 27–50.

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.

Selten, R. (1980). A note on evolutionarily stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology 84*(1), 93–101.

Selten, R. (1983). Evolutionary stability in extensive two-person games. *Mathematical Social Sciences 5*(3), 269–363.

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.

Van Huyck, J. B., R. C. Battalio, and R. O. Beil (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review 80*, 234–248.

von Neumann, J. and O. Morgenstern (1944). *Theory of games and economic behavior*. Princeton university press, NJ.

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.

Weibull, J. W. (1995). *Evolutionary Game Theory*. Cambridge, Massachusetts: MIT Press.

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

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

Online Appendices

A Formal Proofs

A.1 Undominated Action Strategies

In this subsection we show that our restriction to threshold action functions is without loss of generality, in the sense that each generalized action function 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 in this game 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 $\xi : U \times M \times M \to \triangle\{L,R\}$ (see Remark 2.1), 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., *equivalent*), which we denote by $\eta \approx \tilde{\eta}$, if they induce the same behavior with probability one, i.e., if

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

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 η'.

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 η is *strictly dominated* by generalized strategy $\tilde{\eta}$ if $\pi(\eta,\eta') < \pi(\tilde{\eta},\eta')$ for any generalized strategy η' of the opponent.

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

Proof. If $\mathbb{E}_{u\sim F}[\eta_u(L)] = 1$ (resp., $\mathbb{E}_{u\sim F}[\eta_u(L)] = 0$), then η is equivalent to the cutoff strategy of always playing L (resp., R). Thus, suppose that $\mathbb{E}_{u\sim F}[\eta_u(L)] \in (0,1)$. Let $x \in (0,1)$ be such that $F(x) = \mathbb{E}_{u\sim F}[\eta_u(L)] = \int_u \eta_u(L) f(u)du$. Let $\tilde{\eta}$ then be the cutoff strategy with cutoff x, i.e.,

$$\tilde{\eta}_u(L) = \begin{cases} 1 & u \leq x \\ 0 & u > x. \end{cases}$$

Assume that η and $\tilde{\eta}$ are not equivalent, i.e., $\eta \not\approx \tilde{\eta}$. Let η' be an arbitrary generalized strategy of the opponent. By construction, strategies η and $\tilde{\eta}$ induce the same average probability of choosing L. Strategies $\tilde{\eta}$ and η 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., $\eta_u(L) \le \tilde{\eta}_u(L)$ for any type $u \le x$ and $\eta_u(L) \ge \tilde{\eta}_u(L)$ for any type $u > x$. Since $\eta \not\approx \tilde{\eta}$ and $\mathbb{E}_{u\sim F}[\eta_u(L)] \in (0,1)$, it follows that the inequalities are strict for a positive measure of types, i.e.,

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

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 $\pi(\eta,\eta') < \pi(\tilde{\eta},\eta')$. $\square$

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 σ 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 $m \in \text{supp}(\bar{\mu})$. Consider any pair $m,m' \in \text{supp}(\bar{\mu})$. There are three cases to consider. Suppose first that both $\text{supp}(F_m), \text{supp}(F_{m'}) \subseteq [0,1/2]$. Then as σ is mutual-preference consistent we have that $\xi(m,m') = \xi(m',m) = L$. Thus ξ 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 σ is *coordinated* we have that $\xi(m,m') = \xi(m',m) = L$ or $\xi(m,m') = \xi(m',m) = R$. Action function ξ, 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 σ' in the game $\langle \Gamma(F_m,F_{m'}), M\rangle$ Pareto dominates σ after this message pair. This shows that action function ξ is a best response to μ and to itself given μ and that, moreover, it cannot be CP trumped. It remains to show that the message function μ 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 σ 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 $m' \notin \text{supp}(\mu_u)$. 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 $m, m' \in \operatorname{supp}(\bar{\mu})$,

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

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

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

As σ is an equilibrium, we have $\inf\{u \mid \mu_u(m') > 0\} < \xi(m',m) < \sup\{u \mid \mu_u(m') > 0\}$ because otherwise the sender of m' would play L with probability one or R with probability one, in which case the 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 σ_L, σ_R, or σ_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 σ_R. Consider the player who sent message m.

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

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

where the left-hand side is type a u agent's payoff under strategy profile σ_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

$$(1-u)F_{m'}(\tfrac{1}{2}) + u\left(1 - F_{m'}(\tfrac{1}{2})\right) > u\left(1 - F_{m'}(x')\right),$$

where the left-hand side is a type u agent's payoff under strategy profile σ_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 σ_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 σ_R with σ_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 σ_C. To see this, consider the player who sent message m.

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

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

where the left-hand side is a type u agent's payoff under strategy profile σ_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

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

where the left-hand side is a type u agent's payoff under strategy profile σ_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 $(1-u)\left[F_{m'}(\tfrac{1}{2})+\tfrac{1}{2}\left(1-F_{m'}(\tfrac{1}{2})\right)\right]+u\tfrac{1}{2}\left(1-F_{m'}(\tfrac{1}{2})\right) \geq u$.

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

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

where the left-hand side is a type u agent's payoff under strategy profile σ_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. □

Lemma 3. *Every weakly communication-proof equilibrium strategy σ has binary communication.*

Proof. Let σ be a communication-proof equilibrium strategy. Recall that

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

As σ is coordinated by Lemma 2, the payoff to a type u from sending message $m \in supp(\overline{\mu})$ 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 $\overline{\beta}^{\sigma} \in [0,1]$ such that for all $u < 1/2$ and all $m \in \mathrm{supp}(\mu_u)$, we have $\beta^{\sigma}(m) = \overline{\beta}^{\sigma}$. Analogously, we must have a $\underline{\beta}^{\sigma} \in [0,1]$ such that for all $u > 1/2$ and all $m \in \mathrm{supp}(\mu_u)$, we have $\alpha^{\sigma}(m) = \overline{\beta}^{\sigma}$. Clearly also $\underline{\beta}^{\sigma} \leq \overline{\beta}^{\sigma}$. To extend the argument to unused messages $m \notin supp(\overline{\mu})$ we rely on the full support assumption. Assume to the contrary that there is a message $m \notin supp(\overline{\mu})$ with $\beta^{\sigma}(m) > \overline{\beta}^{\sigma}$ (resp., $\beta^{\sigma}(m) < \underline{\beta}^{\sigma}$). 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 σ is an equilibrium strategy. $\square$

Lemma 4. *Every weakly communication-proof equilibrium strategy σ is mutual-preference consistent.*

Proof. By Lemma 2 a communication-proof equilibrium strategy $\sigma = (\mu, \xi)$ is coordinated. Suppose that it is not mutual-preference consistent. Then there is either a message 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 message 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 σ is CP-trumped by strategy σ_R (resp., σ_L) in the game $\langle \Gamma(F_m, F_{m'}), \{m_L, m_R\} \rangle$ because strategy σ_R (resp., σ_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., $u > 1/2$). $\square$

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 strategy σ's equilibrium payoff is determined by its left tendency $\alpha \equiv \alpha^{\sigma} \in [0,1]$. This equilibrium payoff is given by

$$\pi_u(\sigma, \sigma) = (1-u)\left[F(\tfrac{1}{2}) + \alpha\left(1 - F(\tfrac{1}{2})\right)\right] + u(1-\alpha)\left[1 - F(\tfrac{1}{2})\right],$$

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

$$\pi_u(\sigma, \sigma) = (1-u)\alpha F(\tfrac{1}{2}) + u\left[\left(1 - F(\tfrac{1}{2})\right) + F(\tfrac{1}{2})(1-\alpha)\right].$$

for each type $u \in (1/2, 1]$. The payoff to a type u from a given social choice function ϕ 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 ϕ interim Pareto dominates σ. 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:

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

$$\mathbb{E}_v\phi_{u,v}(R,R) \geq \left(1 - F(\tfrac{1}{2})\right) + F(\tfrac{1}{2})(1-\alpha) \text{ for any } u > 1/2, \tag{2}$$

with at least one of the inequalities holding strictly for a positive measure of types. We can write

$$\mathbb{E}_v \phi_{u,v}(L,L) = F(\tfrac{1}{2})\mathbb{E}_{\{v\le 1/2\}}\phi_{u,v}(L,L) + \left(1-F(\tfrac{1}{2})\right)\mathbb{E}_{\{v>1/2\}}\phi_{u,v}(L,L),$$

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

$$F(\tfrac{1}{2})\mathbb{E}_{\{v\le 1/2\}}\phi_{u,v}(L,L) + \left(1-F(\tfrac{1}{2})\right)\mathbb{E}_{\{v>1/2\}}\phi_{u,v}(L,L) \ge F(\tfrac{1}{2}) + \alpha\left(1-F(\tfrac{1}{2})\right)$$

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

This implies that

$$\mathbb{E}_{\{u<1/2\}}\mathbb{E}_{\{v>1/2\}}\phi_{u,v}(L,L) \ge \alpha \text{ and } \mathbb{E}_{\{u>1/2\}}\mathbb{E}_{\{v<1/2\}}\phi_{u,v}(R,R) \ge 1-\alpha,$$

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

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

which contradicts $\phi_{u,v}$ being a social choice function. □

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

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

Proof. Let $\sigma = (\mu, \xi) \in \mathscr{E}$ 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:

$$p_m = \sum_{m' \in M} \mu\left(\bar{m}'\right) 1_{\{\xi(m,m')=L\}}.$$

As σ 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, $\underline{p} \le \bar{p}$. As σ 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 $u \le 1/2$ 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 $\pi_u(\sigma,\sigma) = \underline{p}(1-u) + \left(1-\underline{p}\right)u$. 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 σ 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 σ 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 σ' be a communication-proof strategy with left tendency α. Then $\underline{p} \geq \alpha F(1/2)$ and by construction σ is either interim (pre-communication) payoff equivalent to or Pareto dominated by σ'. □

Proof of Proposition 2. By Lemma 5 we have that every coordinated equilibrium strategy σ is interim (pre-communication) Pareto-dominated by some communication-proof strategy with some left tendency $\alpha \in [0,1]$ denoted by σ_α. 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 σ_α is given by

$$\pi_u(\sigma_\alpha,\sigma_\alpha) = (1-u)\left[F(\tfrac{1}{2}) + \alpha\left(1-F(\tfrac{1}{2})\right)\right] + u(1-\alpha)\left(1-F(\tfrac{1}{2})\right) \text{ for } u \leq 1/2 \text{ and}$$

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

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

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 ξ 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 $u \leq 1/4$ 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 ξ be such that $\xi(m_L, m_L) = L$, $\xi(m_R, m_R) = R$, $\xi(m_L, m_R) = 1/4$, and $\xi(m_R, m_L) = 3/4$. 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

$$(F(3/4) - F(1/2))\,/\,(1 - F(1/2)) = 1/4 textrmand F(1/4)/F(1/2) = 3/4.$$

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., $\text{supp}(\bar{\mu}) = M_L \dot{\bigcup} M_R$) 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 σ 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 Non-binary Communication Equilibrium

We here formally present the example, which is discussed informally at the end of Section 3, of an equilibrium in which agents reveal some information about the cardinality of their preferences.

Suppose that $|M| \geq 4$ and consider the game with two rounds of communication. Let $e, m_L, m_R \in M$ and let $\sigma = (\mu_1, \mu_2, \xi)$, with $\mu_1 : U \to \Delta(M)$, $\mu_2 : M \times M \times U \to \Delta(M)$, and $\xi : (M \times M)^2 \to U$ be as follows. For the first round of messages there is an $x \in [0,1]$ such that

$$\mu_1(u) = \begin{cases} e & \text{if } u \leq x \text{ or } u > 1-x \\ m_L & \text{if } x < u \leq \frac{1}{2} \\ m_R & \text{if } \frac{1}{2} < u \leq 1-x. \end{cases}$$

The second round of messages depends on the outcome of the first round and is best described in the following table.

	e	m_L	m_R
e	e	μ^*	μ^*
m_L	m_L	m_L	μ_C
m_R	m_R	μ_C	m_R

Each entry in this table describes the message function that a player follows if her first-stage message is the one indicated on the left and her opponent's first-stage message is the one indicated at the top. The message function μ^* (after for instance a message pair of (e, m_L)) is just as in the definition of σ_L and σ_R (in Section 3). The message function μ_C is as in the definition of σ_C with an appropriate relabeling of four messages in M. The action function is also best given in table form as a function of the result of the first round of communication (or the second round when so indicated).

	e	m_L	m_R
e	$\begin{cases} L & \text{if } u \leq \frac{1}{2} \\ R & \text{if } u > \frac{1}{2} \end{cases}$	$\begin{cases} L & \text{if } u \leq \frac{1}{2} \\ R & \text{if } u > \frac{1}{2} \end{cases}$	$\begin{cases} L & \text{if } u \leq \frac{1}{2} \\ R & \text{if } u > \frac{1}{2} \end{cases}$
m_L	$\begin{cases} L & \text{if } \mu_2 = (m_L, m_L) \\ R & \text{otherwise} \end{cases}$	L	ξ_C
m_R	$\begin{cases} R & \text{if } \mu_2 = (m_R, m_R) \\ L & \text{otherwise} \end{cases}$	ξ_C	R

Action function ξ_C is as defined for σ_C applied to the second round of communication only.

We can complete the description of this strategy by requiring that all other messages in M be treated exactly the same as one of the messages e, m_L, m_R.

Proposition 3. *Let F be a nondegenerate symmetric distribution around $1/2$, i.e., $F(x) = 1 - F(1-x)$*

for all $x \in [0,1]$. Then there is an $x \in (0, 1/2)$ such that the above-defined strategy of the coordination game, with two rounds of communication and with $|M| \geq 4$, is a Nash equilibrium.

Proof. Consider the given strategy for an arbitrary $x \in (0, 1/2)$. First note that whenever messages lead the players to coordinate their action then clearly both players are best replying to each other with their actions. This is so in all cases except when both players send message e in the first round. In this case players choose L if their type $u \leq 1/2$ and R otherwise. Each player, in this case, faces an opponent that either has $u \leq x$ or $u > 1-x$. In the first (second) case the opponent plays L (R). Given that $F(x) = 1 - F(1-x)$ both cases are equally likely. Given this, players' actions are indeed best replies.

Thus, all action choices are best responses to the given strategy. We now turn to message choices. Consider the second round. After moderate messages in the first round messages in the second round either do not affect play at all (after (m_L, m_L) and (m_R, m_R)) or do so as in strategy σ_C. In either case players are indifferent between all messages. After message pairs (m_L, e) and (m_R, e) the sender of the moderate message has a strict incentive to send the same message again, while the sender of the extreme message has a strict incentive to send m_L if her type $u < 1/2$ or to send m_R if her type $u > 1/2$ as this induces coordination on her preferred outcome. After both players send message e, play will not depend on messages in the second round either and so both players will be indifferent between all messages. Thus, the behavior in the second round of communication is a best response to the given strategy.

Finally, we need to consider the incentives to send messages in the first round. It is obvious that any type $u < 1/2$ prefers sending message m_L to sending message m_R and vice versa for types $u > 1/2$. The only remaining thing to show is that types $u \leq x$ and $u > 1-x$ and only these weakly prefer to send message e in the first round. Given the symmetry it is without loss of generality to consider a type $u \leq 1/2$. Given the strategy, sending message e yields to this type a payoff of

$$F(x)(1-u) + \left(F(\tfrac{1}{2}) - F(x)\right)(1-u) + \left(F(1-x) - F(\tfrac{1}{2})\right)(1-u) + (1 - F(1-x))0,$$

where $F(x)$ is the probability that her opponent is an extreme left type, $(F(1/2) - F(x))$ is the probability that the opponent is a moderate left type, $(F(1-x) - F(1/2))$ is the probability that the opponent is a moderate right type, in all of which cases both players eventually play L, and where $(1 - F(1-x))$ is the probability that her opponent is an extreme right type, in which case the two players miscoordinate. Sending message m_L yields a payoff of

$$F(x)(1-u) + \left(F(\tfrac{1}{2}) - F(x)\right)(1-u) + \left(F(1-x) - F(\tfrac{1}{2})\right)\tfrac{1}{2} + (1 - F(1-x))u.$$

A type $u \leq 1/2$ therefore weakly prefers sending message e to sending message m_L if and only if

$$D_x(u) \equiv \left(F(1-x) - F(\tfrac{1}{2})\right)(1-u) - \left(F(1-x) - F(\tfrac{1}{2})\right)\tfrac{1}{2} - (1 - F(1-x))u \geq 0.$$

Using the symmetry of F we can rewrite $D(x)$ as $D_x(u) = (1/2 - F(x))(1/2 - u) - F(x)u$.

Note that $D_x(u)$ is linear and downward sloping in u if $x \in (0, 1/2)$. In an equilibrium we then must have that $D_x(x) = 0$. This implies $D_x(x) = \left(\frac{1}{2} - F(x)\right)\left(\frac{1}{2} - x\right) - F(x)x = 0$, or, equivalently, $D_x(x) = \frac{1}{4} - \frac{1}{2}F(x) - \frac{1}{2}x = 0$. As $D_0(0) = 1/4 > 0$, $D_{1/2}(1/2) = -1/4 < 0$, and $D_x(x)$ is a continuous function in x, there is an $x \in (0, 1/2)$ such that $D_x(x) = 0$. For this x the given strategy is thus an equilibrium. □

D Evolutionary Stability Analysis

In this appendix we analyze the stability of strategies σ_L and σ_R (the results can be extended to other communication-proof equilibria, but we omit the details here for brevity).

D.1 Evolutionary/Neutral Stability

We say that two strategies are almost surely realization equivalent (abbr., equivalent) if they induce the same behavior in almost all types (regardless of the opponent's behavior).

Definition 5. A condition holds for *almost all types* if the set of types that satisfy the condition $\tilde{U} \subseteq U$ has mass one (as measured by the distribution f), i.e., $\int_{u \in \tilde{U}} f(u)\,du = 1$.

Definition 6. Strategies $\sigma = (\mu, \xi)$ and $\tilde{\sigma} = \left(\tilde{\mu}, \tilde{\xi}\right)$ are *almost surely realization equivalent* (abbr., *equivalent*) if for almost all types $u \in [0,1]$: $\mu_u(m) = \tilde{\mu}_u(m)$ for every message $m \in M$, and $F_m(\xi(m,m')) = F_m\left(\tilde{\xi}(m,m')\right)$ for all messages $m, m' \in \text{supp}(\bar{\mu})$.

If σ and $\tilde{\sigma}$ are equivalent strategies we denote this by $\sigma \approx \tilde{\sigma}$. It is immediate that equivalent strategies always obtain the same ex-ante expected payoff.

An equilibrium strategy σ is neutrally (evolutionarily) stable if it achieves a weakly (strictly) higher ex-ante expected payoff against any (non-equivalent) best-reply strategy, relative to the payoff that the best-reply strategy achieves against itself.

Definition 7 (adaptation of Maynard Smith and Price, 1973)**.** Equilibrium strategy $\sigma \in \mathscr{E}$ is neutrally stable if $\pi(\tilde{\sigma}, \sigma) = \pi(\sigma, \sigma) \Rightarrow \pi(\sigma, \tilde{\sigma}) \geq \pi(\tilde{\sigma}, \tilde{\sigma})$ for any nonequivalent strategy $\tilde{\sigma} \not\approx \sigma$. It is evolutionarily stable if this last weak inequality is replaced by a strict one.

The refinement of neutral stability is arguably a necessary requirement for an equilibrium to be a stable convention in a population (see, e.g., Banerjee and Weibull, 2000). If σ is an equilibrium strategy that is not neutrally stable, then a few experimenting agents who play a best-reply strategy σ' can invade the population. These experimenting agents would fare the same against the incumbents, whereas they would outperform the incumbents when being matched with other experimenting agents. This implies that, on average, these experimenting agents would be more successful than the incumbents, and their frequency in the population would increase in any payoff-monotone learning dynamics. This, in turn, implies that the population will move away from σ.

Our first result shows that both σ_L and σ_R are neutrally stable, and, moreover, they are evolutionarily stable if there are two feasible messages.

Proposition 4. *Strategies σ_L and σ_R are neutrally stable strategies of the coordination game with communication $\langle \Gamma, M \rangle$. Moreover, if $|M| = 2$, then σ_L and σ_R are evolutionarily stable strategies.*

Proof. We here prove this result for σ_L. The proof for σ_R proceeds analogously and is omitted. In order to prove this result we first characterize all strategies σ that are best responses to σ_L and thus satisfy $\pi(\sigma, \sigma_L) = \pi(\sigma_L, \sigma_L)$. Consider the message and action choice of a type u when her opponent uses strategy σ_L. If our type u chooses any message other than m_R, her opponent, sending message m_L or m_R, plays action L in either case. Our type u could then choose action L (as prescribed by σ_L), which provides a payoff of $1 - u$, or action R, which provides a payoff of zero. Thus all types $u < 1$ are strictly better off choosing action L in this case. Also note that sending any message other than m_L leads to a best possible payoff of $1 - u$.

If our type u chooses to send message m_R then there are two cases. First, suppose that her opponent sends message m_L, in which case her opponent chooses action L. Our type u could then choose action L (as prescribed by σ_L), which provides a payoff of $1 - u$, or action R which provides a payoff of zero. Thus all types $u < 1$ are strictly better off choosing action L in this case. Second, suppose that her opponent sends message m_R, in which case her opponent chooses action R. Our type u could then choose action R (as prescribed by σ_L), which provides a payoff of u, or action L which provides a payoff of zero. Thus all types $u > 0$ are strictly better off choosing action R in this case. Note that sending message m_R thus provides a best possible payoff of $F(1/2)(1-u) + (1 - F(1/2))\,u$.

For type u it is then a strict best response to send message m_R if $F(1/2)(1-u) + (1-F(1/2))\,u > 1 - u$, which is the case if and only if $u > 1/2$ (as $F(1/2) \in (0,1)$ by assumption). For the case of $|M| = 2$ we then have that any best response to σ_L is equivalent to σ_L as only three possible types have an alternative best reply: types $u = 0$, $u = 1/2$, and $u = 1$ (all zero measures under the assumption of an atomless distribution F). Any strategy that differs from σ_L for a positive measure of types yields a strictly inferior payoff against σ_L than σ_L does. This proves that σ_L is evolutionarily stable in the case of $|M| = 2$ simply by virtue of the fact that there are no nonequivalent strategies σ that satisfy $\pi(\sigma, \sigma_L) = \pi(\sigma_L, \sigma_L)$.

Suppose from now on that $|M| > 2$. Our type u then has a choice of messages $m \neq m_R$ when $u < 1/2$. All of these messages can at best lead to a payoff of u (from playing L,L) and, therefore, all of them are equally good when playing against σ_L. As her opponent never chooses any message other than m_L or m_R (each has probability zero under σ_L) our type $u < 1/2$ when best responding can play anything after any message pair (m,m') when both $m,m' \notin \{m_L,m_R\}$. Let σ be a strategy that satisfies all the previous restrictions, where all types u play a (in most cases unique and strict) best response against σ_L. Then we have that $\pi_u(\sigma,\sigma) = \pi_u(\sigma_L,\sigma)$ for all $u \geq 1/2$ (as the behavior under σ for types $u \geq 1/2$ (except for possibly types $u = 1/2$ and $u = 1$) is identical to that under σ_L), $\pi_u(\sigma,\sigma) \leq 1-u$ for all $u < 1/2$ (since this type can achieve at best $1-u$), and $\pi_u(\sigma_L,\sigma) = 1-u$ (since σ similarly to σ_L prescribes playing L in this case). We thus have for any such σ by construction $\pi(\sigma,\sigma_L) = \pi(\sigma_L,\sigma_L)$. We also have that $\pi(\sigma_L,\sigma) \geq \pi(\sigma,\sigma_L)$ for any such σ. Finally any best-reply strategy to strategy σ_L must be equivalent to some such strategy σ and thus σ_L is neutrally stable. □

D.2 Message Function is Dominant

Next we show that the first-stage behavior induced by strategy σ_L (resp., σ_R), namely, the message function μ^*, is weakly dominant (and strictly dominant when $|M| = 2$), when taking as given that the behavior in the second stage is according to the action function ξ_L (resp., ξ_R). This suggests that the behavior in the first stage that is induced by σ_L (resp., by σ_R) is robust to any perturbation that keeps the behavior in the second stage unchanged. Specifically, it implies that even if the message function used by the population is perturbed in an arbitrary (and possibly significant) way, then the original function μ^* yields a weakly higher payoff than any other message function, which suggests that the behavior in the first stage would converge back to play μ^* under any payoff-monotone learning dynamics.

Proposition 5 shows that message function μ^* yields a weakly higher payoff relative to any other message function when the action function is given by ξ_L or ξ_R. Moreover, the inequality is strict whenever the alternative message function is essentially different from μ^* in the sense of inducing low types to play m_R or inducing high types to play $m \neq m_R$.

Proposition 5. *Let μ',μ'' be arbitrary message functions. Then for, $\xi \in \{\xi_L,\xi_R\}$ and for any type $u \neq 1/2$,*

$$\pi_u\left(\left(\mu^*,\xi\right),\left(\mu',\xi\right)\right) \geq \pi_u\left(\left(\mu'',\xi\right),\left(\mu',\xi\right)\right).$$

This inequality is strict for $\xi = \xi_L$ if $\mu'_u(m_R) > 0$ for a positive measure of types u and, either $\mu''_u(m_R) > 0$ for a positive measure of types $u < 1/2$, or $\mu''_u(m_R) < 1$ for a positive measure of types $u > 1/2$. This inequality is strict for $\xi = \xi_R$ if $\mu'_u(m_L) > 0$ for a positive measure of types u and, either $\mu''_u(m_L) > 0$ for a positive measure of types $u > 1/2$, or $\mu''_u(m_L) < 1$ for a positive measure of types $u < 1/2$.

Proof. Consider the case of $\xi = \xi_L$ (the other case is proven analogously). Let γ denote the probability that a player following strategy (μ', ξ_L) sends message m_R. Then sending any message other than m_R when the partner sends (μ', ξ_L) yields a payoff of $1-u$, and sending message m_R yields a payoff of $\gamma u + (1-\gamma)(1-u)$. Thus any type $u > 1/2$ weakly prefers sending message m_R to sending any other message (and strictly prefers this if $\gamma > 0$), while any type $u < 1/2$ weakly prefers sending any message other than m_R to sending message m_R (and strictly prefers this if $\gamma < 1$). Thus, for every message function μ' of the opponent, μ^* optimizes the message choice for every type u universally. □

D.3 Action Function is a Neighborhood Invader Strategy

In the induced second-stage game $\Gamma(F_m, F_{m'})$ (the game played after players observe a pair of messages (m, m')), players choose a cutoff to determine whether to play action L (if their type is below or equal to that cutoff) or action R (otherwise). Thus, players essentially choose a number (their cutoff) from the unit interval. Note also that this induced game is asymmetric whenever the message profile is asymmetric, i.e., when $m \neq m'$. As argued by Eshel and Motro (1981) and Eshel (1983), when the set of strategies is a continuum, a stable convention should be robust to perturbations that slightly change the strategy played by all agents in the population. Cressman (2010) formalizes this requirement using the notion of neighborhood invader strategy (adapting the related notion of Apaloo, 1997). In what follows we show that the action function induced by σ_L and σ_R is a neighborhood invader strategy in any induced game $\Gamma(F_m, F_{m'})$ on the path of play.

Fix a message function μ and a pair of messages $m_1, m_2 \in \text{supp}(\bar{\mu})$. We identify a strategy in the induced game $\Gamma(F_{m_1}, F_{m_2})$ with thresholds x_i, which is interpreted as the maximal type for which player $i \in \{1,2\}$ plays L. We say that strategy x_i of player i is equivalent to x_i' (denoted by $x_i \approx x_i'$) in the induced game $\Gamma(F_{m_1}, F_{m_2})$, if $F_{m_i}(x_i) = F_{m_i}(x_i')$, which implies that both thresholds induce the same behavior with probability one. Let $\pi^{m_1,m_2}(x_1, x_2)$ denote the expected payoff of an agent with a random type sampled from f_{m_1} who uses threshold x_1 when facing a partner with a random unknown type sampled from f_{m_2} who uses threshold x_2.

A strategy profile (x_1, x_2) is a strict equilibrium if any best reply to x_j is equivalent to x_i, i.e., $\pi^{m_1,m_2}(x_1', x_2) \geq \pi^{m_1,m_2}(x_1, x_2) \Rightarrow x_1' \approx x_1$, and $\pi^{m_2,m_1}(x_2', x_1) \geq \pi^{m_2,m_1}(x_2, x_1) \Rightarrow x_2' \approx x_2$.

We say that the strict equilibrium (x_1, x_2) is a neighborhood invader strategy in the induced game $\Gamma(F_{m_1}, F_{m_2})$ if the population converges to (x_1, x_2) from any nonequivalent nearby strategy profile (x_1', x_2') in two steps: (1) strategy x_i yields a strictly higher payoff against x_j relative to the payoff of x_i' against x_j (which implies convergence from $\left(x_i', x_j'\right)$ to $\left(x_i, x_j'\right)$), and (2) due to (x_1, x_2) being a strict equilibrium, strategy x_j yields a strictly higher payoff against x_i relative to the payoff of x_j' against x_i (which implies the convergence from $\left(x_i, x_j'\right)$ to (x_i, x_j)).

Definition 8 (Adaptation of Cressman, 2010, Def. 5)**.** Fix a message function μ and a pair of messages $m_1, m_2 \in \text{supp}(\bar{\mu})$. A strict Nash equilibrium (x_1, x_2) is a *neighborhood invader strategy profile* in $\Gamma(F_{m_1}, F_{m_2})$ if there exists $\varepsilon > 0$, such that for each (x_1', x_2') satisfying $x_1' \not\approx x_1$, $x_2' \not\approx x_2$, $|x_1' - x_1| < \varepsilon$ and $|x_2' - x_2| < \varepsilon$, then either $\pi^{m_1,m_2}(x_1, x_2') > \pi^{m_1,m_2}(x_1', x_2')$ or $\pi^{m_2,m_1}(x_2, x_1') > \pi^{m_2,m_1}(x_2', x_1')$.

Proposition 6 shows that the profile of action functions induced by σ_L (or, similarly, by σ_R) is a neighborhood invader strategy in any induced game.

Proposition 6. *Let $m_1, m_2 \in supp(\bar{\mu}^*)$. Then both strategy profiles $(\xi_L(m_1, m_2), \xi_L(m_2, m_1))$ and $(\xi_R(m_1, m_2), \xi_R(m_2, m_1))$ are strict equilibria and neighborhood invader strategy profiles in $\Gamma_{F_{m_1}, F_{m_2}}$.*

Proof. We present the proof for $(\xi_L(m_1, m_2), \xi_L(m_2, m_1))$ (the proof for ξ_R is analogous). Observe that $m_1, m_2 \in \text{supp}(\bar{\mu}^*)$ implies one of three cases: $m_1 = m_2 = m_L$, $m_1 = m_2 = m_R$, or $m_1 = m_R$, $m_2 = m_L$. We analyze each case as follows.

Suppose first that $m_1 = m_2 = m_L$. This implies that $\xi_L(m_1, m_2) = \xi_L(m_2, m_1) = 1$ and $F_{m_1}(1/2) = F_{m_2}(1/2) = 1$. Let $\bar{x} < 1/2$ be sufficiently close to $1/2$ such that $F_{m_1}(\bar{x}), F_{m_2}(\bar{x}) > 1/2$. Observe that $\pi^{m_1,m_2}(1, x) > \pi^{m_1,m_2}(y, x)$ for any $x > \bar{x}$ and any $y \not\approx 1$. This proves that $(\xi_L(m_1, m_2), \xi_L(m_2, m_1))$ is a strict equilibrium and a neighborhood invader strategy profile. Now suppose that $m_1 = m_2 = m_R$. This implies that $\xi_L(m_1, m_2) = \xi_L(m_2, m_1) = 0$ and $F_{m_1}(1/2) = F_{m_2}(1/2) = 0$. Let $\bar{x} > 1/2$ be sufficiently close to $1/2$ such that $F_{m_1}(\bar{x}), F_{m_2}(\bar{x}) < 1/2$. Observe that $\pi^{m_1,m_2}(0, x) > \pi^{m_1,m_2}(y, x)$ for any $x < \bar{x}$ and any $y \not\approx 0$. This proves that $(\xi_L(m_1, m_2), \xi_L(m_2, m_1))$ is a strict equilibrium and neighborhood invader.

Suppose finally that $m_1 = m_R$, $m_2 = m_L$. This implies that $\xi_L(m_1, m_2) = \xi_L(m_2, m_1) = 1$, $F_{m_1}(1/2) = 0$, and $F_{m_2}(1/2) = 1$. Observe that $\pi^{m_1,m_2}(1, 1) > \pi^{m_1,m_2}(x, 1)$ for any $x \not\approx 1$ and $\pi^{m_2,m_1}(1, 1) > \pi^{m_2,m_1}(x, 1)$ for any $x \not\approx 1$, which implies that $(\xi_L(m_1, m_2), \xi_L(m_2, m_1))$ is a strict equilibrium. Let $\bar{x} > 1/2$ be sufficiently close to $1/2$ such that $F_{m_1}(\bar{x}) < 1/2$. Observe that $\pi^{m_2,m_1}(1, x) > \pi^{m_1,m_2}(y, x)$ for any $x < \bar{x}$ and any $y \not\approx 1$. This proves that $(\xi_L(m_1, m_2), \xi_L(m_2, m_1))$ is a neighborhood invader strategy profile. □

D.4 Remark on Evolutionary Robustness

Oechssler and Riedel (2002) present a strong notion of stability, called evolutionary robustness, that refines both evolutionary stability and the neighborhood invader strategy. An evolutionary robust strategy σ^* is required to be robust against small perturbation in the strategy played by the population, which may comprise both (1) a few experimenting agents who follow arbitrary strategies, and (2) many agents who follow strategies that are only slightly different than σ^*. Specifically, if σ is a distribution of strategies that is sufficiently close to σ^* (in the L_1 norm induced by the weak topology), evolutionary robustness à la Oechssler and Riedel requires that $\pi(\sigma^*, \sigma) > (\sigma, \sigma)$.

One can show that σ_L and σ_R do not satisfy this condition (and, we conjecture, that no strategy can satisfy this condition in our setup). However, we conjecture that one can show that σ_L and

σ_R satisfy a somewhat weaker notion of evolutionary robustness: for each strategy distribution σ sufficiently close to σ_L (σ_R), there exists a finite sequence of strategy distributions $\sigma_1, \sigma_2, \ldots, \sigma_k$, such that $\pi(\sigma_1, \sigma) \geq (\sigma, \sigma)$, $\pi(\sigma_2, \sigma_1) \geq (\sigma_1, \sigma_1)$, $\ldots$, $\pi(\sigma_k, \sigma_{k-1}) \geq (\sigma_{k-1}, \sigma_{k-1})$, and $\pi(\sigma_L, \sigma_1) \geq (\sigma_1, \sigma_1)$ (resp., $\pi(\sigma_R, \sigma_1) \geq (\sigma_1, \sigma_1)$), with strict inequalities if $|M| = 2$ and σ is not realization equivalent to σ_L (σ_R).

E Analysis of the Extensions

E.1 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 $\mathscr{M} = \bigcup_{t=0}^{T-1} (M \times M)^{\mathrm{t}}$, where $(M \times M)^0 = \emptyset$.

A (pure) *message protocol* is a function $\mathfrak{m} : \mathscr{M} \to M$ that describes the message sent by an agent as a deterministic function of the message profiles observed in the previous rounds of communication. Let $\mathfrak{M}$ be the set of all message protocols. A strategy $\sigma = (\mu, \xi)$ is a pair where $\mu : U \to \Delta(\mathfrak{M})$ denotes the *message function*, prescribing a (possibly random) message protocol for each type, and $\xi : (M \times M)^T \to U$ 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 $\mathfrak{m} \in \mathfrak{M}$, let $\beta^\sigma(\mathfrak{m})$ denote the expected probability of a player's opponent playing L conditional on the player following message protocol $\mathfrak{m} \in \mathfrak{M}$ and the opponent following strategy $\sigma = (\mu, \xi) \in \Sigma$. We say that strategy σ has *binary communication* if there are two numbers $0 \leq \underline{\beta}^\sigma \leq \overline{\beta}^\sigma \leq 1$ such that for all message protocols $\mathfrak{m} \in \mathfrak{M}$ we have $\beta^\sigma(\mathfrak{m}) \in [\underline{\beta}^\sigma, \overline{\beta}^\sigma]$, for all message protocols $\mathfrak{m} \in \mathfrak{M}$ such that there is a type $u < 1/2$ with $\mu_u(\mathfrak{m}) > 0$ we have $\beta^\sigma(\mathfrak{m}) = \overline{\beta}^\sigma$, and for all message protocols $\mathfrak{m} \in \mathfrak{M}$ such that there is a type $u > 1/2$ with $\mu_u(\mathfrak{m}) > 0$ we have $\beta^\sigma(\mathfrak{m}) = \underline{\beta}^\sigma$. 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.

E.2 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. Note that any affine transformation of all payoffs neither changes the player's incentives nor changes how she compares any two outcome distributions $\in \Delta(\{L,R\})$. We can thus subtract $\min\{u_{RL}, u_{LR}\}$ from all payoffs and then divide all payoffs by some number such that the sum of the diagonal entries is equal to one. This leaves two parameters to describe a payoff vector in $\hat{U}$. This means that for our purposes the set $\hat{U}$ is two-dimensional. Let $\hat{\Gamma} = \hat{\Gamma}(G)$ denote the coordination game with the two-dimensional type space $\hat{U}$, endowed with an atomless CDF G over $\hat{U}$ with a density g. Similarly, let $\langle \hat{\Gamma}, M \rangle$ 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 φ_u, where φ_u is a number always between 0 and 1, prefers to play L (R) if her partner plays L with probability larger (smaller) than φ_u. In other words, for a given probability of her partner playing L, a type u prefers to play L if and only if φ_u is less than that probability. Thus, the indifference threshold φ_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 φ_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 $\varphi_u \le \xi(m, m')$.

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 (one-dimensional) main model is that for any type of player the action that she prefers to coordinate on is also risk-dominant for her.

Definition 9. 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 iff 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 σ be a strategy of a game $\langle \hat{\Gamma}, M \rangle$ that satisfies unambiguous coordination preferences. Then the following three statements are equivalent:*

1. *σ is mutual-preference consistent, coordinated, and has binary communication.*
2. *σ is a strongly communication-proof equilibrium strategy.*
3. *σ is a weakly communication-proof equilibrium strategy.*

The proof is presented in Appendix E.7.1. 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_{L_1}	L	R
L	2	0
R	0	1

u_{L_2}	L	R
L	2	-15
R	0	1

u_{R_1}	L	R
L	1	0
R	0	2

u_{R_2}	L	R
L	1	0
R	-15	2

The distribution of types F is such that[20] $P(u_{L_1}) = P(u_{R_1}) = 1/18$ and $P(u_{L_2}) = P(u_{R_2}) = 8/18$.

[20] This distribution is discrete, but could be modified to a nearby atomless distribution without changing the result.

Let $M = \{m_L, m_R\}$ and let $\sigma = (\mu, \xi)$ be such that $\mu(u_{L_1}) = \mu(u_{L_2}) = m_L$ and $\mu(u_{R_1}) = \mu(u_{R_2}) = m_R$ (making σ mutual-preference consistent), and $\xi(m_L, m_L) = L$, $\xi(m_R, m_R) = R$, $\xi(u_{L_1}, m_L, m_R) = \xi(u_{R_2}, m_R, m_L) = L$, and $\xi(u_{L_2}, m_L, m_R) = R$ as well as $\xi(u_{R_1}, m_R, m_L) = R$.

It is straightforward to verify that σ is an equilibrium with expected payoffs $\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot \left(\frac{8}{9} \cdot 2 + \frac{1}{9} \cdot 0\right) = 1 + \frac{8}{9}$ for types u_{L_1} and u_{R_1}, and $\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot \frac{1}{9} \cdot 1 = 1 + \frac{1}{18}$ for types u_{L_2} and u_{R_2}. Observe that no type wants to misreport her preferred outcome in round one. In particular, a misreporting type u_{L_2} will get a payoff of $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{8}{9} \cdot 1 = \frac{17}{18} < 1 + \frac{1}{9}$.

Proposition 7. *The equilibrium given in Example 2 is not strongly communication proof, but it is weakly communication proof.*

We prove Proposition 7 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* $L_1, L_2; R_1, R_2$	α	β
L, L	L, L	$2, 2; 1, 1$	$\in [0,1]$	$\in [0,1]$
R, R	R, R	$1, 1; 2, 2$	$\in [0,1]$	$\in [0,1]$
mix, R	mix, L	$\frac{2}{3}, \frac{2}{3}; \frac{2}{3}, \frac{2}{3}$	$\geq \frac{2}{3}$	$\geq \frac{2}{3}$
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	$\frac{16}{9}, \frac{1}{9}; \frac{2}{3}, \frac{2}{3}$	$\leq \frac{2}{3}$	$\geq \frac{1}{9}$
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	$2(1-\beta), \beta; 2(1-\alpha), \alpha$	$\in [\frac{1}{9}, \frac{2}{3}]$	$\in [\frac{1}{9}, \frac{2}{3}]$

The last two columns provide the range of α and β 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* $L_1, L_2; R_1, R_2$	α	β	$\alpha+\beta$
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	$2(1-\beta), \beta; 2(1-\alpha), \alpha$	$\in [\frac{1}{9}, \frac{7}{18}]$	$\in [\frac{1}{9}, \frac{7}{18}]$	$\leq \frac{1}{2}$

Proof. All mixed equilibria except ((L,mix),(R,mix)) are Pareto dominated by either (L,L) or (R,R), see Claim 1. Equilibrium ((L,R),(R,L)) (when α, β 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 2β 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)) with respect to (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 $\frac{16}{9}$. 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 $\frac{16}{9}$. But then R_1 can only obtain a payoff of at most $3 - \frac{16}{9} = \frac{11}{9}$ (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 $\alpha = P(L_1|m) \in [\frac{1}{9}, \frac{1}{2}]$ and $\beta = P(R_1|m) \in [\frac{1}{9}, \frac{1}{2}]$ that also satisfy $\alpha + \beta \leq \frac{1}{2}$.*

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 $\frac{1}{9}$ (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 $\alpha = P(L_1|m) \in [\frac{1}{9}, \frac{1}{2}]$ and $\beta = P(R_1|m) \in [\frac{1}{9}, \frac{1}{2}]$ that also satisfy $\alpha + \beta \leq \frac{1}{2}$. Otherwise (L,R),(R,L) is Pareto dominated, a contradiction. □

Claim 8. *Consider the stage game with $\mu = P(L_1|m) \in [\frac{1}{9}, \frac{1}{2}]$ and $\nu = P(R_1|m) \in [\frac{1}{9}, \frac{1}{2}]$ that also satisfy $\mu + \nu \leq \frac{1}{2}$. 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 $\frac{1}{3}$ and m_2 with probability $\frac{2}{3}$. Continuation play is given by the following table:

	m_1	m_2
m_1	$((L,R),(R,L))$	$((L,L),(L,L))$
m_2	$((R,R),(R,R))$	$\frac{1}{2}((L,L),(L,L)) + \frac{1}{2}((R,R),(R,R))$

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 $\frac{1}{9} \leq \alpha = P(L_1|m) = \frac{\nu}{\nu+(1-\nu)\frac{1}{3}} \leq \frac{2}{3}$ (which is true for $\nu \leq \frac{2}{5}$) as $\nu \leq \frac{1}{2} - \frac{1}{9} =$

$\frac{7}{18} < \frac{2}{5}$ and analogously $\frac{1}{9}\beta \leq \frac{2}{3}$. 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. □

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 7.

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).

E.3 More Than Two Players

Consider a variant of the coordination game in which there are $n \geq 2$ players who play a symmetric coordination game (with private values) with pre-play communication. The action set is $\{L, R\}$ for every player and the payoff to player i is equal to u_i if every player chooses action R, equal to $1-u_i$ if every player chooses L, and equal to zero otherwise. The payoff to type u_i is independent and identically drawn from some given distribution F with support in the unit interval $[0,1]$. Before players choose actions, they simultaneously send messages from a finite set of messages M and observe all these messages. Let $\langle \Gamma_n, M \rangle$ denote this n-player coordination game with pre-play communication.

In this setting the appropriate version of Theorem 1 still holds.

Theorem 3 (Theorem 1 adapted to more than two players)**.** *Let σ be a strategy of the n-player coordination game $\langle \Gamma_n, M \rangle$. Then the following three statements are equivalent:*

1. *σ is mutual-preference consistent, coordinated, and has binary communication.*

2. *σ is a strongly communication-proof equilibrium strategy.*

3. *σ is a weakly communication-proof equilibrium strategy.*

Sketch of proof; for the formal proof see Appendix E.7.2. The proof of the "$3 \Rightarrow 1$" direction has to be adapted (the proof of the "$1 \Rightarrow 2$" direction remains, essentially, the same). In this setting it is not generally true that any play that involves miscoordination is CP-trumped by σ_L, σ_R,or σ_C. The proof instead first establishes that miscoordination after all players send the *same* message must be Pareto-dominated by either σ_L or σ_R (Lemma 7). This is then used to show that a weakly communication-proof equilibrium strategy must be mutual-preference consistent (Lemma 8). Then one can show that a weakly communication-proof strategy must be coordinated and must have binary communication (Lemma 9). □

Prop. 1 and Cor. 1 also hold in the multi-player setting: communication-proof equilibrium strategies are interim Pareto-undominated and Pareto-improving relative to all symmetric equilibria of the game without communication. By contrast, Prop. 2 does not extend to this setting: with three players, for instance, for some distributions of values F, the strategy that determines the fallback option by majority vote (in the case of messages that indicate different preferred actions) is an ex-ante payoff improvement over a simple fallback norm of choosing, say, action L in every case of disagreement.

E.4 Asymmetric Coordination Games

Adapted Model Consider a setup similar to our baseline model except that the distributions of the types of the two players' positions differ: the type of player 1 is distributed according to F_1 and the type of player 2 is distributed according to F_2. As in the baseline model, both distributions are assumed to be atomless with full support in $[0,1]$. Let $\langle \Gamma(F_1,F_2), M\rangle$ denote the asymmetric coordination game with communication (to ease notation, we assume that both players have the same set of messages at their disposal). Let Σ^i denote the set of all strategies of player $i \in \{1,2\}$. We let i denote the index of one player and j denote the index of the opponent.

Remark 3. The game $\langle \Gamma(F,F), M\rangle$ in which both players have the same distribution of types corresponds to a setup, in which the payoff-irrelevant position of player 1 or player 2 is identifiable, and the players can condition their play on their positions.

Given a strategy profile (σ_1,σ_2), let $\pi_u^i(\sigma_1,\sigma_2)$ denote the (interim) payoff of type u of player $i \in \{1,2\}$, and let $\pi^i(\sigma_1,\sigma_2) = \mathbb{E}_{u\sim F_i}\left[\pi_u^i(\sigma_1,\sigma_2)\right]$ denote the ex-ante payoff of player $i \in \{1,2\}$. A strategy profile (σ_1,σ_2) is an *equilibrium* if $\pi_u^1(\sigma_1,\sigma_2) \geq \pi_u^1(\sigma_1',\sigma_2)$ for each strategy $\sigma_1' \in \Sigma^1$ and for each type u of player 1, and $\pi_u^2(\sigma_1,\sigma_2) \geq \pi_u^2(\sigma_1,\sigma_2')$ for each strategy $\sigma_2' \in \Sigma^2$ and for each type u of player 2.

Adapted Properties We adapt the three key properties of Section 3 as follows. Let $\mu_u^i(m_i)$ denote the probability, given message function μ^i, that player i sends message m_i if she is of type u_i. Let $\mu^i(m_i) = \mathbb{E}_{u\sim F_i}\left[\mu_u^i(m_i)\right]$ be the average (ex-ante) probability of player i sending message m_i. A strategy profile (σ_1,σ_2) is *mutual-preference consistent* if whenever $u_1,u_2 < 1/2$ then $\xi_1(m_1,m_2) = \xi_2(m_1,m_2) = L$ for all $m_1 \in \text{supp}\left(\mu_u^1\right)$ and $m_2 \in \text{supp}\left(\mu_u^2\right)$, and whenever $u_1,u_2 > 1/2$ then $\xi_1(m_1,m_2) = \xi_2(m_1,m_2) = R$ for all $m_1 \in \text{supp}\left(\mu_u^1\right)$ and $m_2 \in \text{supp}\left(\mu_u^2\right)$.

A strategy profile (σ_1, σ_2) is *coordinated* if $\xi_1(m_1, m_2) = \xi_2(m_1, m_2) \in \{L, R\}$ for each pair of messages $m_1 \in \text{supp}\left(\mu^1\right)$ and $m_2 \in \text{supp}\left(\mu^2\right)$.

For any strategy profile $\sigma = \left((\mu^1, \xi_1), (\mu^2, \xi_2)\right) \in \Sigma^1 \times \Sigma^2$ and any message $m_j \in M$, define

$$\beta_i^\sigma(m_j) = E_{u \sim F_i}\left[\sum_{m_i \in M} \mu_u^i(m_i) 1_{\{u \leq \xi_i(m_i, m_j)\}}\right]$$

as the expected probability of player i playing L conditional on player j sending message $m_j \in M$. We say that strategy profile $\sigma = (\sigma_1, \sigma_2)$ has *(essentially) binary communication* if there are two pairs of numbers $0 \leq \underline{\beta}_1^\sigma \leq \overline{\beta}_1^\sigma \leq 1$ and $0 \leq \underline{\beta}_2^\sigma \leq \overline{\beta}_2^\sigma \leq 1$ such that for all messages $m \in M$ and each player $i \in \{1, 2\}$ we have $\beta_i^\sigma(m) \in [\underline{\beta}_i^\sigma, \overline{\beta}_i^\sigma]$; for all messages $m \in M$ such that there is a type $u < 1/2$ with $\mu_u^j(m) > 0$ we have $\beta_i^\sigma(m) = \overline{\beta}_i^\sigma$; and for all messages $m \in M$ such that there is a type $u > 1/2$ with $\mu_u(m) > 0$ we have $\beta_i^\sigma(m) = \underline{\beta}_i^\sigma$.

Consider a strategy profile $\sigma = (\sigma_1, \sigma_2)$ that is coordinated and mutual-preference consistent and has binary communication. Then there are $\alpha_1^\sigma, \alpha_2^\sigma \in [0, 1]$ such that, for each $i \in \{1, 2\}$,

$$\underline{\beta}_i^\sigma = \left(1 - F_j(\tfrac{1}{2})\right) \alpha_i^\sigma \text{ and } \overline{\beta}_i^\sigma = F_j(\tfrac{1}{2}) + \left(1 - F_j(\tfrac{1}{2})\right) \alpha_i^\sigma,$$

where α_i^σ is the probability of coordination on L conditional on player i having type $u_i < 1/2$ and player j having type $u_i > 1/2$. We refer to $\alpha^\sigma = \left(\alpha_1^\sigma, \alpha_2^\sigma\right)$ as the *left-tendency profile* of a strategy profile σ that is coordinated and mutual-preference consistent and has binary communication. It is simple to see that the set of strategies satisfying the above three properties (coordination, mutual-preference consistency, and binary communication) is essentially two-dimensional because the left-tendency profile $\alpha^\sigma = \left(\alpha_1^\sigma, \alpha_2^\sigma\right)$ of such a strategy profile σ describes all payoff-relevant aspects. Two such strategy profiles σ and σ' with the same left-tendency profile (i.e., with $\alpha^\sigma = \alpha^{\sigma'}$) 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 not payoff-relevant, as the probability of the joint lottery inducing the players to play L remains the same.

Adaptation of communication-proofness Given a strategy profile of the game $\langle \Gamma, M \rangle$ we denote the induced "renegotiation" game after a positive probability message pair $m_1, m_2 \in M$ is sent by $\langle \Gamma(F_{m_1}, F_{m_2}), \tilde{M} \rangle$. For a strategy profile σ' of such a renegotiation game $\langle \Gamma(G_1, G_2), \tilde{M} \rangle$, define the *post-communication* expected payoffs for a player i of type u by

$$\pi_u^{i, G_2}\left(\sigma'\right) = \mathbb{E}_{v \sim G_2}\left[\pi_{u,v}^i\left(\sigma'\right)\right] \equiv \int_{v=0}^1 \pi_{u,v}^i\left(\sigma'\right) g_2(v)\, dv.$$

Define $\mathscr{E}(G_1, G_2)$ as the set of all (possibly asymmetric) equilibrium profiles of the coordination game with communication $\langle \Gamma(G_1, G_2), \tilde{M} \rangle$ for some finite message set $\tilde{M}$.

Adapted Results Our main result remains the same in the setup of asymmetric coordination games. The proof, which is analogous to the proof of Theorem 1, is omitted for brevity.

Theorem 4. *Let σ be a strategy profile of $\langle \Gamma(F_1, F_2), M \rangle$. Then the following three statements are equivalent:*

1. *σ is mutual-preference consistent, coordinated, and has binary communication.*

2. *σ is a strongly communication-proof equilibrium strategy.*

3. *σ is a weakly communication-proof equilibrium strategy.*

Prop. 1–2 and Cor. 1 can be adapted to the present setup analogously (proofs are omitted for brevity). In addition, it is straightforward to see that the asymmetric equilibrium with the left-tendency profile $(1,0)$ (resp., $(0,1)$) that always coordinates on the action preferred by Player 1 (resp., Player 2) is ex-ante Pareto efficient (in contrast to the symmetric case, in which sometimes none of the symmetric communication-proof equilibria are ex-ante Pareto efficient).

E.5 Coordination Games with More Than 2 Actions

Next we extend our main model to games with more than two actions. We consider a coordination game with two players in which the two players first send one message from a finite message set M and then, after observing the message pair, choose one action from the ordered set $A = (a_1, \ldots, a_k)$ with $2 < k < \infty$.

A player's type is now a vector $u = (u_1, \ldots, u_k) \in [0,1]^k$, where we interpret the i-th component u_i as the payoff of the agent if both players choose action a_i. If the players choose different actions (miscoordinate), then they both get a payoff of zero. We assume that the distribution of types F is a continuous (atomless) distribution with full support in $[0,1]^k$. For each action a_i, let p_i be the probability that the preferred action of a random type is a_i (i.e., the probability that $u_i = \max(\{u_1, \ldots, u_k\})$). Let $\langle \Gamma_A, M \rangle = \langle \Gamma_A(F), M \rangle$ be the coordination game with set of actions A and pre-play communication.

A player's (ex-ante) strategy is a pair $\sigma = (\mu, \xi)$, where $\mu : U \to \Delta(M)$ is a *message function* that describes which message is sent for each possible realization of the player's type, and $\xi : M \times M \times U \to \Delta(A)$ is an *action function* that describes the distribution of actions chosen as a measurable function of the player's type and the observed message profile. That is, when a player of type u who follows strategy (μ, ξ) observes a message profile (m, m'), then this player plays action a_i with probability $\xi_u(m, m')(a_i)$.

As in the main model, this game has many equilibria. For every action there is a babbling equilibrium in which players of all types after observing any message pair play this action. For every pair of actions a_i, a_j there are also equilibria in which players send only one of two messages, one message indicating a preference for a_i and another for not a_i with play coordinated on a_i if both players send the appropriate message and play coordinated on a_j otherwise. None of these equilibria are communication-proof as they are not mutual-preference consistent and mutual-preference consistency is a necessary condition for a strategy to be communication-proof also in the present context, as we shall see below.

It is more difficult to find equilibria that are mutual-preference consistent, that is equilibria in which each player indicates her most preferred action out of all k actions and play is coordinated on that action if both players indicate a preference for it. Simple adaptations of σ_L and σ_R are not equilibria in the present context. To see this, consider a strategy in which there is a "fallback" action, say action a_1, in which players indicate their most preferred action (the action with the highest u_i), and in which play is coordinated on either action a_i if both players indicate a preference for it, or coordinated on action a_1 otherwise. Suppose that the distribution of types is such that there are two actions a_i and a_j (unequal to each other and unequal to a_1) with $p_j > p_i$. But then there is a player type $u = (u_1, u_2, \ldots, u_k)$ with $u_i = \max\{u_1, \ldots, u_k\}$, u_j very close to u_i, and $u_1 < u_j$, who would prefer to indicate a preference for action a_j. Indicating a preference for action a_i, under the given strategy, provides her with a payoff of $p_i u_i + (1 - p_i) u_1$. Indicating a preference for a_j yields a payoff of $p_j u_j + (1 - p_j) u_1$. But then for a suitably chosen vector $u = (u_1, u_2, \ldots, u_k)$ the latter expression is greater than the former, which contradicts the supposition that the given strategy is an equilibrium.

Next we show that a simple adaptation of σ_C remains a strongly communication-proof equilibrium strategy also in this setup. Let $m_1^0, m_2^0, \ldots, m_k^0, m_1^1, m_2^1, \ldots, m_k^1 \in M$ be $2k$ distinct messages, where the index i of message m_i^b is interpreted as denoting that the agent's preferred outcome is the i-th outcome, and the index $b \in \{0, 1\}$ is interpreted as a random binary number. Let $\sigma_C = (\mu_C, \xi_C)$ be extended to the current setup as follows. Define $\mu_C(u) = {}^1\!/\!{}_2 m_i^0 \oplus {}^1\!/\!{}_2 m_i^1$, where $i = \operatorname{argmin}_j \{u_j \mid u_j = \max\{u_1, .., u_k\}\}$. Thus, the message function μ_C induces each agent to reveal her preferred outcome, and to uniformly choose a binary number (either, zero or one). In the second stage, if both agents share the same preferred outcome they play it. Otherwise, they coordinate on the preferred action with the smaller index if both agents have chosen the same random number, and they coordinate on the preferred outcome with the larger index if both agents have chosen different random numbers, i.e.,

$$\xi_C\left(m_i^b, m_j^c\right) = \begin{cases} a_i & (i \leq j \text{ and } b = c) \;\; OR \;\; (i \geq j \text{ and } b \neq c) \\ a_j & \text{otherwise.} \end{cases}$$

We then have the following proposition.

Proposition 8. *Strategy σ_C is a strongly communication-proof equilibrium strategy in the game*

$\langle \Gamma_A, M \rangle$.

Proof. Observe that an agent who sends message m_i^b obtains an expected payoff of $1/2 u_i + 1/2 \sum_{j=1}^{k} p_j u_j$ when facing a partner who follows strategy σ_C. As the second term in this sum is the same for all messages, an agent of type u sends this message only if $u_i = \max\{u_1, \ldots, u_k\}$, as required. The remaining arguments as to why the second-stage behavior is a best reply and why σ_C is strongly communication-proof are analogous to the proof of the "if" part of Theorem 1 and are omitted for brevity. □

A strategy $\sigma = (\mu, \xi)$ is *same-message coordinated* if for all messages $m \in \text{supp}(\bar{\mu})$ there is an action a_i such that for all u with $\mu_u(m) > 0$ we have $\xi(u, m, m) = a_i$. In what follows we show that a necessary condition for a strategy to be a communication-proof equilibrium strategy is that this strategy is same-message coordinated and mutual-preference consistent.

Proposition 9. *If strategy σ of the game $\langle \Gamma_A, M \rangle$ with action set A is strongly communication-proof, then it is same-message coordinated and mutual-preference consistent.*

Sketch of proof; for the formal proof see Appendix E.7.3. To show that a communication-proof equilibrium strategy is same-message coordinated we cannot, in fact, use the proof of the main theorem because Lemma 2 crucially depends on the game having only two actions. Instead, we suppose to the contrary that there is a communication-proof equilibrium strategy in which play is not coordinated after both players have sent the same message m. This strategy thus induces some nondegenerate probability distribution over actions after both players send message m. We then construct a CP-trumping equilibrium that is fully coordinated and has a probability of coordination on every action exactly equal to the probability of this action being played under the original strategy conditional on observing (m, m), which contradicts the supposition. The construction is achieved by players sending random messages in such a way that they are indifferent between all messages and this joint lottery is implemented.[21] The proof that a strongly communication-proof equilibrium strategy must be mutual-preference consistent is then achieved straightforward (by a simple adaptation of Lemma 8). □

We are able neither to show nor to provide a counterexample that a strongly communication-proof strategy must be coordinated after observing a pair of different messages, and that it must have binary communication.

E.6 Extreme Types with Dominant Actions

In this subsection we show how to extend our analysis to a setup in which some types have an extreme preference for one of the actions such that it becomes a dominant action for them.

[21] The original construction of Aumann and Maschler (1968) relies on the probabilities of different actions being rational numbers. In Appendix E.7.3 we present a more elaborate implementation that allows to deal also with irrational numbers in the current setup.

Let $a < 0$ and $b > 1$. We extend the set of types to be the interval $[a,b]$. Observe that action L (R) is a dominant action for any type $u < 0$ ($u > 1$) as coordinating on R (L) yields to such a type a negative payoff of $u < 0$ ($1-u < 0$). We call types with a dominant action (i.e., $u < 0$ or $u > 1$) *extreme*, and types that do not have a strictly dominant action (i.e., $u \in [0,1]$) *moderate*. We assume that the cumulative distribution of types F is continuous (atomless) and has full support in the interval $[a,b]$.

We further assume that the extreme types are a minority both among the agents who prefer action R and among the agents who prefer L, i.e., $F(0) < {}^1\!/\!_2 F({}^1\!/\!_2)$ and $1 - F(1) < {}^1\!/\!_2 (1 - F({}^1\!/\!_2))$. Next, we adapt the definitions of coordination and binary communication to the current setup. The original definition of coordination is too strong in the current setup, as, clearly, when extreme types with different preferred outcomes meet they must miscoordinate. Thus, we present a milder notion. A strategy is *weakly coordinated* if whenever two moderate types meet they never miscoordinate. Note that the definition does not impose any restriction on what happens when an extreme type meets a moderate type.

The original definition of binariness is too weak in the current setup. This is because coordinated strategies must allow for some miscoordination between extreme types, which implies that an agent cares not only about the average probability of the opponent playing left (i.e., $\beta^{\sigma}(m)$), but also about the total probability of miscoordination. Thus, we strengthen binariness by requiring that there exist two distributions of messages, which are used by all types below ${}^1\!/\!_2$ and all types above ${}^1\!/\!_2$, respectively. Formally, a strategy $\sigma = (\mu, \xi)$ has *strongly binary communication* if $\mu(u) = \mu(u')$ if either $u, u' \leq {}^1\!/\!_2$ or $u, u' > {}^1\!/\!_2$. It is easy to see that the strategies $\sigma_L, \sigma_R, \sigma_C$ defined in Section 3 all satisfy strongly binary communication. Moreover, one can show, for any $\alpha \in [0,1]$, that if there exists a strategy σ that is coordinated, mutual-preference consistent, and has binary communication with left tendency α, then there also exists strategy $\tilde{\sigma}$ with the same properties that is strongly binary communication.

Our next result shows that there exists, essentially, a unique communication-proof equilibrium strategy that is coordinated, mutual-preference consistent, and has strongly binary communication.

Proposition 10. *In a coordination game with communication and with dominant action types, a strategy σ that is coordinated, mutual-preference consistent, and has strongly binary communication is a strongly communication-proof equilibrium strategy if and only if it has a left tendency of* $\alpha = \frac{F(0)}{F(0)+(1-F(1))}$.

The formal proof is presented in Appendix E.7.4. The key intuition is that given the frequency of dominant action types $F(0) > 0$ (of L-dominant action types) and $1 - F(1) > 0$ (of R-dominant action types) to make the agent of type $u = {}^1\!/\!_2$ indifferent between signaling a lower than half or higher than half type we must have a strategy that counterbalances these frequencies of dominant action types. To see this, consider an adaptation of $\sigma_L = (\mu^*, \xi_L)$ to this setting by having extreme types follow their dominant action in the second stage (and moderate types play in the same way as in the baseline model). Note that σ_L is no longer an equilibrium with extreme types. Observe that having a moderate type send message m_R leads to coordination with probability one (sometimes on R and sometimes on L depending on the opponent's message), while having a moderate type send message m_L leads to coordination (on L only) with probability $F(1) < 1$. This implies that agents of type $u < 1/2$ sufficiently close to $1/2$ strictly prefer sending message m_R to sending message m_L (as the former induces a higher probability of coordination on the same action as the partner), which contradicts the supposition that σ_L is an equilibrium strategy.

Appendix E.7.4 also shows that a left tendency α communication-proof strategy that is coordinated and has strongly binary communication can be implemented whenever α is a rational number and the set of messages M is sufficiently large (and irrational α-s can be approximately implemented by ε-equilibria).

Observe that in the symmetric case ($F(0) = 1 - F(1)$), strategy σ_C is the essentially unique strongly communication-proof strategy with the above two properties. Further observe that in the asymmetric case, the moderate types gain if the extreme types with the same preferred outcome are more frequent than the extreme types of the opposite preferred outcome. Specifically, if there are more extreme "leftists" than extreme "rightists" (i.e., $F(0) > 1 - F(1)$), then the essentially unique strongly communication-proof strategy with properties of coordination and strongly binary communication induces higher probability to coordinate on action L (rather than on action R) whenever two moderate agents with different preferred outcomes meet.

E.7 Formal Proofs of Extensions

E.7.1 Proof of Theorem 2 (Multidimensional types, Section E.2)

The proof of Thm. 2 mimics the proof of Thm. 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 $m, m' \in \operatorname{supp}(\bar{\mu})$,

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

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

$$\inf\{\varphi_u \mid \mu_u(m) > 0\} < \xi(m, m') < \sup\{\varphi_u \mid \mu_u(m) > 0\}.$$

As σ is an equilibrium, we must have $\inf\{\varphi_u \mid \mu_u(m') > 0\} < \xi(m', m) < \sup\{\varphi_u \mid \mu_u(m') > 0\}$. (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 σ_L, σ_R, or σ_C (all based on φ_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 σ_R. Consider the player who sent message m.

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

$$u_{LL}F_{m'}(x') + \left(1 - F_{m'}(x')\right)u_{LR} \leq u_{LL}F_{m'}(\tfrac{1}{2}) + u_{LR}\left(1 - F_{m'}(\tfrac{1}{2})\right) \leq u_{LL}F_{m'}(\tfrac{1}{2}) + u_{RR}\left(1 - F_{m'}(\tfrac{1}{2})\right),$$

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 σ_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

$$u_{RL}F_{m'}(x') + u_{RR}\left(1 - F_{m'}(x')\right) \leq u_{LL}F_{m'}(x') + u_{RR}\left(1 - F_{m'}(x')\right) \leq u_{LL}F_{m'}(\tfrac{1}{2}) + u_{RR}\left(1 - F_{m'}(\tfrac{1}{2})\right),$$

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 σ_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 $u_{LL} > u_{RR}$.

Case 1c: Finally, consider a type u with $\varphi_u > 1/2$. Then we have $u_{RR} > u_{RL}F_{m'}(x') + u_{RR}(1 - F_{m'}(x'))$, where the right-hand side is the type u agent's payoff under (x, x') and the left-hand side is her payoff under σ_R. 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 σ_R with σ_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 σ_C. Consider the player who sent message m.

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

$$u_{LL}\left[F_{m'}(\tfrac{1}{2})+\tfrac{1}{2}\left(1-F_{m'}(\tfrac{1}{2})\right)\right]+u_{RR}\tfrac{1}{2}\left(1-F_{m'}(\tfrac{1}{2})\right) > u_{LL}F_{m'}(x')+u_{LR}\left(1-F_{m'}(x')\right),$$

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 σ_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

$$u_{RL}F_{m'}(x')+u_{RR}\left(1-F_{m'}(x')\right) \leq u_{LL}F_{m'}(x')+u_{RR}\left(1-F_{m'}(x')\right) \leq u_{LL}\left[\tfrac{1}{2}+\tfrac{1}{2}F_{m'}(x')\right]+u_{RR}\tfrac{1}{2}\left(1-F_{m'}(\tfrac{1}{2})\right),$$

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 σ_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 $\varphi_u \leq 1/2$ and $F_{m'}(x') = x$ by (x,x') being an equilibrium and $x < 1/2$. The inequality is strict if $u_{LL} > u_{RL}$ or $u_{LL} > u_{RR}$.

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

$$\begin{aligned} u_{RL}F_{m'}(x')+u_{RR}\left(1-F_{m'}(x')\right) &< u_{RL}\tfrac{1}{2}F_{m'}(\tfrac{1}{2})+u_{RR}\left[\left(1-F_{m'}(\tfrac{1}{2})\right)+\tfrac{1}{2}F_{m'}(\tfrac{1}{2})\right] \\ &\leq u_{LL}\tfrac{1}{2}F_{m'}(\tfrac{1}{2})+u_{RR}\left[\left(1-F_{m'}(\tfrac{1}{2})\right)+\tfrac{1}{2}F_{m'}(\tfrac{1}{2})\right], \end{aligned}$$

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 σ_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. □

E.7.2 Proof of Theorem 3 (Multiple Players, Section E.3)

The "1 ⇒ 2" part is analogous to the proof of the "1 ⇒ 2" part of Theorem 1. The proof of the "3 ⇒ 1" part does not extend directly and has to be adapted as follows. The following lemma states that play is coordinated whenever all players send the same message.

Lemma 7. *Let $\sigma = (\mu,\xi)$ be a weakly communication-proof equilibrium strategy. Let $m \in supp(\bar{\mu})$ and let $\mathbf{m} = (m,\ldots,m)$ be the vector with n identical entries of m, which represents the case of all n*

players sending message m. Then either $\xi(\mathbf{m}) \geq \sup\{u \mid \mu_u(m) > 0\}$ *or* $\xi(\mathbf{m}) \leq \inf\{u \mid \mu_u(m) > 0\}$.

Proof. Let $m \in \operatorname{supp}(\bar{\mu})$ and assume to the contrary that $\inf\{u \mid \mu_u(m) > 0\} < \xi(\mathbf{m}) < \sup\{u \mid \mu_u(m) > 0\}$. Let $x = \xi(\mathbf{m})$. We now show that the symmetric equilibrium in which all players use cutoff x after sending the identical message m, denoted by $\mathbf{x} = (x, \ldots, x)$, is Pareto-dominated by σ_L or σ_R.

There are two cases to be considered. Case 1: Suppose that $x \leq 1/2$. We now show that in this case the equilibrium $\mathbf{x}$ is Pareto-dominated by σ_R.

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

$$(1-u)\left(F_m(\tfrac{1}{2})\right)^{n-1} + u\left(1 - \left(F_m(\tfrac{1}{2})\right)^{n-1}\right) \geq (1-u)\left(F_m(x)\right)^{n-1},$$

where the left-hand side is the type u agent's payoff under strategy profile σ_R and the right-hand side is her payoff under strategy profile $\mathbf{x}$. The inequality follows from $u\left(1 - (F_m(1/2))^{n-1}\right) \geq 0$ and $F_m(1/2) \geq F_m(x)$ by the fact that F_m is nondecreasing (as it is a cumulative distribution function). Note also that this inequality is strict for all u except for $u = 0$ in the case of $x = 1/2$.

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

$$(1-u)\left(F_m(\tfrac{1}{2})\right)^{n-1} + u\left(1 - \left(F_m(\tfrac{1}{2})\right)^{n-1}\right) \geq u\left(1 - \left(F_m(x)\right)^{n-1}\right),$$

where the left-hand side is the type u agent's payoff under strategy profile σ_R and the right-hand side is her payoff under strategy profile $\mathbf{x}$. The inequality follows from the fact that given $u \leq 1/2$ we have that $1 - u \geq u$ and therefore

$$(1-u)\left(F_m(\tfrac{1}{2})\right)^{n-1} + u\left(1 - \left(F_m(\tfrac{1}{2})\right)^{n-1}\right) \geq u.$$

Note that this inequality actually holds strictly for all u.

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

Case 2: Suppose that $x \geq 1/2$. The analysis is analogous to Case 1 if we replace σ_R with σ_L. □

Lemma 8. *Every weakly communication-proof equilibrium strategy* $\sigma = (\mu, \xi)$ *is mutual-preference consistent.*

Proof. The proof of this lemma involves two steps. In the first step we show that a communication-proof equilibrium strategy σ is ordinal preference revealing, i.e., such that for any message $m \in \operatorname{supp}(\bar{\mu})$, $F_m(1/2) \in \{0, 1\}$. We then use this to show that σ is mutual-preference consistent.

Assume, first, that σ is a communication-proof but not ordinal preference-revealing equilibrium strategy. That is, suppose to the contrary that $F_m(1/2) \in (0, 1)$. Then there are types $u < 1/2$ as well as types $u > 1/2$ who both send message m with positive probability. By Lemma 7 play after message pair (m,m) must be either L or R. If it is L then the equilibrium strategy σ_L Pareto-dominates playing L, with a strict payoff improvement for all types $u > 1/2$ (and unchanged payoffs for all types $u \leq 1/2$). If it is R then the equilibrium strategy σ_R Pareto-dominates playing R, with a strict payoff improvement for all types $u < 1/2$ (and unchanged payoffs for all types $u \geq 1/2$).

Given a communication-proof equilibrium strategy $\sigma = (\mu, \xi)$, we can classify messages in the support of μ into two distinct sets, $M_L = M_L(\sigma) = \{m \in \operatorname{supp}(\mu) \mid F_m(1/2) = 1\}$ and $M_R = M_R(\sigma) = \{m \in \operatorname{supp}(\mu) \mid F_m(1/2) = 0\}$, where $M_L \cap M_R = \emptyset$ and $M_L \cup M_R = \operatorname{supp}(\bar{\mu})$.

To show that a communication-proof equilibrium strategy $\sigma = (\mu, \xi)$ is mutual-preference consistent, then consider any profile of types $(u_1, u_2, \ldots, u_n)$ such that $u_i < 1/2$ for all $i \in \{1, \ldots, n\}$. They must each send a message in M_L, which we denote by the profile $\mathbf{m} = (m_1, \ldots, m_n)$. Any play after message profile m that is not coordinated on L is now clearly Pareto-dominated (given that all types $\leq 1/2$) by playing the equilibrium strategy L. The case for a profile of types $u_i > 1/2$ for all players is proven analogously. □

The following lemma shows that, in a communication-proof equilibrium strategy, agents never miscoordinate after observing any message profile.

Lemma 9. *Every weakly communication-proof equilibrium strategy σ is coordinated.*

Proof. Suppose that $\sigma = (\mu, \xi)$ is a communication-proof equilibrium strategy. Given Lemmas 7 and 8 it only remains to prove that play under σ is coordinated even after mixed messages are sent, i.e., when there is at least one player who sends a message in M_L and another player who sends a message in M_R, where M_L and M_R are as defined in the proof of Lemma 8. Suppose that this is the case. Then let $I \subset \{1, \ldots, n\}$ be the set of all players who send a message $m_i \in M_L$. Let I^c denote its complement. By Lemma 8 all $i \in I^c$ satisfy $m_i \in M_R$. Let $x_i = \xi(m_i, m_{-i})$ be the cutoff used by player i after observing message profile $(m_1, m_2, \ldots, m_n)$. Then by Lemma 8 we have $x_i \leq 1/2$ for all $i \in I$ and $x_i \geq 1/2$ for all $i \in I^c$. For this profile $\mathbf{x} = (x_1, \ldots, x_n)$ to be an equilibrium after the players observe message profile $(m_1, m_2, \ldots, m_n)$, we must have that for each $i = 1, 2, \ldots, n$, the probability that player i's opponents coordinate their action on L conditional on them coordinating (on either L or R) is

$$x_i = \frac{\prod_{j \neq i} F_{m_j}(x_j)}{\prod_{j \neq i} F_{m_j}(x_j) + \prod_{j \neq i} \left(1 - F_{m_j}(x_j)\right)}.$$

But then, all types of all players, after observing message profile $(m_1, m_2, \ldots, m_n)$, weakly (and some strictly) prefer to play σ_C, which is a payoff identical in this case to a public fair coin toss to determine whether coordination should be on L or R. To see this, consider a player i who sent a message in M_L (i.e., $u_i \leq 1/2$, which implies that $x_i \leq 1/2$) and consider the following two cases.

Case 1: Suppose that $u_i \leq x_i$. Under the given strategy, this type's payoff is $(1-u_i)\prod_{j\neq i} F_{m_j}(x_j)$ with $\prod_{j\neq i} F_{m_j}(x_j) \leq 1/2$. The equilibrium strategy σ_C yields to this type a payoff of $1/2(1-u_i)+1/2u_i$, which exceeds the former payoff, which contradicts the supposition that σ is communication-proof.

Case 2: Suppose that $x_i < u_i \leq 1/2$. Then, under the given strategy, this type's payoff is $u_i \prod_{j\neq i}\left(1-F_{m_j}(x_j)\right)$. The equilibrium σ_C yields to this type a payoff of $1/2(1-u_i)+1/2u_i$, which, by virtue of $1-u_i \geq u_i$, again exceeds the former payoff, which contradicts the supposition that σ is communication-proof.

The analysis for a player who sent a message in M_R is proven analogously. □

To complete the proof of Theorem 1 for the case of many players, we need to prove that any communication-proof equilibrium strategy also has binary communication. The proof of this statement is analogous to the proof of Lemma 3 and therefore omitted.

E.7.3 Proof of Proposition 9 (Multiple Actions, Section E.2)

For the proof, two lemmas about approximating real numbers by rational numbers will be useful. The first lemma shows that any discrete distribution with at least three elements in its support can be approximated from below by a vector of rational numbers, such that the profile of differences (between the irrational exact probability and its rational approximation from below) is roughly uniform in the sense that no difference is larger than the half the sum of all the differences.

Lemma 10. *Let $p \in \Delta(A)$ be a distribution satisfying $|supp(p)| \geq 3$. Then there exists a function $q : A \to \mathbb{R}^+$ such that, for each $1 \leq i \leq k$, $q(a_i)$ is a rational number, $q(a_i) \leq p(a_i)$, and*

$$p(a_i) - q(a_i) \leq \tfrac{1}{2}\textstyle\sum_{1\leq j\leq k}\left(p(a_j) - q(a_j)\right).$$

Proof. Let $\delta < \min\{p(a)/2 \mid a \in \mathrm{supp}(p)\}$. As the rational numbers are dense in the reals, for each real number $\hat{p} > \delta$, there exists a rational number $\hat{q} \in (0,\hat{p})$ such that $\hat{p}-\hat{q} \in ((9/10)\delta, \delta)$. Call this $\hat{q}$ a *rational approximation* of $\hat{p}$. For each $a \in \mathrm{supp} p$, let $q(a)$ be a rational approximation of $p(a)$. For each $a \notin \mathrm{supp}(p)$ let $q(a) = q(a) = 0$. Then it follows that, for each $1 \leq i \leq k$, $q(a_i)$ is a rational number, and $q(a_i) \leq p(a_i)$. Finally we get, for each $1 \leq i \leq k$, that

$$p(a_i) - q(a_i) \leq \delta \leq \tfrac{1}{2}\left|\mathrm{supp}(p)\right| \tfrac{9}{10}\delta \leq \tfrac{1}{2}\textstyle\sum_{1\leq j\leq k}\left(p(a_j) - q(a_j)\right),$$

where the first inequality follows directly from the definition of a rational approximation, the second one follows from the assumption that $|\mathrm{supp}(p)| \geq 3$, and the last one from the assumption that, for each $a_j \in \mathrm{supp}(p)$, $p(a_j) - q(a_j) > (9/10)\delta$ by the definition of a rational approximation. □

Note that the closer δ is to zero, the better the rational approximation constructed in this proof. Note, however, that this does not matter in the proof of Proposition 9 below, which simply uses any (possibly quite rough) rational approximation.

The second lemma is utilized in the proof of Proposition 9 for the case where the distribution in question has exactly two elements in its support.

Lemma 11. *Let $p, q \in (0,1)$. Then there exists a rational number $\alpha \in (0,1)$ satisfying*

$$\frac{p-q}{1-q} < \alpha < \frac{p}{q}.$$

Proof. Note that, as $p < 1$ the following inequality holds: $\leq (p-q)^2 = p^2 - 2pq + q^2 < p - 2pq + q^2$. The inequality $0 < p - 2pq + q^2$ then implies that

$$qp - q^2 < p - pq \Leftrightarrow q(p-q) < p(1-q) \Leftrightarrow \frac{p-q}{1-q} < \frac{p}{q}.$$

The result then follows from the fact that $p - q < 1 - q$. □

We now turn to the proof of Proposition 9. Let $\sigma = (\mu, \xi)$ be a communication-proof equilibrium strategy of $\langle \Gamma_A, M \rangle$. We begin by showing that σ is same-message coordinated. Let $m \in \operatorname{supp}(\bar{\mu})$ let and $p \in \Delta(A)$ be the distribution of play under σ conditional on message pair (m,m) being observed. Assume to the contrary that $p(a) < 1$ for each $a \in A$ (i.e., that there is miscoordination).

Case I: Assume that $|\operatorname{supp}(p) \geq 3|$. Let $q : A \to \mathbb{R}^+$ be a rational approximation of p satisfying the requirements of Lemma 10. The fact that all $q(a_i)$-s are rational numbers implies that there are $l_1, \ldots, l_k, n \in \mathbb{N}$, such that $q(a_i) = l_i / n$ for each i and $l_1 + \ldots + l_k \leq n$. Consider the following equilibrium strategy $\tilde{\sigma} = \left(\tilde{\mu}, \tilde{\xi}\right)$ of the game induced after players observe message pair (m,m) with an additional communication round with the set of messages $\tilde{M} = \left\{ m_{B,i,b} \mid 1 \leq B \leq n,\ 1 \leq i \leq k,\ b \in \{0,1\} \right\}$. We let $1 \leq B \leq n$ denote a random integer used for a joint lottery, $1 \leq i \leq k$ denote the index of the player's preferred outcome (i.e., $u_i = \max \left\{ u_j \mid 1 \leq j \leq k \right\}$, and $b \in \{0,1\}$ denotes a random bit). The message function $\tilde{\mu}$ induces each agent to choose the indexes B and b randomly (uniformly, and independently of each other), and to choose i such that $u_i = \max \left\{ u_j \mid 1 \leq j \leq k \right\}$ is her preferred outcome.

The action function $\tilde{\xi}\left(m_{B,i,b}, m_{B',i',b'}\right)$ is defined as follows. Let $\hat{B} = (B + B') \mod n$ be the sum of the random B-s sent by the players. Both players play action a_j if $l_1 + \ldots + l_{j-1} \leq \hat{B} < l_1 + \ldots + l_j$. If $\hat{B} \geq l_1 + \ldots + l_k$, then both players play the action in $\{a_i, a_{i'}\}$ with the smaller index if action if $b = b'$ and the action with the larger index if $b \neq b'$. Strategy $\tilde{\sigma}$ induces both players to coordinate on a random action with probability $\bar{q} \equiv q(a_1) + \ldots + q(a_k)$ (and, conditional on that, the random action is chosen to be a_j with probability $q\left(a_j\right)/\bar{q}$), and to coordinate on the preferred action of one of the two players (chosen uniformly at random) with probability $1 - \bar{q}$, which can be written as $1 - \bar{q} = \sum_{1 \leq j \leq k} \left(p\left(a_j\right) - q\left(a_j\right) \right)$, by the fact that p is a probability distribution and, thus, $\sum_{1 \leq j \leq k} p\left(a_j\right) = 1$.

The proof that $\tilde{\sigma}$ is an equilibrium is analogous to the proof of Prop. 8 and thus omitted. The expected payoff that the original equilibrium σ yields to each type $u=(u_1,\ldots,u_k)\in U$ in the game induced after observing (m,m) is $\max_j\left\{p\left(a_j\right)u_j\right\}$. The expected payoff that $\tilde{\sigma}$ induces for each type u is at least

$$\sum_j q\left(a_j\right)u_j+\tfrac{1}{2}\left(1-\bar{q}\right)\max_j\left\{u_j\right\}\geq \max_j\{q\left(a_j\right)u_j\}+\tfrac{1}{2}\max_j\left\{u_j\right\}\textstyle\sum_j\left(p\left(a_j\right)-q\left(a_j\right)\right).$$

Thus, the difference between the payoff of $\tilde{\sigma}$ and the payoff of σ for a type u is at least

$$\begin{aligned}\tfrac{1}{2}\max_j\left\{u_j\right\}\textstyle\sum_j\left(p\left(a_j\right)-q\left(a_j\right)\right)-\left(\max_j\left\{p\left(a_j\right)u_j\right\}-\max_j\left\{q\left(a_j\right)u_j\right\}\right) &\geq \\ \tfrac{1}{2}\max_j\left\{u_j\right\}\textstyle\sum_j\left(p\left(a_j\right)-q\left(a_j\right)\right)-\tfrac{1}{2}\sum_{1\leq j\leq k}\left(p\left(a_j\right)-q\left(a_j\right)\right)u_j &\geq 0,\end{aligned}$$

where the first inequality is due to

$$\max_j\left\{p\left(a_j\right)u_j\right\}-\max_j\left\{q\left(a_j\right)u_j\right\}=p\left(a_l\right)u_l-\max_j\left\{q\left(a_j\right)u_j\right\}\leq p\left(a_l\right)u_l-q\left(a_l\right)u_l,$$

where $l=\operatorname{argmax}_j\left\{p\left(a_j\right)u_j\right\}$; and the second inequality is due to q being a rational approximation of p as given by Lemma 10.

This then implies that σ is CP-trumped by $\tilde{\sigma}$, which contradicts the supposition that σ is a communication-proof equilibrium.

Case II: We are left with the case of $|\mathrm{supp}(p)|=2$. Let $\mathrm{supp}(p)=\left\{a_i,a_j\right\}$. Let $q\in(0,1)$ be the posterior probability of a player having a type u with $u_i\geq u_j$, conditional on sending m. Let $p\equiv p\left(a_i\right)$. By Lemma 11, there exists a rational number $\alpha\equiv {}^k/_n\in(0,1)$ satisfying ${}^{p-q}/_{1-q}<\alpha<{}^p/_q$. Consider the following symmetric equilibrium $\tilde{\sigma}=\left(\tilde{\mu},\tilde{\xi}\right)$ of the game induced after players observe message pair (m,m) with an additional communication round with the set of messages $\tilde{M}=\{i,j\}\times\{1,\ldots,n\}$.

The first component of the message of each player is interpreted as her preferred coordinated outcome of a_i and a_j, and the second component is a random number between 1 and n. When following strategy $\tilde{\sigma}$ the players send message component i if and only if $u_i\geq u_j$, send a random number between 1 and n according to the uniform distribution, play a_i after observing $((i,a),(i,b))$ for any numbers a and b, play a_j after observing $((j,a),(j,b))$ for any numbers a and b, play a_i after observing $((i,a),(j,b))$ if $a+b<k \mod n$, and play a_j after observing $\left((a_i,a),\left(a_j,b\right)\right)$ if $a+b\geq k$ mod n.

Observe that $\tilde{\sigma}$ is indeed an equilibrium of the induced game, because following any pair of messages the players coordinate for sure, each agent with $u_i>u_j$ (resp., $u_j>u_i$) strictly prefers to report that her preferred outcome is a_i (resp., a_j) as this induces her to coordinate on a_i (resp., a_j) with a high probability of $q+\alpha(1-q)$ (resp., $1-q+(1-\alpha)q$) instead of with a low probability of αq

(resp., $(1-q)(1-\alpha)$), and each agent is indifferent between sending any random number, as this has no effect on the probability of coordinating on a_i (which is equal to $\alpha = {}^{k}\!/\!_{n}$), given that her opponent chooses his random number uniformly as well.

Recall that the payoff of each type who follows σ in the game induced after observing (m,m) is equal to $\max\left\{u_i p, u_j(1-p)\right\}$. The payoff of each type u with $u_i \geq u_j$ in equilibrium $\tilde{\sigma}$ is given by

$$(q+\alpha(1-q))\,u_i + (1-(q+\alpha(1-q)))\,u_j > pu_i + (1-p)\,u_j \geq \max\left\{u_i p, u_j(1-p)\right\},$$

where the first inequality is implied by $u_i \geq u_j$ and $\frac{p-q}{1-q} < \alpha \Leftrightarrow p < q + \alpha(1-q)$.

The payoff of each type u with $u_i < u_j$ in equilibrium $\tilde{\sigma}$ is given by

$$((1-q)+(1-\alpha)\,q)\,u_j + (1-((1-q)+(1-\alpha)\,q))\,u_i > (1-p)\,u_j + pu_i \geq \max\left\{(1-p)u_j, pu_i\right\},$$

where the first inequality is implied by $\frac{q}{p} > \alpha \Leftrightarrow 1-\alpha > \frac{q-p}{q} \Leftrightarrow (1-q)+(1-\alpha)\,q > 1-p$ and $u_i < u_j$. This implies that all types obtain a strictly larger payoff in $\tilde{\sigma}$ (relative to the expected payoff of σ in the game induced after players observe message pair m,m), which implies that $\tilde{\sigma}$ Pareto-dominates σ, which contradicts the supposition that σ is communication-proof.

Next, we show that $\sigma = (\mu, \xi)$ is mutual-preference consistent. For each $i \in \{1,\ldots,k\}$, let $U_i \subset [0,1]^k$ be the set of types such that $u_i \geq \max_{j\neq i} u_j$. Assume, first, that σ is not ordinal preference-revealing. That is, suppose that there is a message $m \in \mathrm{supp}(\bar{\mu})$ such that there are action indices i,j with $i \neq j$ and $\mu_u(m) > 0$ for some $u \in U_i$ and some $u \in U_j$. We have shown above that the play after players observe (m,m) must be coordinated on some action $a_l \in A$. Now consider the following strategy with new message space $\tilde{M} = \{m_i, m_{\neg i}\}$ in which players of type $u \in U_i$ send message m_i, while all others send $m_{\neg i}$ and play is a_l unless both players send message m_i, in which case it is a_i. This is an equilibrium strategy of the induced game after players observe (m,m) and it Pareto-dominates σ, a contradiction. This proves that a communication-proof equilibrium strategy σ must be ordinal preference-revealing.

Given a communication-proof equilibrium strategy $\sigma = (\mu, \xi)$, we can classify messages in the support of μ into k distinct sets $M_i = M_i(\sigma) = \{m \in \mathrm{supp}(\bar{\mu}) \mid u \in U_i\}$ for each $i = 1,\ldots,k$, where for each i,j with $i \neq j$ $M_i \cap M_j = \emptyset$ and $\bigcup_{i=1}^{k} M_i = \mathrm{supp}(\bar{\mu})$.

To show that a communication-proof equilibrium strategy σ is mutual-preference consistent, consider a message pair (m,m') with $m,m' \in M_i$ for some $i = 1,\ldots,k$. Since σ is ordinal preference-revealing, the updated support of types who observe either m or m' is then in U_i. But then any joint action distribution that the two players could play after (m,m') is Pareto-dominated by the equilibrium of playing action a_i.

E.7.4 Proof of Proposition 10 (Extreme Types, Section E.6)

Proof of Proposition 10. Any strategy σ that is coordinated, mutual-preference consistent, and has strongly binary communication can be characterized by its left tendency (see Section 3) as follows. Under such a mutual-preference consistent strategy players indicate whether their type is below or above $1/2$. This means that there are two disjoint sets of messages, M_L and M_R, such that players of type $u \leq 1/2$ send a message in M_L and players of type $u > 1/2$ send a message in M_R. Also whenever two players both send messages in M_L they then play L and if both send messages in M_R they both play R. The left tendency $\alpha = \alpha^{\sigma}$ then describes how moderate players coordinate if one of them sends a message from M_L and the other sends a message from M_R. The left tendency α is then the probability that the two players coordinate on L (through random message selection within the respective sets of messages), while $1-\alpha$ is then the remaining probability that they coordinate on R.

To prove the "only if" part, consider an arbitrary left tendency of $\alpha \in [0,1]$. Then consider a player of type $1/2$ who needs to be indifferent between sending a message in M_L and sending a message in M_R for this strategy to be an equilibrium strategy. If she sends a message in M_L she coordinates on L whenever either her opponent sends a message in M_L (which happens with probability $F(1/2)$), or her moderate opponent sends a message in M_R (which happens with probability $F(1) - F(1/2)$) and the joint lottery yields the outcome L (which happens with probability α). By contrast, she coordinates on R whenever her opponent sends a message in M_R and the joint lottery yields the outcome R (which happens with probability $1-\alpha$). Therefore, her expected payoff from sending a message in M_L is given by $\frac{1}{2}F\left(\frac{1}{2}\right) + \frac{1}{2}\alpha\left(F(1) - F\left(\frac{1}{2}\right)\right) + \frac{1}{2}(1-\alpha)\left(1 - F\left(\frac{1}{2}\right)\right)$. Similarly, her expected payoff from sending a message in M_R is given by $\frac{1}{2}\alpha F\left(\frac{1}{2}\right) + \frac{1}{2}(1-\alpha)\left(F\left(\frac{1}{2}\right) - F(0)\right) + \frac{1}{2}\left(1 - F\left(\frac{1}{2}\right)\right)$. Clearly, her expected payoff from sending a message in M_L is equal to her expected payoff from sending a message in M_R iff $\alpha = F(0)/\left(F(0) + (1 - F(1))\right)$, as required. This proves the "only if" direction.

To prove the "if" direction, we need to show that a coordinated and mutual-preference consistent strategy with strongly binary communication and with a left tendency of $\alpha = F(0)/\left(F(0) + (1 - F(1))\right)$ is both an equilibrium and a communication-proof strategy. To prove the latter condition the same arguments as in the relevant parts of the proof of the "if" direction of Theorem 1 apply directly. It remains to show that such a strategy is an equilibrium strategy. We have already shown that the message function is a best reply to itself and the action function. All that remains to prove is that the action function is a best reply to the given strategy. It is easy to see that playing L is the optimal strategy when both players send a message in M_L and thus are of type $u < 1/2$. In doing so, they coordinate on their most preferred outcome with probability one. Similarly, playing R after two messages in M_R is clearly optimal. Now suppose that one player sends a message in M_L and the other player sends a message in M_R. There are two possibilities. Either they are now supposed to both play L (unless they are an extreme R type) or they are now supposed to both play R (unless they are an extreme L type). Consider first the person who sends a message in M_L and therefore be of type $u < 1/2$. Suppose that the two players are expected to coordinate on R. Since her opponent sent a message in M_R, our M_L sender expects R with a probability of one (as all M_R senders are of type $u > 1/2$, which excludes L-dominant action types). But then our M_L sender of any type $u > 0$ has a strict incentive to play R as well. Now suppose that the two players are expected to coordinate on L. Then our M_L sender expects her opponent to play L with a probability of $(F(1) - F(1/2)) / (1 - F(1/2))$, which is the conditional probability of an R-type to be moderate, which by assumption is greater than or equal to $1/2$. Playing L in this case is therefore optimal for all M_L senders. That the M_R sender has the correct incentives in her choice of action after any mixed-message pair (one in M_R and one in M_L) is proven analogously and requires the assumption that $F(0)/F(1/2) \leq 1/2$. $\square$

We now show when one can implement a coordinated, mutual-preference consistent strategy with binary communication with the required left tendency of $\alpha = F(0) / (F(0) + (1 - F(1)))$. This implementation requires two things. First, α needs to be a rational number, and second, the message space needs to be sufficiently large.[22] Note that in the case of a symmetric distribution F (i.e., $F(0) = 1 - F(1)$) the required left tendency is exactly $\alpha = 1/2$ and the required strategy is σ_C, as described in Section 3. More generally, let $\alpha = k/n$, and assume that $|M| \geq 2n$. Denote $2n$ distinct messages as $\{m_{L,1}, \ldots, m_{L,n}, m_{R,1}, \ldots, m_{R,n}\} \in M$, where we interpret sending messages $m_{L,i}$ as expressing a preference for L and sending messages $m_{R,i}$ as expressing a preference for R and choosing at random the number i from the set of numbers $\{1, \ldots, n\}$ in the joint lottery described below. We arbitrarily interpret any message $m \in M \setminus \{m_{L,1}, \ldots, m_{L,n}, m_{R,1}, \ldots, m_{R,n}\}$ as equivalent to $m_{L,1}$. Given message $m \in M$, let $i(m)$ denote its associated random number, e.g., $i(m_{L,j}) = j$. Let $M_R = \{m_{R,1}, \ldots, m_{R,n}\}$ and $M_L = M \setminus M_R$. Then $\sigma_\alpha = (\mu_\alpha, \xi_\alpha)$ can be defined as follows:

$$\mu_\alpha(u) = \begin{cases} \frac{1}{n} m_{L,1} + \ldots + \frac{1}{n} m_{L,n} & u \leq \frac{1}{2} \\ \frac{1}{n} m_{R,1} + \ldots + \frac{1}{n} m_{R,n} & u > \frac{1}{2}, \end{cases} \quad \xi_\alpha(m, m') = \begin{cases} 0 & (m, m') \in M_R \times M_R \\ 0 & (m, m') \notin M_L \times M_L \\ & \text{ and } (i(m) + i(m') \mod n) > k \\ 1 & \text{otherwise.} \end{cases}$$

Thus, μ_α induces each agent to reveal whether her preferred outcome is L or R, and to uniformly choose a number between 1 and n. In the second stage, if both agents share the same preferred outcome they play it. Otherwise, moderate types coordinate on L if the sum of their random numbers modulo n is at most k, and coordinate on R otherwise. Extreme types play their strictly dominant action.

[22]The method for implementing a binary joint lottery of α and $1 - \alpha$ is based on Aumann and Maschler (1968) and relies on α being a rational number. In order to deal with irrational α-s one needs either to slightly weaken the result to show that there exists a communication-proof ε-equilibrium strategy (in which each type of each player gains at most ε from deviating) for any $\varepsilon > 0$, or to allow an infinite set of messages or a continuous "sunspot."