Language and Coordination Games

Pei-yu Lo
Brown University

March, 2007

Abstract

Intuitively, if players can communicate, they should be able to reach coordinated play in a coordination game. However, simply adding a communication stage before the play of the game does not render coordination as a unique prediction. To further refine the set of equilibria, Farrell suggested that a self-committing cheap talk statement about one's planned behavior should be believed and thus would for sure lead to coordinated play. Aumann, however, argued that the statement has to be both self-committing and self-signaling for it to guarantee coordination. In this paper, the concept of common knowledge of language is formally incorporated into the cheap talk extension game. This paper shows that, if the stage game satisfies both the self-committing and the self-signaling condition, then every iteratively admissible outcome in the language game constitutes a coordinated play and gives the Sender her Stackelberg outcome. On the other hand, this paper identifies a class of generic games that violate self-signaling condition where every strategy profile of the stage game is an iteratively admissible outcome of the language game. This result supports Aumann's argument that the self-signaling condition is necessary for coordinated play to be guaranteed by one-sided communication.

1 Introduction

This paper applies the idea of common knowledge of language to complete-information games with one-sided communication. There is a debate in the literature over what criterion for a cheap talk statement makes it credible. Farrell (1988) argues that a cheap talk statement about one's planned behavior is credible if it is *self-committing,* that is, if the speaker believes that the statement will be believed, she will have the incentive to carry it out. A self-committing statement should be believed because, if the speaker is sure that it will be believed, the speaker will indeed carry it out. Aumann (1990), on the other hand, argues that self-committing criterion is not enough; a credible cheap talk statement about one's planned behavior has to be *self-signalling* as well, that is, the speaker would want it to be believed only if she indeed plans to carry it out.

The difficulty in formalizing the credibility criterion lies in how to incorporate the strategy of the hypothetical speaker who intends to *not* carry out her statement into the analysis . Baliga and Morris (2002) tackle this problem by expanding the original game into one in which the Sender has private information. In this expanded game, each action that the Sender may take in the original game is the dominant action of one Sender type. Given any claim about planned behavior, every type of Sender whose dominant action is not equal to this claimed action represents a hypothetical speaker who intends *not* to carry out her claim. This transforms the question of when the Sender could credibly transmit information about her intended action into the question of when a fully-separating perfect Bayesian equilibrium exists, i.e. an equilibrium where the informed player fully reveals her type. Since the common prior puts positive weight on every Sender type, the strategy of every Sender type has to be taken into consideration by the Receiver in a perfect Bayesian equilibrium. Baliga and Morris (2002) show that the self-committing condition alone is not sufficient for establishing a credible Sender claim by demonstrating that there is no communication in a class of games which are self-committing but not self-signaling. In this class of games, the Receiver has only two actions. The self-signaling condition is violated in this class of games because the Sender's preference of the Receiver's actions is independent of her own actions.

We notice that every Sender action in the original stage game that is not strictly dominated is associated with a belief about the Receiver's actions, and that every rationalizable Sender action is a best response to a possibly mixed Receiver action. In addition, if the Receiver puts positive weight on every belief that the Sender holds, the Receiver has to take into account the strategy

of every hypothetical Sender with different intentions. Iterative admissibility is a solution concept with this property.

In this expanded game, an instruction is actually a recommendation for the Receiver to take an action in a specified subset. Thus, an opposite instruction is then a recommendation of actions in the complement of that subset. An instruction is more precise if the recommended subset of Receiver actions is smaller. We assume that the language is rich enough to contain every possible sequence of instructions with increasing precision. Two such sequences may share the first several instructions. So, we can think of the common instructions as the common ancestor of the original sequences. Roughly speaking, if the common ancestor of a pair of such messages contains the common ancestor of another pair as a subsequence, we say that the former pair is more similar to each other than the latter pair. With this relationship, we apply two conditions that define the set of strategies consistent with language: (1) literal meaning condition, i.e.: if the Receiver reacts to a message with a specific action, then he reacts with the same action to the related messages that literally recommends that specific action; (2) convexity condition, i.e.: if the Receiver takes the same action after receiving two different messages, then he takes that same action after any message that may have been delivered with some component of the original message. Our language assumption combined with weak dominance enables messages to convey some information about the Sender's preference regarding the actions of the Receiver.

We focus on stage games where the best response correspondences are functions. The stage game is self-signaling when the Sender always prefers the Receiver to take his best response, and the stage game is self-committing if the Receiver should take an action whenever one is recommended, given that the Sender believes that the recommendation will be followed. With these definitions in mind, we find that if the stage game is self-committing and self-signalling, there is a unique iterative admissible outcome of the language game which gives the Sender her Stackelberg payoff. On the other hand, if the stage game is self-committing, but the Sender's preference over the Receiver's actions does not depend her own action, every rationalizable action profile is the outcome of an iterative admissible strategy profile in the language game.

The rest of this paper is structured as follows. Section 2 provides three simple examples to illustrate the role of the self-signalling condition and to motivate the language assumption. Section 3 describes the model and the language assumptions. Section 4 presents our main results described above. Section 4.3 briefly reviews the main results in Baliga and Morris (2002) and

		Receiver's actions	
		Opera	Club
Sender's actions	Opera	2,1	0,0
	Club	0,0	1,2

Table 1: Battle of Sex Game

	"opera"	"club"
Always Opera	Opera	Opera
Always Club	Club	Club
Literal	Opera	Club
Perverse	Club	Opera

Table 2: Receiver's Strategies in Battle-of-the-Sex Game

compares theirs with ours. Section 5 concludes.

2 Motivating Examples

The main idea of this paper is best understood through examples. The battle-of-the-sex game example in section 2.1 illustrates that self-signalling is sufficient to guarantee Stackleberg payoff for the speaker. The investment game example in section 2.2 shows that a severe violation of the self-signalling criterion makes communication ineffective. The partial-common-interest game in section 2.3 motivates the hierarchical messages and language assumptions formally described in section 3.2.

2.1 Coordination without positive spillovers

In the Battle-of-the-Sexes game in table 2.1, there are two Nash equilibria: both go to the Opera and both go to the Club. The Sender prefers the first equilibrium and the Receiver prefers the second. The promise "I will go to the opera" is self-committing because if the Sender believes that the Receiver will believe this statement and play his best response *Opera*, the Sender would prefer to go to the *Opera* and carry out her promise. The promise is self-signalling as well because had the Sender not intended to go to the *Opera*, i.e. had she intended to go to the *Club*, she would prefer the Receiver to go to the *Club* instead of the *Opera* and hence she would not want the Receiver to believe the promise "I will go to the *Opera*".

Suppose $M = \{$"*opera*", "*club*"$\}$. It can be interpreted as a promise to carry out a certain action, or a recommended action for the Receiver. The Sender

sends a message $m \in M$, and then plays an action $a^S \in A^S$ in the stage game. The set of strategies for the Sender is thus

$$S^S := \left\{ \begin{array}{l} (\text{“}Opera\text{”}, Opera), (\text{“}Club\text{”}, Club), \\ (\text{“}Opera\text{”}, Club), (\text{“}Club\text{”}, Opera) \end{array} \right\},$$

while the set of strategies for the Receiver is listed in table 2.1. Both the *Always Opera* and *Always Club* strategies ignore the messages completely. *Literal* strategy and *Opposite* strategy both respond to a message by going to the *Opera* and the other message by going to the *Club*, and hence are essentially the same up to their renaming.

In addition, We can see from table 2.1 that message "*opera*" and message "*club*" are complete symmetric in the sense that if we swap the names of these two messages, we end up with exactly the same strategy set S^R as in table 2.1. This should not be surprising because in traditional economic models of communication, messages have no inherent meanings — the meaning is determined by the equilibrium.

However, the idea that messages have no inherent meanings is counter-intuitive. If the Receiver does respond differently to the two messages "*opera*" and "*club*," it's generally common knowledge how he is going to respond. Suppose, the Sender says the messages in a very sincere and literal way, it is natural that if the Receiver responds differently to different messages, he will use the *Literal* strategy, not the *Opposite* strategy. Suppose, to the contrary, the Sender says "You'd better go to the Opera" in a sarcastic way. If this sarcasm is commonly understood by the Sender and the Receiver, possibly through the tone or the gesture, then it is natural that there is common knowledge that the Receiver would use the *Opposite* strategy if he decides to respond differently to the two different messages.

If we assume that the two players are both native English speakers and come from the same cultural background, and thus they perfectly understand the meaning that the other person tries to convey from the words uttered, the tone, and the body language, then it is without loss of generality to consider only the sincere tone. Suppose it is common knowledge that the Receiver follows the convention of language and never uses *Opposite*. We are thus describing a different game which I call the language game G_L, where the set of strategies for the Receiver is

$$S_L^R := \{Always\ Opera, Always\ Club, Literal\}.$$

We will show that the unique outcome that survive three rounds of deletion of weakly dominated strategies is for both players to go to the *Opera*.

In the first round of deletion of weakly dominated strategies, sending the message "opera" and going to the $Club$ is weakly dominated for the Sender by sending the message "*club*" and going to the $Club$. This is because if the Sender is going to the $Club$, she prefers the Receiver to go to the $Club$. If what the Sender says affect what the Receiver does, she gets her preferred action only if she says "*club*." Likewise, the strategy ("$\{Club\}$", $Opera$) is weakly dominated for the Sender by the strategy ("$\{Club\}$", $Opera$).

Therefore, in the second round of deletion of weakly dominated strategies, the strategy *Always Opera* is weakly dominated by the strategy *Literal*, and the strategy *Always Club* is weakly dominated by *Literal* strategy. The only Receiver strategy that survives the second round is thus the *Literal* strategy.

In the third round, the Sender knows that if she says "*club*," the Receiver will go to the $Club$ and thus it's best for her to go to the $Club$, and if she says "opera", the Receiver will go to the $Opera$ and thus it's best for her to go to the $Opera$ as well. Since she likes $(Opera, Opera)$ better than $(Club, Club)$, the optimal strategy for her is to say "opera" and go the the $Opera$. Thus, we obtain the unique outcome that they coordinate on the Sender's preferred equilibrium.

2.2 Coordination with positive spillovers

To understand the role of the self-signalling criterion, let's look at the Investment game in figure 2.2. As in the Battle-of-the-Sexes game, there are two Nash equilibria in this game: $(Invest, Invest)$ and (Not, Not). The promise "I'm going to invest" is self-committing because if the Sender believes that the Receiver is going to believe the statement and play his best response, it is optimal for the Sender to carry out the promise and play the strategy $Invest$. In Farrell's point of view, this message is thus credible and should be believed. Aumann argues that this promise is not self-signalling and hence is not credible. Even if the Sender intends to play Not, possibly due to lack of confidence that Receiver is really going to $Invest$, she still prefers the Receiver to use the strategy $Invest$. Therefore, she would like the Receiver to believe her promise regardless of her intended action. If she is pessimistic about the effect of communication and believes that, with high probability, the Receiver is going to $Not\ Invest$ regardless of what she says, then she would prefer to $Not\ Invest$. However, if the probability that the Receiver uses the strategy $Invest$ is higher after hearing the promise "I'm going to invest", the Sender would like to make that promise even though she does not intend to carry it out.

Let's look at the cheap talk extension game in detail. Suppose $M =$

		Receiver	
		$Invest$	$NotInvest$
Sender's actions	$Invest$	$2, 2$	$-1, 1$
	$NotInvest$	$1, -1$	$0, 0$

Table 3: Investment Game

	"*invest*"	"*not*"
Always Invest	Invest	Invest
Never Invest	Not	Not
Literal	Invest	Not
Opposite	Not	Invest

Table 4: Receiver's Strategies in Investment Game

{"*invest*", "*not*"}. Then the set of strategies for the Sender is

$$S^S := \{(\text{"}invest\text{"}, Invest), (\text{"}not\text{"}, Not), (\text{"}invest\text{"}, Not), (\text{"}not\text{"}, Invest)\},$$

while the set of strategies for the Receiver is listed in table 2.2. Suppose it is common knowledge that the Receiver follows the language convention and never uses the strategy *Opposite*. In the transformed game G_L, the set of strategies for the Receiver is thus

$$S_L^R := \{Always\ Invest, Never\ Invest, Literal\}.$$

We will now show that every outcome remains after one round of deletion of weakly dominated strategies, when the iterative process stops. Sending the message "*not*" and using the strategy $Invest$ is weakly dominated by sending the message "invest" and using the strategy $Invest$, because when the Sender invests, she prefers the Receiver to invest, and whenever talking affects the outcome, she gets her preferred action only by saying "invest." Since the Sender has the same preference over the Receiver's actions regardless of the action she takes, the same argument shows that ("*not*", Not) is weakly dominated by ("*invest*", Not). Thus, after the first round of deletion, only the message "invest" survives. The process of iterative deletion of weakly dominated strategies stops after the first round, because the Receiver, after receiving the message, still does not know what the Sender is going to play, and thus might play $Never\ Invest$ if he is pessimistic about the Sender's intention, and either *Literal* or *Always Invest* if he is optimistic. After the first round of deletion, the two Receiver strategies, *Literal* and *Always Invest*, are payoff-equivalent for the Receiver because they differ only in the action taken after the message "not", which is reached with probability zero.

Unlike in the Battle-of-the-Sexes game, when there are positive spillovers, pre-game communication does not eliminate strategic uncertainties. These two examples illustrate the role of the self-signalling criterion.

2.3 Partial Common Interest

The Fighting-Couple game in table 5 shows that communication can help players avoid bad equilibria, even though their preferences are not fully aligned. The game has one pure strategy equilibrium: $(Home, Home)$. In one of the two mixed strategy equilibria, both go to the $Opera$ with probability $\frac{1}{2}$ and go to the $Club$ with probability $\frac{1}{2}$. In the other mixed strategy equilibrium, both go to the $Opera$ with probability $\frac{1}{8}$, go to the $Club$ with probability $\frac{1}{8}$ and stay $Home$ with probability $\frac{3}{4}$. The mixed-strategy equilibrium where both staying $Home$ with probability 0 is the efficient one. Both going to the $Opera$ and going to the $Club$ is consistent with going out, as opposed to staying home. One prefers to stay $Home$ if and only if the other stays $Home$. Moreover, avoiding staying $Home$ and restricting themselves to the submatrix $\{Opera, Club\} \times \{Opera, Club\}$ is mutually beneficial for both players.

Now suppose the Sender has an opportunity to leave a voice message before they play the one-shot game in table 5. She cannot possibly persuade the Receiver to go to the $Opera$, nor can she persuade the Receiver to go to the $Club$, because they have conflict of interest regarding the two actions. However, it is self-committing for her to say "you should go out," in the sense that if the Receiver is persuaded and goes out, the Sender will go out, i.e., she will choose to either go to the $Opera$ or go to the $Club$, in which case, the Receiver prefers to go out. In addition, the suggestion "you should go out" is also self-signalling in the sense that the Sender prefers the Receiver to go out only if she plans to go out herself.

Consider the suggestion "Definitely go out tonight, dear. Regarding where to go, you should go to the opera." We can write this suggestion as a 2-sequence of decreasing subset: $\{Opera, Club\}\{Opera\}$. The suggestion "Definitely go out tonight, dear. Regarding where to go, you should go to the club" is slightly different from the previous one. Another possible suggestion, "You should stay home," on the other hand, is drastically different from the previous two. We can write this message as "$\{Home\}$". If the Receiver plays the same action after receiving both suggestions "$\{Opera, Club\}\{Opera\}$" and "$\{Home\}$", then the Receiver ignores the first layer of literal distinction between going out and staying home. Intuitively, the fine literal difference between the two messages "$\{Opera, Club\}\{Opera\}$" and "$\{Opera, Club\}\{Club\}$" should also

be ignored by the Receiver. That is, the Receiver should play exactly the same action after receiving the message "$\{Opera, Club\}\{Club\}$", the messages "$\{Opera, Club\}\{Opera\}$" and "$\{Home\}$". Suppose the set of messages is

$$M = \{\text{"}\{Opera, Club\}\{Opera\}\text{"}, \text{"}\{Opera, Club\}\{Club\}\text{"}, \text{"}\{Fight\}\text{"}\}.$$

Then the preceding discussion suggests the type of language assumption that restricts the Receiver's strategies to those in table 6.

We will now show that, in the language game G_L, no player uses the action $Home$ after three rounds of deletion of weakly dominated strategies. In the first round of deletion, the strategy ("$\{Opera, Club\}\{Opera\}$", $Home$) is weakly dominated by ("$\{Home\}$", $Home$) for the Sender. In addition, Both

$$(\text{"}\{Opera, Club\}\{Opera\}\text{"}, Opera)$$

and ("$\{Home\}$", $Opera$) are weakly dominated by the Sender strategy

$$(\text{"}\{Opera, Club\}\{Club\}\text{"}, Opera).$$

Therefore, after the first round of deletion of weakly dominated strategies, if the Sender suggests any non-violent action (either $Opera$ or $Club$), she definitely plans to not fight; if the Sender suggests to fight, she definitely plans to fight. Thus, it is weakly dominated for the Receiver to fight after a non-violent suggestion; it is also weakly dominated for the Receiver to play a non-violent action after the suggestion to fight. However, it is also weakly dominated in the second round of deletion for the Receiver to play the *Completely Literal* strategy, because they have opposing interests when restricting the game to $\{Opera, Club\} \times \{Opera, Club\}$. It can be easily checked that the set of strategies that survive the second round of deletion of weakly dominated strategies is thus$\{Opera\ Home, Club\ Home\}$. Therefore, the Sender can be guaranteed a non-violent response if she says either "$\{Opera, Club\}\{Opera\}$" or "$\{Opera, club\}\{Club\}$". Since she prefer any outcome in the submatrix

$$\{Opera, Club\} \times \{Opera, Club\}$$

to anyone outside of that matrix, the strategy ("$\{Fight\}$", $Fight$) is strictly dominated in the third round.

The strategy set for the Sender that survives iterative admissibility is

$$\{(\text{"}\{Opera, Club\}\{Opera\}\text{"}, Club), (\text{"}\{Opera, Club\}\{Club\}\text{"}, Opera)\},$$

while the strategy set for the Receiver that survives iterative admissibility is

$$\{Opera\ Home,\ Club\ Home\}.$$

Pregame communication guarantees both players a payoff of at least 2.

		Receiver's actions		
		Opera	*Club*	*Home*
Sender's actions	*Opera*	2,4	4,2	0,0
	Club	4,2	2,4	0,0
	Home	0,0	0,0	1,1

Table 5: Fighting-Couple Game

	"{opera,club} {opera}"	"{opera,club} {club}"	"{home}"
Always Opera	Opera	Opera	Opera
Always Club	Club	Club	Club
Always Home	Home	Home	Home
Opera & Home	Opera	Opera	Home
Club & Home	Club	Club	Home
Completely Literal	Opera	Club	Home

Table 6: Receiver's Strategies in the Fighting Couple Game

3 The Model

In this paper, we focus on one-sided communication extension to finite two-player games with complete information. The Sender (S) and the Receiver (R) simultaneously choose an action a^S, a^R from a finite set A^S and A^R respectively. Their payoffs are given by $g^S : A^S \times A^R \to R$ and $g^R : A^S \times A^R \to R$. Write $g = \left(g^S, g^R\right)$. We will abuse the notation and denote the stage game also by g. In the one-sided cheap talk extension game G, the Sender sends a message from a finite set M before they play the stage game g. A strategy for the Sender in the reduced-form cheap talk extension game G, denoted by s^S, is a message $m \in M$ and an action $a^S \in A^S$. A strategy for the Receiver in G, denoted by s^R, is a mapping from M to A^R. To characterize the set of communication outcomes with an existing language, we first transform the cheap talk game G into the language game G_L by directly restricting the set of strategies for the Receiver to S_L^R. We motivate and describe the restricted set of Receiver strategies in section 3.2. We then characterize the set of iteratively admissible outcomes in the language game G_L.

We focus on games where changing exactly one player's action in the action profile changes the payoff. This assumption will be carried throughout the paper.

Assumption The stage game payoff $g^i : A^S \times A^R \to R$ is such that

$$\begin{aligned} g^i\left(a^S, a^R\right) &\neq g^i\left(a'^S, a^R\right); \\ g^i\left(a^S, a^R\right) &\neq g^i\left(a^S, a'^R\right) \end{aligned}$$

for any $a^S \neq a'^S$, $a^R \neq a'^R$, and $i = S, R$.

This assumption implies that, in particular, the stage-game best response correspondences for both players are well-defined functions. This condition is weaker than genericity, which is a common assumption, and does not exclude any of the motivating games in section 2.

We denote the stage-game best response functions by $b^i : A^i \to A^j$ where $i, j \in \{S, R\}$ and $i \neq j$. When we mention "best response" later in the paper, it refers to best response in the language game. Let $X^S \times X^R$ be a subset of $S^S \times S_L^R$, and $i, j \in \{S, R\}$ where $i \neq j$. Two strategies s_1^i and s_2^i are equivalent w.r.t. X^j if they give the same payoff to i under any of j's strategies that belong to X^j. An observation that will be useful later is that, if every Receiver strategy in X^R is constant on a message subset E, then no two Sender strategies that both use messages in E but take different actions can be equivalent w.r.t. X^R. A strategy s^i is said to be a strict best response to σ^j w.r.t. $X^S \times X^R$ if

$$U^i\left(s^i, \sigma^j\right) \geq U^i\left(\tilde{s}^i, \sigma^j\right)$$

for every $\tilde{s}^i \in X^i$, and strict inequality holds for every $\tilde{s}^i$ which is not equivalent to s^i w.r.t. X^j.

3.1 Solution Concept

For ease of exposition, we rewrite here the definition of the solution concept of iterative admissibility taken from Brandenburger et al (2004).

Definition 1 *Fix $\left(X^j\right)_{j \in I} \subseteq \left(S^j\right)_{j \in I}$. A strategy s^i is weakly dominated with respect to X^{-i} if there exists $\hat{\sigma}^i \in \Delta X^i$ such that $U^i\left(\hat{\sigma}^i, s^{-i}\right) \geq U^i\left(s^i, s^{-i}\right)$ for every $s^{-i} \in X^{-i}$ and that $U^i\left(\hat{\sigma}^i, \hat{s}^{-i}\right) > U^i\left(s^i, \hat{s}^{-i}\right)$ for some $\hat{s}^{-i} \in X^{-i}$. Otherwise, say that s^i is admissible with respect to $\left(X^j\right)_{j \in I}$. If s^i is admissible w.r.t. $\left(S^j\right)_{j \in I}$, simply say that s^i is admissible.*

Definition 2 *Set $S^i(0) = S^i$ for $i \in I$ and iteratively define*

$$S^i(k+1) = \left\{ \begin{array}{c} s^i \in S^i(k): \\ s^i \text{ is not weakly dominated with respect to } \left(S^i(k)\right)_{i \in I} \end{array} \right\}.$$

Write $\cap_{k=0}^{\infty} S^i(k) = S^i(\infty)$ and $\cap_{k=0}^{\infty} S(k) = S(\infty)$. A strategy $s^i \in S^i(\infty)$ is called iteratively admissible.

Denote by ΔX the set of probability distribution on X, and by $\Delta^+ X$ the set of probability distribution which puts positive weight on every element of X.

Brandenburger et al (2004) show that if there are only two players, say player S and player R, a strategy is weakly dominated if and only if it is never a best response to a totally mixed strategy. For completeness of arguments, this equivalence result is restated as Lemma 1 below.

Lemma 1 (Brandenburger et al (2004)) *A strategy $\hat{s}^i \in X^R$ where $i \in \{S, R\}$ is admissible with respect to $X^S \times X^R$ if and only if there exists $\hat{\sigma}^j \in \Delta^+ S^j$ where $j \neq i$ such that $U^R\left(\hat{\sigma}^S, \hat{s}^R\right) \geq U^R\left(\hat{\sigma}^S, s^R\right)$ for every $s^R \in X^R$.*

For our purpose we strenghten the charaterization of admissible strategies as follows.

Corollary 1 *A strategy $\hat{s}^i \in X^R$ where $i \in \{S, R\}$ is admissible with respect to $X^S \times X^R$ if only if there exists $\hat{\sigma}^j \in \Delta^+ S^j$ where $j \neq i$ to which $\hat{s}^i$ is a strict best response w.r.t. $X^S \times X^R$.*

3.2 Incorporating Language

Consider a language L. Suppose L contains an expression for a subset of Receiver actions B. Denote this expression by ξ_0. If L also contains an expression for logical negation "not," then L contains an expression for the idea "do not do B." Denote this expression by ξ_1. In another language L', the expression ξ_0 may mean "please do B," while the expression ξ_1 may mean "please do not do B." Since messages are costless and are only means to convey information, it does not matter which language the Sender and the Receiver are speaking, as long as it is common knowledge that they speak the same language. Suppose the common language that the Sender and the Receiver speak is L. If the Receiver decides to ignore the Sender's messages, whatever the Sender says does not matter and the Receiver takes the same action regardless. If the Receiver decides to respond to ξ_0 and ξ_1 differently because he thinks the Sender conveys information through her messages, he refers to his own knowledge L and responds to message ξ_0 with an action in B while to message ξ_1 with an action not in B.[1]

[1] Some may argue that the Receiver would want to take the *Opposite* strategy in a matching penny game. However, if the Sender knows the payoff structure of a matching penny game, and if she knows that the Receiver uses language L' in a matching penny game, she will give recommendations according to L', thereby destroying the incentive for the Receiver to use language L'. In this case, one may argue that the Receiver randomizes his actions after

In the Battle-of-the-Sexes game, if we let B refer to going to the opera, then "not B," i.e. "not go to the opera," is equivalent to "go to the club," since this is the only choice other than going to the opera. In the Fighting-Couple game, the expression, "go out", is saying exactly the same thing as, "go to the opera or go to the club", and the expression, "do not go out", says the same thing as, "go home". If the Receiver responds differently to the two recommendations, "go out", and, "go home", then we see from previous discussion that he responds to, "go out", by going out. However, the Receiver still has to decide whether to go to the opera or the club. Carrying this idea forward, lets suppose that the subset of Receiver actions B contains a strict subset B_2, and the language L contains an expression for B_2. Suppose further that L contains an expression ξ_{00} that is simply a concatenation of ξ_0 and the expression for B_2. Then with the expression for logical negation, L contains an expression ξ_{01} which is the concatenation of ξ_0 meaning ,"do not do B", and the expression for, "within B_1, do not do B_2". That the receiver may decide that the messages, "go out", and, "go home", convey separate information, but decide to ignore the finer differences between, "go out; furthermore, go to the opera", and "go out; and then go to the club'. Then the Receiver takes the same action after receiving both message ξ_{00} and ξ_{01}. However, if the Receiver decides not to ignore the finer difference between the recommendations ξ_{00} and ξ_{01}, he refers to his knowledge of L and responds to message ξ_{00} with action B_2 and to message ξ_{01} with an action in B_1 but not in B_2.

Let M denote every message that the Sender could possibly utter. We also assume that the language L the players commonly speak contains an expression for every subset of Receiver actions, an expression for logical negation, and an expression for concatenation. Then, the language L contains an expression for every strictly decreasing sequence of subsets of Receiver actions $A_1A_2...A_n$. As a convention, let $A_0 = A^R$ and $A_{n+1} = \emptyset$. Each sequence can be seen as a sequence of instructions with finer and finer details. The set of all such sequences where the last subset has only one element is called the set of hierarchical

receiving a message. This could be achieved by randomizing between the *Always B* strategy, and the *Never B* strategy.

Some also argue that the Receiver plays the *Literal* and *Opposite* strategies at the same time. This argument is supported by observing the game being played many times. Throughout these observations, there are incidents where the Receiver takes action B after message ξ_0 and after message ξ_1. There are also incidents where the Receiver takes action not in B after message ξ_0 and after message ξ_1. These observations do not refute the hypothesis that the Receiver does not play the *Opposite* strategy because all of the aforementioned outcomes may be realizations of a Receiver strategy that randomizes between *Always B* and *Never B*. Finally, in a matching penny game, the Receiver actually has no incentive to respond differently with the Sender's messages, because he knows that the Sender will not convey any information about her intention.

recommendations, denoted by M^h. Given a hierarchical recommendation $m = A_1...A_n$, we call A_j the j^{th} level of instruction. Define $M(A_1...A_j)$ to be the set of all messages that start with the strictly decreasing sequence $A_1...A_j$. Every message m in $M(A_1...A_j)$ express the same idea of "Do A_1. Further more, take an action in A_2. ....To be even more precise, do A_j".

Let s^R be a mapping from the set of messages M to the set of Receiver actions A^R, and $m = A_1...A_n$ a hierarchical recommendation. Let $a^R = s^R(m)$. Let γ be the highest level of instruction the action a^R is consistent with according to m. That is, $a^R \in A_\gamma \backslash A_{\gamma+1}$. Therefore, within the subset of A_γ, the action a^R is "opposite to" the instruction of $A_{\gamma+1}$. Our previous discussion suggests that, if s^R is a language-based Receiver strategy, then either s^R takes the same action after both expressions for $A_1...A_\gamma(A_{\gamma+1})$ and expressions for $A_1...A_\gamma(A_\gamma \backslash A_{\gamma+1})$, or s^R responds to expressions for $A_1...A_\gamma(A_{\gamma+1})$ with actions in $A_{\gamma+1}$ and expressions for $A_1...A_\gamma(A_\gamma \backslash A_{\gamma+1})$ with actions in $A_\gamma \backslash A_{\gamma+1}$. Therefore, if s^R is language-based and $s^R(m) \in A_\gamma \backslash A_{\gamma+1}$, s^R must ignore differences between $A_{\gamma+1}$ and $A_\gamma \backslash A_{\gamma+1}$, and takes the same action after both expressions for $A_1...A_\gamma(A_{\gamma+1})$ and expressions for $A_1...A_\gamma(A_\gamma \backslash A_{\gamma+1})$. That is, s^R takes the "opposite" action to the instruction $A_{\gamma+1}$. The preceding discussion suggests that, if s^R is a language-based Receiver strategy, then

$$s^R(m') = s^R(m)$$

for every message $m' \in M(A_1...A_\gamma A_{\gamma+1}) \cup M(A_1...A_\gamma(A_\gamma \backslash A_{\gamma+1}))$. Call the set

$$M(A_1...A_\gamma A_{\gamma+1}) \cup M(A_1...A_\gamma(A_\gamma \backslash A_{\gamma+1}))$$

the constrained message set given message m and action a^R. Formally, given $m \in M$ and $a^R \in A^R$, define

$$M^{cstr}(m, a^R) \equiv \begin{cases} M(A_1...A_\gamma A_{\gamma+1}) & \text{if } m \in M^h \text{ and } \gamma \text{ such that} \\ \cup M(A_1...A_\gamma(A_\gamma \backslash A_{\gamma+1})) & a^R \in A_\gamma \backslash A_{\gamma+1} \\ m & \text{otherwise} \end{cases}.$$

Now we formally define our language assumptions.

Definition 3 *$s^R : M \to A$ is a language-based Receiver strategy, denoted by $s^R \in S_L^R$, if and only if s^R is constant on $M^{cstr}(m, s^R(m))$, for every $m \in M$.*

This definition is best illustrated with graphs. Suppose $A^R = \{A, B, C, D\}$. Figure 2 shows some hierarchical recommendations in this game. There are many different ways to group A^R. The first level of instruction can be about taking action D or not taking action D, as shown by the two branches $\{D\}$

and $\{A, B, C\}$ that diverge from each other. The first layer of instruction can also tell you whether to take actions in $\{A, C\}$ or not, as shown by the two branches $\{A, C\}$ and $\{B, D\}$. Given a first-layer instruction $\{A, B, C\}$, the second layer of instruction could be about whether to take action A or not, as shown by the two branches $\{A\}$ and $\{B, C\}$ that diverge one node on the branch of $\{A, B, C\}$. In general , expressions that are "opposite to" each other at some level of instruction are drawn to diverge from the same node. We call all the messages that diverge from the same node a message bundle. A message branch $A_1...A_j$, on the other hand, consists of all messages that start with instructions $A_1...A_j$. For example, all the messages in the circle in figure 2 constitutes message branch $\{A, B, C\}$, while the message "$\{D\}$" combined with all messages in the red circle constitute a message bundle. There can be several parallel message bundles on a branch, which represent different ways to subdivide the set of Receiver actions relevant for the branch.

Suppose we choose the branch of $\{A, B, C\}$ and then choose $\{B\}$, we end up with the message "$\{A, B, C\}\{B\}$." The set $M(\{A, B, C\}\{A, C\})$ consists of two messages: "$\{A, B, C\}\{A, C\}\{A\}$" and "$\{A, B, C\}\{A, C\}\{C\}$." Within the broad instruction $\{A, B, C\}$, these two messages are both "opposite to" message "$\{A, B, C\}\{B\}$."

Given message $\{A, B, C\}\{B\}$ and Receiver action C, we first find that action C belongs to the subset $\{A, B, C\}$ but not to the subset $\{B\}$. According to the definition, the constrained message set given m and a^R is thus the following set of messages:

$$\{\text{"}\{A, B, C\}\{B\}\text{"}, \text{"}\{A, B, C\}\{A, C\}\{A\}\text{"}, \text{"}\{A, B, C\}\{A, C\}\{C\}\text{"}\}.$$

If a language-based Receiver strategy s^R responds to message "$\{A, B, C\}\{B\}$" with action C, then by definition, s^R takes action C after receiving message $\{A, B, C\}\{B\}$," "$\{A, B, C\}\{A, C\}\{A\}$" and "$\{A, B, C\}\{A, C\}\{C\}$."

We call $M^{cstr}\left(m, b^R\left(a^S\right)\right)$ the constrained message set of the pure Sender strategy $\left(m, a^S\right)$. Every message in $M^{cstr}\left(m, b^R\left(a^S\right)\right)$ recommends actions in A_j which includes the Receiver's best response to a^S. Let $\left(m_1, a_1^S\right)$ and $\left(m_2, a_2^S\right)$ be two Sender strategies with overlapping constrained message sets where $b^R\left(a_1^S\right) \neq b^R\left(a_2^S\right)$. It is esay to see that the constrained message set of one Sender strategy must contain that of the other. Let A_j be the last common level of instruction in the larger of the two constrained message sets. In a way, both message m_1 and m_2 share the recoomendation of A_j. If we see all messages in the larger constrained message set as one compound message, then these two Sender strategies essentially use the same message for different

intentions. It is immediate from the definition of constrained message set that no Receiver strategy consistent with language can be a best response to both $\left(m_1, a_1^S\right)$ and $\left(m_2, a_2^S\right)$.

On the other hand, if $\left(m_1, a_1^S\right)$ and $\left(m_2, a_2^S\right)$ has disjoint constrained message set and $b^R\left(a_1^S\right) \neq b^R\left(a_2^S\right)$, given any Receiver strategy in a certain class, we can modify it in a minimal way so that it is a best response to both $\left(m_1, a_1^S\right)$ and $\left(m_2, a_2^S\right)$.

Lemma 2 *Let $\left(m_1, a_1^S\right)$ and $\left(m_2, a_2^S\right)$ be two Sender strategies in $S^S(j)$ with disjoint constrained message set where $b^R\left(a_1^S\right) \neq b^R\left(a_2^S\right)$. Let E and B be the smallest message branch and message bundle respectively, containing the constrained message sets of both Sender strategies. Then*

1. *there exists a Receiver strategy in $S^R(j+1)$ non-constant on B, and*
2. *there exists a mapping $\psi_B : S^R(j) \to S^R(j)$ such that $\psi_B\left(s^R\right) = s^R$ for every $s^R \in S^R(j)$ constant on B, while for $s^R \in S^R(j)$ non-constant on B, $\psi_B\left(s^R\right)$ is equal to s^R outside of E and $\psi_B\left(s^R\right)$ responds to message m_i with $b^R\left(a_i^S\right)$, $i = 1, 2$.*

4 Results

In this section we generalize the intuition gained from the contrast between the Battle-of-the-Sex game and the Investment game. Section 4.1 gives sufficient conditions for one-sided pre-game communication to guarantee coordinated play in a coordination game. Section 4.2 shows that, when the Sender's preference over the Receiver's actions is independent of the Sender's own action, every rationalizable outcome in the stage game is an iteratively admissible outcome of the language game.

4.1 A Sufficient Condition to Guarantee Stackelberg Payoff for the Sender

In Farrell's definition, messages are about intended actions. In this chapter, we focus on messages that serve as recommendations of actions to the Receiver. We can easily translate a message about the speaker's intended action into a recommendation for the Receiver, since the payoff matrix of the stage game is common knowledge, and thus the Receiver can infer from the speaker's claim about her intended action what the speaker wants the Receiver to do. For

example, the message, "I will take action a^S", is equivalent to a recommendation for the Receiver to take his best response to a^S.

Let b^i denote the best reply correspondence for player i in the stage game, $i = S, R$. Since we focus on games where changing only one player's action changes both players' payoff, the aforementioned best response correspondence b^i is in fact a function.

For ease of comparison, we re-write the formal definition of the condition of self-committing by Baliga and Morris (2002) in the following. We then give our version of the definition.

Definition 4 (Baliga and Morris (2002)) *Claim about intended action a^S is self-committing if $b^S\left(b^R\left(a^S\right)\right) = a^S$.*

Definition 5 *Recommendation $a^R \in A^R$ is self-committing if $b^R\left(b^S\left(a^R\right)\right) = a^R$.*

Definition 6 *The stage game g is self-committing if every recommendation $a^R \in A^R$ is self-committing.*

It is straightforward to see that the recommendation a^R is self-committing if and only if the claim about intended action $b^S\left(a^R\right)$ is self-committing because $b^S\left(b^R\left(b^S\left(a^R\right)\right)\right) = b^S\left(a^R\right)$.

The definition Aumann (1990) gives for self-signalling criterion is as follows. A statement is self-signalling if the speaker would want it to be believed only if it is true. We can thus say that a recommendation is self-signalling if the speaker would want it to be followed only if she plans to take the action which makes the recommendation optimal for the Receiver. This definition implies that the speaker would NOT want her recommendation $b^R\left(a^S\right)$ to be followed if her planned action would not make this recommendation optimal for the Receiver, that is, if she planned to take an action different from a^S. This suggests that the self-signalling condition is a property on the stage game as whole, not one about individual actions.

Baliga and Morris formalizes the definition as follows.

Definition 7 (Baliga and Morris (2002)) *The game g is self-signalling (for the Sender) if $g^S\left(a^S, b^R\left(a^S\right)\right) > g^S\left(a^S, a^R\right)$ for every $a^S \in A^S$, and $a^R \in A^R$ where $a^R \neq b^R\left(a^S\right)$.*

The following proposition gives a sufficient condition for the Sender to be guaranteed her Stackelberg payoff.

		Receiver's actions		
		A	B	C
Sender's actions	a	$2,3$	$1,2$	$-1,-9$
	b	$0,0$	$4,3$	$-1,2$
	c	$1,-9$	$2,2$	$3,3$

Table 7: A Stage Game with Three Receiver Actions

Proposition 1 *If the stage game g is self-signalling and self-committing, then any strategy profile $\left(\left(m,a^S\right),s^R\right)$ that survives iterative deletion of weakly dominated strategies in the language game gives the Sender her Stackelberg payoff, that is,*

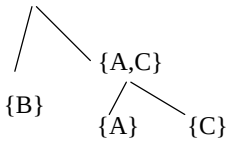
$$\begin{aligned} & g^S\left(a^S,a^R\right) \\ = & \max_{a^S} u^S\left(a^S,b^R\left(a^S\right)\right) \end{aligned}$$

for every $\left(\left(m,a^S\right),s^R\right)\in S_L\left(\infty\right)$.

4.1.1 An Example

To see the main idea behind the proof, it is easy to start with a simple example. The game is shown in table 7. This game is generic and has three pure strategy Nash-equilibria: $(a,A),(b,B)$ and (c,C). It is obvious that every recommendation is self-committing, and this game is self-signaling.

For ease of exposition, we will assume that the Sender can only give hierarchical recommendations that start with either "$\{B\}$" or "$\{A,C\}$." The very top of figure 1 lists every such message. The bottom table in figure 1 lists every Sender strategy that survives the first, third, and fifth round of deletion of weakly dominated strategies. Action a is listed in the cell at the intersection of row $S^S(1)$ and column "$\{A,C\}\{A\}$", while action b and c is not listed in that cell. This indicates that taking action a after sending the recommendation "$\{A,C\}\{A\}$" survives the first round of deletion of weakly dominated strategies, while taking action b or action c after sending the recommendation "$\{A,C\}\{A\}$" does not. The table in the middle of figure 1 lists Receiver strategies that survive the 0^{th}, the second and the fourth round of deletion of weakly dominated strategies. For example, the Receiver strategy, *First Layer and A*, shown in the fourth row in the right panel of figure 1, responds to message "$\{B\}$" with action B, and to both message "$\{A,C\}\{A\}$" and message "$\{A,C\}\{C\}$" with action A. By definition, every language-based Receiver strategy survives the 0^{th} round of deletion of weakly dominated strategies in the language game.



Receiver strategies	R $S(0)$	Always A	A	A	A
		Always B	B	B	B
		Always C	C	C	C
		First Layer and A	B	A	A
		First Layer and C	B	C	C
		Completely Literal	B	A	C
	R $S(2)$	Always A	A	A	A
		Always C	C	C	C
		Completely Literal	B	A	C
	R $S(4)$	Completely Literal	B	A	C

Sender strategies	S $S(1)$	a, b, c	c	a
	S $S(3)$	b	a	c
	S $S(5)$	b		

Figure 1: The Iterative Process for a Game with Three Receiver Actions

The Sender strategy that takes action c after sending the recommendation "$\{A, C\}\{A\}$" does not survive the first round of elimination for the Sender because by the self-signalling condition, the Sender prefers the Receiver action C to any other Receiver action if the Sender is going to take action c, and thus taking action c after sending the recommendation "$\{A, C\}\{A\}$" is weakly dominated by taking the same action c while sending the recommendation "$\{A, C\}\{C\}$." It can be shown in a similar way that taking action b after sending the recommendation "$\{A, C\}\{A\}$" is weakly dominated by the Sender strategy that takes action b but sends the recommendation "$\{B\}$."

However, it is not the case that every Sender strategy that takes an action which makes the recommendation suboptimal for the Receiver is weakly dominated in the first round of deletion. For example, taking action c after giving the recommendation "$\{B\}$" is not weakly dominated in the first round because it is the best response to the belief

$$(1-\varepsilon)\, Always\ C + \varepsilon\,(1-\varepsilon)\, First\ Layer\ and\ A + \varepsilon^2 \sigma^R$$

for ε sufficiently small and any totally mixed Receiver strategy σ^R.

Proceeding to the second round of deletion, *First Layer and B*, is weakly dominated by the Receiver strategy, *Literal*, in the second round, because 1) these two strategies are not equivalent since they differ only in their response to message "$\{A, C\}\{A\}$" and message "$\{A, C\}\{A\}$" is used by a Sender strategy in $S^S(1)$, and 2) every Sender strategy in $S^S(1)$ that uses message "$\{A, C\}\{A\}$" involves taking action a, and Receiver action A does strictly better than Receiver action C given that the Sender plays action a.

In the third round of deletion, the Sender strategy ("$\{B\}$", c) is weakly dominated by the Sender strategy ("$\{A, C\}\{C\}$", c) because any Receiver strategy surviving the second round of deletion that takes different action after the two messages responds to message "$\{A, C\}\{C\}$" with action C, which is most preferred by the Sender given her action c.

The Sender strategy ("$\{B\}$", b) survives the third round of deletion because it is a best response to the Receiver strategy

$$(1-\varepsilon)\, Completely\ Literal + \varepsilon^2 \sigma^R$$

where ε is very small and σ^R is any totally mixed Receiver strategy in $S^R(2)$. This in turn leads to the deletion of Receiver strategies *Always C* and *Always A* in the fourth round by *Completely Literal*.

It follows that, after four rounds of deletion of weakly dominated strategies, the message "$\{B\}$" will certainly induce action B. Since the strategy profile (b, B) gives the Sender her highest payoff, the Sender strategy that sends

message "$\{B\}$" and takes action b strictly dominates any other Sender strategy remaining after four rounds of deletion of weakly dominated strategies. Therefore, the unique outcome surviving iterative deletion of weakly dominated strategies gives the Sender her Stackelberg payoff.

This example illustrates two points. First, Sender strategies which use a message whose ultimate recommended action is not optimal for the Receiver given the Sender's intention may still survive the first round of deletion, even though the stage game is a pure coordination game. One such Sender strategy in this particular example is ("$\{B\}$", c). This is because the Sender is afraid that the Receiver may follow those layered recommendations only halfway. Once those Receiver strategies that follow recommendations like "$\{A, C\}\{C\}$" halfway are eliminated, those Sender strategies that do not recommend the Receiver's best response to the Sender's intention may subsequently have a chance to be eliminated. Second, to continue the iterative process and eliminate those Sender strategies, we need to show that a Sender strategy that serves as a dominator remains when those Receiver strategies that follow only halfway are eliminated.

4.1.2 The Proof

Denote the size of the set of Receiver strategies A^R by N. We can arbitrarily order Receiver actions and write

$$A^R = \left\{a_1^R, ..., a_N^R\right\}.$$

Define a_i^S to be $b^S\left(a_i^R\right)$. Let ϕ denote a permutation of $\{1, 2, ..., N\}$ and Φ the set of all permutations of $\{1, ..., N\}$, with id being the identity permutation. Define

$$m_{\phi(N-k)} = A_1 ... A_{N-k-1} \left\{a_{\phi(N-k)}^R\right\}$$

where $A_j = A_{j-1} \backslash \left\{a_{\phi(j)}^R\right\}$ for $j = 1, ..., N-k-1$. Define $M_{\phi(N-k)}$ to be the set of hierarchical messages that share in common the first $N-k-1$ levels of instruction which eliminates one action at a time from the previous level according to ϕ. Formally, define

$$M_{\phi(N-k)} := \left\{ \begin{array}{c} m = A_1 ... A_{N-k} A_{N-k+1} ... A_n | \\ A_j \supsetneqq A_{j+1}, \forall j = 1, ..., n-1; \\ A_j = A_{j-1} \backslash \left\{a_{\phi(j)}^R\right\} \ \forall j = 1, ..., N-k \\ n \geq N-k \end{array} \right\}.$$

Figure 2 shows a partial set of hierarchical recommendations. Let ϕ be such that $a_{\phi(1)}^R = D$, $a_{\phi(2)}^R = B$, $a_{\phi(3)}^R = C$ and $a_{\phi(4)}^R = D$. Then the messages in

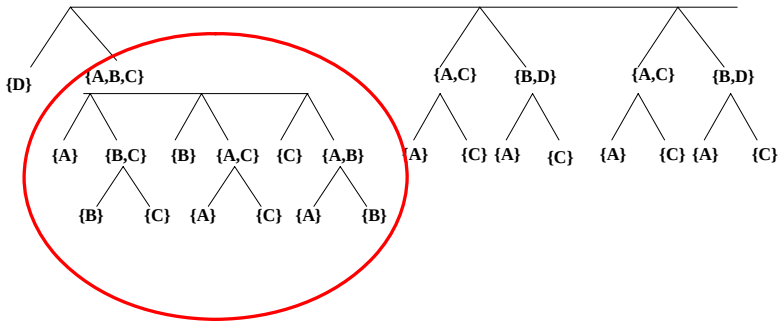


Figure 2: Partial Set of Hierarchical Recommendations, $A^R = \{A, B, C, D\}$.

the circle of figure 2 constitute the set $M_{\phi(1)}$, while the message "$\{D\}$" is the message $m_{\phi(1)}$.

The following observation basically says that, if the Sender intends to take some action $\hat{a}^S$, then it is weakly dominated for the Sedner to use a message that has, through a process of one-by-one exclusion, recommended the Receiver NOT to take his best response to $\hat{a}^S$. For example, it implies that taking action d, to which the Receiver's best response is D, and sending a message in the circle in figure 2 is weakly dominated in the first round by the Sender strategy that takes action d but uses the message "$\{D\}$."

Observation Given a Sender action a^S and a permutation ϕ and any message $m \in M_{\phi(q)}$ where q is such that $b^R\left(a^S\right) = a^R_{\phi(q)}$, then the Sender strategy $\left(m, a^S\right)$ is weakly dominated in the first round by the Sender strategy $\left(m_{\phi(q)}, a^S\right)$.

This follows immediately by noticing that every language-based Receiver strategy that takes different actions after receiving message $m_{\phi(q)}$ and $m \in M_{\phi(q)}$ responds to message $m_{\phi(q)}$ with action $a^R_{\phi(q)}$, which is most preferred by the Sender given that she takes action a^S due to the self-signalling condition, and the two messages are not equivalent in the first round. It implies that for every Sender strategy $\left(m, a^S\right)$ surviving the first iteration where $m \in M_{\phi(q)}$, its constrained message set is contained in $M_{\phi(q)}$.

Communication may fail to achieve coordinated play only if the Receiver makes wrong inference about the Sender's intention from the message he re-

ceived. This may happen only if the message in use is associated with at least two different Sender actions. Proving that every iteratively admissible outcome gives the Sender her Stackelberg payoff involves ruling out situations like this.

Definition 8 *A subset of messages E is intention-clear given a collection of Sender strategies Ξ if*

1. *the constrained message set of every Sender strategy $\left(m, a^S\right)$ in Ξ where $m \in E$ is contained in E, and*

2. *every pair of Sender strategies in this collection Ξ that use messages in E either share the same intention or have disjoint constrained message sets. That is, $\forall \left(m_1, a_1^S\right), \left(m_2, a_2^S\right) \in \Xi$, either $a_1^S = a_2^S$ or*

$$M^{cstr}\left(m_1, b^R\left(a_1^S\right)\right) \cap M^{cstr}\left(m_2, b^R\left(a_2^S\right)\right) = \emptyset.$$

Let E be either a message bundle or a message branch and B be the smallest message bundle strictly containing E. An example is B as the message set consisting of the circle and the message "$\{D\}$" in figure 2 and E as the message set consisting of "$\{B\}$" and messages starting with "$\{AC\}$," or E is the set of messages in the circle. Suppose E is intention-clear given $S^S(k)$. Lemma 2 implies that, for every Receiver strategy s^R in $S^R(k+1)$ that is non-constant on B, $S^R(k+1)$ contains a Receiver strategy $\psi_E\left(s^R\right)$ equal to s^R outside of E but is a best response to every Sender strategy in $S^S(k)$ that uses a message in E. It follows from intention-clearness of E given $S^S(k)$ that if s^R is not a best response to some Sender strategy in $S^S(k)$ that uses a message in E, s^R would be weakly dominated by $\psi_E\left(s^R\right)$, a contradiction to the construction of s^R. This is stated formally in lemma 3.

Lemma 3 *Let E be either a message bundle or a message branch and F be the smallest message bundle strictly containing E. If E is intention-clear given $S^S(k)$, then given any Sender strategy $\left(m, a^S\right)$ in $S^S(k)$ that uses a message in E and any Receiver strategy $s^R \in S^R(k+1)$ that is non-constant on F, we have*

$$s^R(m) = b^R\left(a^S\right).$$

We can deduce that $M_{\phi(N-3)}$ is intention clear given $S^S(3)$ by lemma 3, observation 4.1.2 and the observation that the Sender strategy $\left(m_{\phi(N-l)}, a^S_{\phi(N-l)}\right)$ belongs to $S^S(2)$ for every $l = 0, ..., N-l$ where $a^S_{\phi(N-l)} = b^S\left(a^R_{\phi(N-l)}\right)$.

Lemma 4 is the key step in establishing proposition 3.

Lemma 4 *$\forall k = 1, ..., N-1$, the message set $M_{\phi(N-k-1)}$ is intention-clear given $S^S(4k-1)$.*

Since $M \supset M_{\phi(0)}$, after sufficiently many rounds of deletion, the Sender can convey her intention without fail, and thus every iteratively admissible outcome achieves coordinated play. In particular, this implies that if there exists an admissible Sender strategy that uses a Stackelberg action, then this Sender strategy yields the Stackelberg payoff under any admissible belief. Therefore, if there exists an admissible Sender strategy that uses a Stackelberg action, then every admissible Sender strategy uses a Stackelberg action. Proposition 3 immediately follows because there must exist an admissible Sender strategy that uses a Stackelberg action, since no such Sender strategy can be weakly dominated by a Sender strategy that does not use a Stackelberg action.

We prove lemma 4 by induction. To visualize the inductive process, think of $M_{\phi(N-k-1)}$ as the messages in the circle in figure 2. Assuming that the circle is intention-clearn given $S^S(4k-1)$, we first show that the circle plus the message "$\{D\}$" is intention-clear given $S^S(4k+1)$, and then show that the set of all messages shown in figure 2 ($M_{\phi(N-k-2)}$) is intention clear given $S^S(4k+3)$. We need to show that, given a pair of Sender strategies in $S^S(4k-1)$ which use messages in $M_{\phi(N-k-2)}$ and have overlapping constrained message sets, at least one of them is weakly dominated w.r.t. $S(4k+3)$. To show that $\left(m, a^S\right)$ is weakly dominated at or before the $(4k+3)^{th}$ round of deletion, we need to show the existence of a Sender strategy $\left(\hat{m}, \hat{a}^S\right)$ that may weakly dominate $\left(m, a^S\right)$ w.r.t. $S(4k+3)$. In the case where $m = m_{\phi(N-l)}$ for some $l = 0, ..., N-1$, lemma 5 gives conditions for existence of dominators of $\left(m, a^S\right)$ that use messages in $M_{\phi(N-k-1)}$. In all other cases, the constrained message set of $\left(m, a^S\right)$ belongs to some message bundle E. Lemma 6 gives conditions for existence of dominators of $\left(m, a^S\right)$ that use messages in a message set parallel to E.

To establish existence of dominators of $\left(m, a^S\right)$, we need to show that the desired potential dominator "outlives" $\left(m, a^S\right)$. We do so by minimally modifying an anti-best response of $\left(m, a^S\right)$ and show that its Sender best response gives us the desired properties. Given a message bundle E, let B denote the smallest message bundle strictly containing E. We define Ψ_B to be a mapping from S_L^R to a subset in S_L^R such that $\Psi_B\left(s^R\right) = \left\{s^R\right\}$ for s^R constant on B, while for s^R non-constant on B, $\Psi_B\left(s^R\right)$ is the set of all Receiver strategies in S_L^R that is equal to s^R outside of B. Ψ_B is a collection of minimal modifications.

Let

$$\Psi_B\left(\sigma^R\right) = \left\{ \begin{array}{c} \sigma'^R \in \Delta S_L^R : \\ \forall s^R \in S_L^R,\ \exists \psi_B\left(s^R\right) \in \Psi_B\left(s^R\right) \text{ such that} \\ \sigma'^R\left(\psi_B\left(s^R\right)\right) = \sigma^R\left(s^R\right) \end{array} \right\}.$$

Given any $\sigma^R \in \Delta S_L^R$, $\Psi_B\left(\sigma^R\right)$ is the set of strategies "minimally modified" from σ^R that will give us potential dominators that use messages in B.

Suppose we want a potential dominator of $\left(m, a^S\right)$ to use the same action and send a message in the same message bundle E as m, then the highest payoff to the potential dominator given that the Receiver uses strategies in $\Psi_B\left(\sigma^R\right)$ for some $\sigma^R \in \Delta S_L^R$ is

$$\begin{aligned} \chi\left(\sigma^R, B, a^S\right) \equiv & \sum_{\substack{s^R \text{ constant} \\ \text{on } B}} \sigma^R\left(s^R\right) g^S\left(a^S, s^R(m)\right) \\ & + \sum_{\substack{s^R \text{ non-constant} \\ \text{on } B}} \sigma^R\left(s^R\right) g^S\left(a^S, b^R\left(a^S\right)\right). \end{aligned}$$

This level is obtained if the Sender uses message m in B and the Receiver uses strategy $\psi\left(\sigma^R\right) \in \Psi_B\left(\sigma^R\right)$ such that $\psi\left(s^R\right)(m) = b^R\left(a^S\right)$ for every s^R non-constant on B which receives positive weight from σ^R. In general, the potential dominator uses a message in E and an action $\hat{a}^S$ that maximizes $\chi\left(\sigma^R, B, a'^S\right)$ over a subset of Receiver actions. In the iteration where we aim to show weak dominance of $\left(m, a^S\right)$, typically the payoff to $\left(m, a^S\right)$ given σ^R is strictly lower than $\chi\left(\sigma^R, B, a^S\right)$, and the potential dominator gives a payoff at least as high as $\chi\left(\sigma^R, B, a^S\right)$.

To understand lemma 5, we need to define an action-strict best response.

Definition 9 *$\left(\hat{m}, \hat{a}^S\right)$ is an action strict best response to $\hat{\sigma}^R$ w.r.t. $S^R(k)$ for some iteration k if it gives the Sender a strictly higher payoff against $\hat{\sigma}^R$ than any other Sender strategy that uses some action different from $\hat{a}^S$.*

Lemma 5 *Suppose a Sender strategy $\left(m_{\phi(N-l)}, a^S\right)$ survives the k^{th} round of deletion of weakly dominated strategies, where $b^R\left(a^S\right) = a^R_{\phi(q)}$ and $q \neq N-l$. Write $a^S = a^S_{\phi(q)}$. If*

1. ***(Unclear Intention)*** *$m_{\phi(N-l)} \cup M_{\phi(N-l)}$ is not intention-clear given $S^S(k)$, and*

2. ***(No Sincere Recommendation)*** *no Sender strategy that takes action $a^S_{\phi(q)}$ while using a message in $M_{\phi(N-l)}$ survives the k^{th} round of deletion,*

then for every Receiver strategy $\sigma^R \in \Delta S^R(k-1)$ *to which the Sender strategy*

$$\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right)$$

is a best response, there exists a Sender strategy $\left(\hat{m}, \hat{a}^S\right)$ *in* $S^S(k)$ *and a mixed Receiver strategy* $\psi\left(\sigma^R\right) \in \Delta S^R(k-1)$ *that belongs to*

$$\Psi_{m_{\phi(N-l)} \cup M_{\phi(N-l)}}\left(\sigma^R\right)$$

such that

A $\hat{m} \in M_{\phi(N-l)}$ *and* $\hat{a}^S \neq a^S_{\phi(q)}$*;*

B $\left(\hat{m}, \hat{a}^S\right)$ *is a best response to* $\psi\left(\sigma^R\right)$*, and*

C

$$\begin{aligned} & \chi\left(\sigma^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a^S_{\phi(q)}\right) \\ \leq \quad & \chi\left(\sigma^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, \hat{a}^S\right). \end{aligned}$$

To visually grasp the idea in lemma 6, think of message sets E, F_1 and F_2 as those denoted in figure 2.

Lemma 6 *Let* E *be a message bundle in* $M_{\phi(N-l)}$*, and* F_1 *and* F_2 *be two other parallel message bundles in* $M_{\phi(N-l)}$*. Let* $\left(\hat{m}, \hat{a}^S\right)$ *be a Sender strategy in* $S^S(k)$ *for some iteration* k *where* $\hat{m} \in E$*, and* $\hat{\sigma}^R$ *be a totally mixed strategy in* $S^R(k-1)$ *to which* $\left(\hat{m}, \hat{a}^S\right)$ *is a best response. If there exists* $s^R \in S^R(k)$ *non-constant on* $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ *such that* $s^R(\hat{m}) \neq b^R\left(\hat{a}^S\right)$*, then there exists two Sender strategies* $\left(m_1, a^S_1\right)$*,* $\left(m_2, a^S_2\right)$ *in* $S^S(k)$ *where*

1. *message* m_1 *belongs to* F_1*, message* m_2 *belongs to* F_2*, and*
2. *Sender actions* a^S_1 *and* a^S_2 *both maximize*

$$\chi\left(\hat{\sigma}^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a^S\right)$$

over $\left\{a^S : b^R\left(a^S\right) = a^S_{\phi(i)}, \text{ for some } i > N-l\right\}$.

Proof of lemma 4. It is trivially true for $k = 1$ because $M_{\phi(N-2)} = \left\{m_{\phi(N-1)}, m_{\phi(N)}\right\}$, and if the Sender strategy $\left(m_{\phi(N-1)}, a^S\right)$ belongs to $S^S(1)$, then $a^S = a^S_{\phi(N-1)}$.

Suppose it is true for $k = 1, ..., \bar{k}$.

Claim 1 *$m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$ is intention-clear given $S^S\left(4\bar{k}+1\right)$.*

Proof. Suppose to the contrary that there exists two Sender strategies $\left(m_1, a_1^S\right)$ and $\left(m_2, a_2^S\right)$ in $S^S\left(4\bar{k}+1\right)$ with overlapping constrained message sets where m_1, m_2 belong to $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$ and $b^R\left(a_1^S\right) \neq b^R\left(a_2^S\right)$. We can assume w.l.o.g. that $M^{cstr}\left(m_1, b^R\left(a_1^S\right)\right)$ is larger. By the assumption that $M_{\phi\left(N-\bar{k}-1\right)}$ is intention-clear given $S^S\left(4\bar{k}-1\right)$, it cannot be the case that both m_1 and m_2 belong to $M_{\phi\left(N-\bar{k}-1\right)}$. It follows that the constrained message set of $\left(m_2, a_2^S\right)$ must be $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$. That implies $m_2 = m_{\phi\left(N-\bar{k}-1\right)}$ and $b^R\left(a_2^S\right) = a^R_{\phi(q)}$ where $q \neq N-\bar{k}-1$. Observation 4.1.2 further implies that $q > N-\bar{k}-1$. Either there exists $\left(m_2', a_2^S\right)$ in $S^S\left(4\bar{k}+1\right)$ where $m_2' \in M_{\phi\left(N-\bar{k}-1\right)}$, or the assumptions of lemma 5 holds. It follows that, given any $\sigma_2^R \in \Delta^+ S^R\left(4\bar{k}\right)$ to which $\left(m_2, a_2^S\right)$ is a strict best response w.r.t. $S\left(4\bar{k}\right)$, there exists a Sender strategy $\left(\hat{m}_2, \hat{a}_2^S\right) \in S^S\left(4\bar{k}+1\right)$ where $\hat{m} \in M_{\phi\left(N-\bar{k}-1\right)}$ such that

$$\begin{aligned} & \chi\left(\sigma_2^R, m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}, \hat{a}_2^S\right) \\ \geq \quad & \chi\left(\sigma_2^R, m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}, a_2^S\right). \end{aligned}$$

The assumption that $M_{\phi\left(N-\bar{k}-1\right)}$ is intention-clear given $S^S\left(4\bar{k}-1\right)$ combined with lemma 3 implies that $s^R\left(\hat{m}\right) = b^R\left(\hat{a}^S\right)$ for every $s^R \in S^R\left(4\bar{k}\right)$ non-constant on $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$. It follows that

$$\begin{aligned} & u^S\left(\left(\hat{m}, \hat{a}_2^S\right), \sigma_2^R\right) \\ = \quad & \chi\left(\sigma_2^R, m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}, \hat{a}_2^S\right) \\ \geq \quad & \chi\left(\sigma_2^R, m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}, a^S_{\phi(q)}\right) \qquad (1) \\ \geq \quad & u^R\left(\left(m_{\phi\left(N-\bar{k}-1\right)}, a_2^S\right), \sigma_2^R\right), \qquad (2) \end{aligned}$$

Strict inequality holds in line 2 if $S^R\left(4\bar{k}\right)$ contains a Receiver strategy non-constant on $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$. In that case, we obtain a contradiction to the construction of σ_2^R. If $m_1 = m_{\phi\left(N-\bar{k}-1\right)}$ and $b^R\left(a_1^S\right) = a^R_{\phi\left(N-\bar{k}-1\right)}$, then $\left(m_1, a_1^S\right)$ and $\left(\hat{m}_2, \hat{a}_2^S\right)$ have disjoint constrained message sets, and by lemma 2, $S^R\left(4\bar{k}\right)$ contains a Receiver strategy non-constant on $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$. If $b^R\left(a_1^S\right) \neq a^R_{\phi\left(N-\bar{k}-1\right)}$, then there exists $\left(\hat{m}_1, \hat{a}_1^S\right) \in S^S\left(4\bar{k}+1\right)$ with $\hat{m}_1 \in M_{\phi\left(N-\bar{k}-1\right)}$ that is related to $\left(m_1, a_1^S\right)$ in the same way $\left(\hat{m}_2, \hat{a}_2^S\right)$ is to $\left(m_2, a_2^S\right)$. The assumption of unclear intention implies that either $b^R\left(\hat{a}_1^S\right) \neq b^R\left(\hat{a}_2^S\right)$, in which case we apply lemma 2 and we are done, or $b^R\left(\hat{a}_i^S\right) \neq b^R\left(a_i^S\right)$ for some

$i = 1, 2$. Assume w.l.o.g. that $i = 2$. But then $\hat{a}_2^S \neq a_2^S$, and the construction that $\left(m_2, a_2^S\right)$ is a strict best response to σ_2^R implies that at least one strict inequality holds in line 1 and 2. ■

Claim 2 *$M_{\phi\left(N-\bar{k}-2\right)}$ is intention-clear given $S^S\left(4\bar{k}+3\right)$.*

Proof. Suppose to the contrary that there exists two Sender strategies $\left(m_1, a_1^S\right)$ and $\left(m_2, a_2^S\right)$ in $S^S\left(4\bar{k}+1\right)$ with overlapping constrained message sets where m_1, m_2 belong to $M_{\phi\left(N-\bar{k}-2\right)}$ and $b^R\left(a_1^S\right) \neq b^R\left(a_2^S\right)$. Claim 1 implies that it cannot be the case that m_1 and m_2 both belong to $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$. Observation 4.1.2 and the definition of *constrained message sets* imply that m_1 and m_2 must belong to the same message bundle in $M_{\phi\left(N-\bar{k}-2\right)}$. Denote this message bundle by E. Then from lemma 6, for $i = 1, 2$, given $\sigma_i^R \in \Delta^+ S^R\left(4\bar{k}+2\right)$ to which $\left(m_i, a_i^S\right)$ is a strict best response w.r.t. $S\left(4\bar{k}+2\right)$, there exists $\left(\hat{m}_i, \hat{a}_i^S\right) \in S^S\left(4\bar{k}+3\right)$ where $\hat{m} \in m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$ and

$$\begin{aligned} & \chi\left(\sigma_i^R, m_{\phi\left(N-\bar{k}-2\right)} \cup M_{\phi\left(N-\bar{k}-2\right)}, \hat{a}_i^S\right) \\ \geq \; & \chi\left(\sigma_i^R, m_{\phi\left(N-\bar{k}-2\right)} \cup M_{\phi\left(N-\bar{k}-2\right)}, a_i^S\right). \end{aligned}$$

But from claim 1, $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$ is intention-clear given $S^S\left(4\bar{k}+1\right)$. Lemma 3 then implies that $s^R\left(\hat{m}_i\right) = b^R\left(\hat{a}_i^S\right)$ for every $s^R \in S^R\left(4\bar{k}+2\right)$ non-constant on $M_{\phi\left(N-\bar{k}-2\right)}$. Therefore,

$$\begin{aligned} & u^S\left(\left(\hat{m}_i, \hat{a}_i^S\right), \sigma_i^R\right) \\ = \; & \chi\left(\sigma_i^R, m_{\phi\left(N-\bar{k}-2\right)} \cup M_{\phi\left(N-\bar{k}-2\right)}, \hat{a}_i^S\right) \\ \geq \; & \chi\left(\sigma_i^R, m_{\phi\left(N-\bar{k}-2\right)} \cup M_{\phi\left(N-\bar{k}-2\right)}, a_i^S\right) \qquad (3) \\ \geq \; & u^R\left(\left(m_i, a_i^S\right), \sigma_i^R\right). \qquad (4) \end{aligned}$$

There exists $j \in \{1, 2\}$ such that $b^R\left(\hat{a}_1^S\right) \neq b^R\left(a_j^S\right)$ because $b^R\left(a_1^S\right) \neq b^R\left(a_2^S\right)$. Since $\hat{m}_1$ belongs to a message bundle parallel to E, the constrained message set of $\left(\hat{m}_1, \hat{a}_1^S\right)$ is disjoint from that of $\left(m_j, a_j^S\right)$. Lemma 2 implies that $S^R\left(4\bar{k}+2\right)$ contains Receiver strategies non-constant on $m_{\phi\left(N-\bar{k}-2\right)} \cup M_{\phi\left(N-\bar{k}-2\right)}$. Hence inequality 4 holds strictly for both $i = 1, 2$, contradiction to the construction of σ_i^R. ■ ■

4.2 Games with Positive Spillovers

The self-signalling criterion implies that the Sender's preference over the Receiver's actions differ with her own intention. Our language assumption com-

bined with iterative admissibility connects different messages with different preferences. This then separates one intention from the other and guarantees the Sender her Stackleberg payoff.

It seems natural then that the Sender cannot convey any information about her intention through cheap talk if the Sender's preference over the Receiver's actions is invariant with her own intention.

If the stage game is self-committing, then for every $a^R \in A^R$, $b^R\left(b^S\left(a^R\right)\right) = a^R$. Therefore, $A^R(1) = A^R$. It follows that $A^R(\infty) = A^R$ and $A^S(\infty) = A^S(1)$ that contains $b^S\left(A^R\right)$.

Theorem 1 *If the stage game is self-committing and the Sender's preference over the Receiver's actions is independent of her own action, then for every* $\left(a^S, a^R\right) \in A(\infty)$*, there exists* $\left(m, a^S\right) \in S^S(\infty)$ *and* $s^R \in S^R(\infty)$ *such that* $s^R(m) = a^R$*.*

Proof. Define $M(k)$ to be the projection of $S^S(k)$ onto M. So $M(k)$ is the set of messages that are used in $S^S(k)$.

Claim 3 $s^R(1) = S_L^R$*.*

Proof. Given $s^R \in S_L^R$, let σ^S be a totally mixed Sender strategy in

$$\Delta\left\{\left(m, b^S\left(s^R(m)\right)\right) : m \in M\right\}.$$

For every $m \in M$, $\left(m, b^S\left(s^R(m)\right)\right)$ is the only Sender strategy given positive weight under σ^S. Therefore, s^R is a best response to every pure strategy given strictly positive weight by σ^S, and thus is a strict best response to σ^S. It follows that $s^R \in S^R(1)$. ■

Claim 4 $S^S(1) = M(1) \times A^S(1)$*.*

Proof. Suppose $\left(\hat{m}, \hat{a}^S\right) \in S^S(1)$. Write $\hat{m} = A_1 ... A_n$, and define $A_0 = A^R$. Then for every $j = 1, ..., n$, there exists $\bar{a}_j \in A_j$ and $\underline{a}_j \in A_{j-1} \backslash A_j$ where $g^S\left(\hat{a}^S, \bar{a}_j\right) > g^S\left(\hat{a}^S, \underline{a}_j\right)$. Otherwise, $\left(\hat{m}, \hat{a}^S\right)$ would be weakly dominated by $\left(m', \hat{a}^S\right)$ for any $m' \in M\left(A_1 ... A_{j-1}\left(A_{j-1} \backslash A_j\right)\right)$.

Define a partial relation $>$ on A^R by the preference order of the Sender. That is, $a_2^R > a_1^R$ iff $g^S\left(a^S, a_2^R\right) > g^S\left(a^S, a_1^R\right)$. Define

$$s_1^R(m) = \begin{cases} \bar{a}_1 & m \in M(A_1) \\ \underline{a}_1 & m \in M\left(A_1^c\right) \\ \min A^R & \text{otherwise} \end{cases}$$

and

$$s_j^R(m) = \begin{cases} \bar{a}_j & m \in M(A_1...A_{j-1}A_j) \\ \underline{a}_j & m \in M(A_1...A_{j-1}(A_{j-1} \backslash A_j)) \\ s_{j-1}^R(m) & m \notin M(A_1...A_{j-1}) \\ \min A_{j-1} & otherwise \end{cases}$$

for $j = 2, ..., n$. It follows that, for $j = 1, ..., n$, for every a^S,

$$\begin{aligned} u^S\left(\left(m, a^S\right), s_j^R\right) &= g^S\left(a^S, \bar{a}_j\right) \\ &> g^S\left(a^S, \underline{a}_j\right) \\ &= u^S\left(\left(m, a^S\right), s_j^R\right) \end{aligned}$$

for every $m \in M(A_1...A_{j-1}(A_{j-1} \backslash A_j))$ and

$$\begin{aligned} u^S\left(\left(\hat{m}, a^S\right), s_j^R\right) &= g^S\left(a^S, \bar{a}_j\right) \\ &> g^S\left(a^S, \min A^R\right) \\ &= u^S\left(\left(m, a^S\right), s_j^R\right) \end{aligned}$$

for every $m \notin M(A_1) \cup M(A_1^c)$. Define

$$\hat{\sigma}_\varepsilon^R := \sum_{j=1}^{n-1} \varepsilon^{j-1}(1-\varepsilon) s_j^R + \varepsilon^{n-1} s_n^R.$$

Therefore,

$$u^S\left(\left(\hat{m}, a^S\right), \hat{\sigma}_\varepsilon^R\right) > u^S\left(\left(m, a^S\right), \hat{\sigma}_\varepsilon^R\right)$$

for every $m \neq \hat{m}$, every a^S and every ε sufficiently small. Let a^R denote also the constant Receiver strategy that takes the action a^R upon receiving every message. There exists $\alpha \in \Delta A^R$ to which a^S is a best response in the stage game, since $a^S \in A^S(1)$. Let α also denote the Receiver strategy that puts weights $\alpha\left(a^R\right)$ on the constant strategy a^R. Then for ε sufficiently small,

$$\begin{aligned} & u^S\left(\left(\hat{m}, \tilde{a}^S\right), (1-\varepsilon)\alpha + \varepsilon\hat{\sigma}_\varepsilon^R\right) \\ > \; & u^S\left(m, a^S\right) \end{aligned}$$

for every $\left(m, a'^S\right) \neq \left(\hat{m}, a^S\right)$. We have thus established that $\left(\hat{m}, a^S\right) \in S^S(1)$ for every $a^S \in A^S(1)$. ■

Since $S^R(1) = S_L^R$, it is immediate that $S^S(\infty) = S^S(1)$ and $S^R(\infty) = S_L^R$. We are done. ■

		Receiver's Action	
		Invest	Not
Sender's	Invest	10+x,10+x	-90,x
Action	Not	x,-90	0,0

Table 8: leading example in Baliga Morris (2002)

		Receiver's Action	
		Invest	Not
Sender's	Invest	10+x,10+x	-90,x
Action	Not	x,-90	0,0
	Low Cost		

		Receiver's Action	
		Invest	Not
Sender's	Invest	-10+x,10+x	-110,x
Action	Not	x,-90	0,0
	High Cost		

Table 9: Incomplete Information Investment Game

4.3 Comparison with Baliga and Morris

To formally formulate the role of the self-signalling criterion, Baliga and Morris (2002) transforms the complete information game into a coordination game with incomplete information, and use the solution concept of perfect Bayesian equilibrium. The counterfactual "what would the Sender have said had she intended to play action a' instead of a" does not really have a role in the solution concept of Nash equilibrium in complete information games. However, the solution concept of perfect Bayesian equilibrium addresses the question "what would the Sender have said were she of type t'.?

The easiest way to see the comparison is to look at the leading example in Baliga and Morris (2002). The game is shown in table 8.

In this stage game, both action $Invest$ and action Not are self-committing. If $x < 0$, the stage-game is self-signalling, while the game exhibits positive spillovers if $x > 0$. To formally study the role of self-signalling, Baliga and Morris (2002) study the following incomplete information game where with probability $1 - p$ the Sender is of *Low Cost* and with probability $p > 0$ the Sender is of $High\ Cost$. The *Low Cost* type has the same payoff matrix as in the complete information game of table 8. However, the $High\ Cost$ Sender has a dominant strategy to not invest. The Receiver's payoff depends only on the action taken by the Sender, not on the Sender's type. Therefore, the Receiver

cares about the type of the Sender only insofar as it conveys information about the action the Sender would take. For example, if the Receiver knew that the Sender is of *High Cost*, the Receiver would infer that the Sender would not invest, and thus his best response would be to not invest. Hence, the hypothetical Sender who intends to not invest is equated with the *High Cost* Sender who has a dominant strategy to not invest. Since the prior puts strictly positive weight on the *High Cost* type, the strategy of the *High Cost* type, or equivalently, the strategy of the hypothetical Sender who intends to not invest, has to be taken into account by the Receiver.

They show that when $x < 0$, there exists a perfect Bayesian equilibrium where the *Low Cost* Sender sends a different message from the *High Cost* Sender and both the Sender and the Receiver invest when the Sender is of *Low Cost*, while neither of them invest when the Sender is of *High Cost*. However, when $x > 0$, there can be no perfect Bayesian equilibrium where the outcomes are type-dependent. Conditional on the Sender being *Low Cost* type, when $x < 0$, there exists an equilibrium where the outcome is $(Invest, Invest)$. On the other hand, when $x > 0$, conditional on the Sender being *Low Cost* type, the unique equilibrium outcome is (Not, Not) if the probability of the *High Cost* type (p) is greater than $\frac{1}{10}$, while the equilibrium outcome could be either $(Invest, Invest)$ or (Not, Not) if $p < \frac{1}{10}$. Since the stage-game is self-signalling only when $x < 0$, this illustrates the role of the self-signalling criterion.

When $x < 0$, our approach predicts that the unique outcome is for both to invest, which coincides with the prediction of Baliga and Morris (2002). When $x > 0$, our approach predicts that every action profile is possible. This is natural because there is not a fixed probability attached to the pessimistic Sender who is going to not invest, and players may have incorrect belief about each other.

The formal model in Baliga and Morris (2002) is as follows. The Sender is one of a finite set of possible types T. The Sender's utility function is $\tilde{g}^S : A^S \times A^R \times T \to R$; the Receiver's utility function is $g^R : A^S \times A^R \to R$. For ease of comparison, I rewrite the positive result in Baliga and Morris (2002) here.

Proposition 2 *If (1) for each $a^S \in A^S$, there exists a type $\tau\left(a^S\right) \in T$ such that a^S the dominant strategy for the Sender in the game $g^S(.,t)$; and (2) for each action $a^R \in A^R$, there exists $a^S \in A^S$ such that $a^R = b^R\left(a^S\right)$, then there exists a full revelation perfect Bayesian equilibrium in the one-sided cheap talk game if and only if*

1. *a^S is a self-committing action for the Sender in the game $g^S\left(.,\tau\left(a^S\right)\right)$;*

2. *a^S is the Stackelberg action for the Sender in the game $g^S\left(.,\tau\left(a^S\right)\right)$;*

3. *a^S is self-signalling for the Sender in the game $g^S\left(.,\tau\left(a^S\right)\right)$.*

Let's take the complete information stage game $g=\left(A^S,A^R,g^S,g^R\right)$ where $A^R=\left\{b^R\left(a^S\right):a^S\in A^S\right\}$. Let $T=A^S$. For clarity, let τ be the bijective function from A^S to T. Let $a_*^S=\arg\max_{a^S} g^S\left(a^S,b^R\left(a^S\right)\right)$. Then a_*^S is the Stackleberg action for the Sender in the game G. Define

$$\bar{d}:=\max_{\left(a^S,a^R\right)\neq\left(a'^S,a'^R\right)}\left|g^S\left(a^S,a^R\right)-g^S\left(a'^S,a'^R\right)\right|.$$

$\bar{d}$ is thus the maximum payoff difference for the Sender. Expand the utility function $g^S:A^S\times A^R\to R$ into $\tilde{g}^S:A^S\times A^R\times T\to R$ as follows. $\tilde{g}^S\left(.,\tau\left(a_*^S\right)\right)=g\left(.\right)$, and for every $a^S\neq a_*^S$, $\tilde{g}^S\left(a^S,a'^R,\tau\left(a^S\right)\right)=g^S\left(a^S,a'^R\right)$ for every $a'^R\in A^R$, and $\tilde{g}^S\left(a'^S,a'^R,\tau\left(a^S\right)\right)=g^S\left(a'^S,a'^R\right)-2\bar{d}$, for every $a'^S\neq a^S$ and every $a'^R\in A^R$. Denote the one-sided cheap talk extension game of g with language by G_L, and the one-sided cheap talk extension game of $\tilde{g}$ by $\tilde{G}$. Then proposition 2 implies that $\tilde{G}$ has a full revelation prefect Bayesian equilibrium if and only if the complete information stage game g is self-signalling and every action a^S for the Sender is self-committing. In this equilibrium, the type $\tau\left(a_*^S\right)$, whose payoff matrix is g, gets her Stackelberg payoff. Our positive result equivalently states that if every a^S in g is self-committing and g is self-signalling, the unique iterative admissible outcome in G_L gives the Sender her Stackelberg payoff. The negative result in Baliga and Morris (2002) says that there is no communication in any equilibrium of $\tilde{G}$ if g exhibits binary action positive spillovers. Equilibrium outcomes of $\tilde{G}$ in such games depend on the common prior over T. If there is no common prior, and we allow any prior over T, we can span every rationalizable outcome. Our negative result relaxes the condition to any finite games with positive spillovers, and states that every rationalizable outcome is consistent with iterative admissibility in G_L.

5 Conclusion

By modeling the idea that there exists common knowledge of language, we are able to formally distinguish coordination games with and without the self-signaling condition first used by Aumann (1990) within the framework of complete information games. We show that, if the stage game is self-committing and self-signalling, every iterative admissible outcome in the language game achieves

coordination and gives the Sender her Stackelberg payoff. On the other hand, if the stage game is self-committing but the Sender's preference for the Receiver's actions does not depend on her intended action, every rationalizable stage game outcome is also an iteratively admissible outcome in the language game.

6 Appendix

6.1 Proof for lemma 2

Every Receiver best response to

$$\frac{1}{2}\left(m_1, a_1^S\right) + \frac{1}{2}\left(m_2, a_1^S\right)$$

responds to message m_i with $b^R\left(a_i^S\right)$ for $i = 1, 2$, and thus is non-constant on B since $b^R\left(a_1^S\right) \neq b^R\left(a_2^S\right)$. At least of of the best responses survive the $(j+1)^{th}$ iteration and statement 1 follows.

Given $s^R \in S^R(j)$ non-constant on B, there exists $\sigma^S \in \Delta^+ S^S(j-1)$ to which s^R is a best response. Then every best response to

$$\frac{1}{2}(1-\varepsilon)\sum_{i=1,2}\left(m_i, a_i^S\right) + \varepsilon\sigma^S|_{M\setminus E}$$

gives us the desired property in statement 2.

6.2 Proof for lemma 6

We need the following lemma for the proof.

Lemma 7 *Let F_1, F_2, and F_3 be three parallel message bundles, and B be the smallest message bundle that strictly contains F_1. If*

1. $\left(m_1, a_1^S\right) \in S^S(j-1)$ *where*

$$M^{cstr}\left(m_1, b^R\left(a_1^S\right)\right) \subset F_1,$$

2. $S^S(j-1)$ *contains Sender strategies that use messages in F_2 and F_3 respectively, and*

3. $S^R(j)$ *contains Receiver strategies s_2^R and s_3^R non-constant on B such that either $s_2^R|_{F_2}$ or $s_3^R|_{F_3}$ is not equivalent w.r.t. $S^S(j-1)$ for the Receiver to a constant of $b^R\left(a_1^S\right)$,*

then there exists a mapping $\psi : S^R(j) \to S^R(j)$ *that belongs to* Ψ_B *such that for every* $s^R \in S^R(j)$ *non-constant on* B, $\psi\left(s^R\right)$ *is equal to* s_i^R *on* F_i *for* $i = 2, 3$, *is equal to* s^R *outside of* $E \cup F_1 \cup F_2$, *and is a best response to the Sender strategy* $\left(m_1, a_1^S\right)$.

Proof. Let σ_i^S be a totally mixed Sender strategy in $S^S(j-1)$ to which s_i^R is a best response, for $i = 2, 3$. Let $\sigma^S|_{\tilde{F}}$ denote the mixed Sender strategy that is the probability distribution of σ^S conditional on sending messages in $\tilde{F}$. Given $s^R \in S^R(j)$ non-constant on B, let σ^S be a totally mixed Sender strategy in $S^S(j-1)$ to which s^R is a best response. Then for ε sufficiently small, every Receiver best response to

$$(1-\varepsilon)\left(m_1, a_1^S\right) + \frac{\varepsilon(1-\varepsilon)}{2}\sigma_2^S|_{F_2} + \frac{\varepsilon(1-\varepsilon)}{2}\sigma_3^S|_{F_3} + \varepsilon^2\sigma^S|_{M\setminus(F_2\cup F_3)} \qquad (5)$$

responds to message m_1 with action $b^R\left(a_1^S\right)$, is equivalent to s_i^R on F_i w.r.t. $S^R(j-1)$ for the Receiver, for $i = 2, 3$, and is equal to s^R outside of $F_1 \cup F_2 \cup F_3$. Condition 3 implies that such a best response must also be non-constant on B. Since the Sender strategy in expression 5 is totally mixed on $S^S(j-1)$, at least one best response to expression 5 belongs to $S^R(j)$. Define $\psi\left(s^R\right)$ to be a best response to expression 5 in $S^R(j)$ and we are done. ■

If $l \leq 2$, then $M_{\phi(N-l)}$ consists of only one parallel message bundle, and lemma 6 holds automatically. Hence we are concerned only with $l \geq 3$. It is easy to see that the statement holds for every $l = 3, .., N$ for $k = 1$. Suppose the statement is true for $l = 3, .., N$ and $k = 1, ..., \bar{k}$. Suppose $S^S\left(\bar{k}+1\right)$ contains a Sender strategy $\left(\hat{m}, \hat{a}^S\right)$ where $\hat{m}$ belongs to a message bundle E in $M_{\phi(N-l)}$ and there exists $s'^R \in S^R\left(\bar{k}\right)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ such that $s^R_{w,(\hat{m},\hat{a}^S)}(\hat{m}) \neq b^R\left(\hat{a}^S\right)$. Let $\hat{\sigma}^R$ be a totally mixed strategy in $\Delta S^R\left(\bar{k}\right)$ to which $\left(\hat{m}, \hat{a}^S\right)$ is a best response. Let $A^{\max}$ denote the set of Sender actions that maximize

$$\chi\left(\hat{\sigma}^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a^S\right)$$

over $\left\{a^S_{\phi(i)} : i > N-l\right\}$. Given any two message bundles F_1 and F_2 parallel to E, by assumption, there exist two Sender strategies $\left(m_1, a_1^S\right)$, $\left(m_2, a_2^S\right)$ in $S^S\left(\bar{k}\right)$ where message m_1 belongs to F_1, message m_2 belongs to F_2, and Sender actions a_1^S and a_2^S both belong to $A^{\max}$. It suffices to show that $S^S\left(\bar{k}+1\right)$ contains a Sender strategy $\left(m_{1*}, a_{1*}^S\right)$ where $m_{1*} \in F_1$ and $a_{1*}^S \in A^{\max}$. We would be done if $\left(m_1, a_1^S\right) \in S^S\left(\bar{k}+1\right)$. Suppose $\left(m_1, a_1^S\right) \notin S^S\left(\bar{k}+1\right)$.

Claim 5 *Given any* $\left(\tilde{m}, \tilde{a}^S\right) \in S^S\left(\bar{k}\right)$ *where* $\tilde{a}^S \in A^{\max}$ *and* $\tilde{m} \in E \cup F_2$, *there exists* $\psi_{d,(\tilde{m},\tilde{a}^S)} : S^R\left(\bar{k}\right) \to S^R\left(\bar{k}\right)$ *that belongs to* $\Psi_{m_{\phi(N-l)} \cup M_{\phi(N-l)}}$

such that for every s^R *non-constant on* $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, $\psi_{d,(\tilde{m},\tilde{a}^S)}\left(s^R\right)$ *is equal to* s^R *outside of* $E \cup F_1 \cup F_2$, *while* $\psi_{d,(\tilde{m},\tilde{a}^S)}\left(s^R\right)(m_1) = b^R\left(a_1^S\right)$ *and* $\psi_{d,(\tilde{m},\tilde{a}^S)}\left(s^R\right)(\tilde{m}) \neq b^R\left(\tilde{a}^S\right)$.

Proof. Suppose that $s^R(\tilde{m}) = b^R\left(\tilde{a}^S\right)$ for every Receiver strategy in $S^R\left(\bar{k}\right)$ which is non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, then

$$\begin{aligned} & u^S\left(\left(\tilde{m},\tilde{a}^S\right),\hat{\sigma}^R\right) \\ = \;& \chi\left(\hat{\sigma}^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a_1^S\right) \\ \geq \;& \chi\left(\hat{\sigma}^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, \hat{a}^S\right) \\ > \;& u^S\left(\left(\hat{m},\hat{a}^S\right),\hat{\sigma}^R\right), \end{aligned}$$

which contradicts the assumption that $\left(\hat{m},\hat{a}^S\right)$ is a best response to $\hat{\sigma}^R$. Therefore, there exists $s^R_{w,(\tilde{m},\tilde{a}^S)} \in S^R\left(\bar{k}\right)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ where $s^R_{w,(\tilde{m},\tilde{a}^S)}(m_j) \neq b^R\left(\tilde{a}^S\right)$. Analogously, there exists $s^R_{w,i} \in S^R\left(\bar{k}\right)$ with the same properties w.r.t. $\left(m_i, a_i^S\right)$.

Assume w.l.o.g. that $\tilde{m} \in E$. If either $s^R_{w,(\tilde{m},\tilde{a}^S)}|_E$ or $s^R_{w,\left(m_2,a_2^S\right)}|_{F_2}$ is not equivalent to a constant Receiver strategy of $b^R\left(a_1^S\right)$ w.r.t. $S^S\left(\bar{k}-1\right)$, then we can apply lemma 7 and we are done. Otherwise, $b^R\left(a_2^S\right) \neq b^R\left(a_1^S\right)$. Since F_1 and F_2 are parallel in $M_{\phi(N-l)}$, constrained message sets of $\left(m_i, a_i^S\right)$ are disjoint, and lemma 2 implies that there exists $s_2^R \in S^R\left(\bar{k}\right)$ which is a best response to $\left(m_i, a_i^S\right)$, for $i = 1, 2$. It follows that s_2^R is non-constant on $M_{\phi(N-l)}$ and $s_2^R|_{F_2}$ is not equivalent to a constant of $b^R\left(a_1^S\right)$. We can again apply lemma 7 and we are done. ■

Following claim 5, we can then define $T_d : S^R\left(\bar{k}\right) \to S^R\left(\bar{k}\right)$ where $T_d\left(s^R\right)$ puts strictly positive probability on every strategy in the set

$$\left\{ \begin{array}{c} \psi_{d,(\tilde{m},\tilde{a}^S)} : \left(\tilde{m},\tilde{a}^S\right) \in S^S\left(\bar{k}+1\right) \\ \tilde{m} \in E \cup F_2 \text{ and } \tilde{a}^S \in A^{\max} \end{array} \right\}.$$

Not that for every s^R non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, $T_d\left(s^R\right)(m_1) = b^R\left(a_1^S\right)$, $T_d\left(s^R\right)$ is equal to s^R outside if $E \cup F_1 \cup F_2$, and

$$T_d\left(s^R\right)(\tilde{m}) \neq b^R\left(\tilde{a}^S\right) \text{ if } \left(\tilde{m},\tilde{a}^S\right) \in S^S\left(\bar{k}\right) \text{ and } \tilde{m} \in E \cup F_2. \tag{6}$$

Therefore, for every Sender strategy $\left(m, a^S\right)$ where $m \notin E \cup F_1 \cup F_2$,

$$\begin{aligned} u^S\left(\left(m, a^S\right), T_d\left(\hat{\sigma}^R\right)\right) &= u^S\left(\left(m, a^S\right), \hat{\sigma}^R\right) \\ &\leq u^S\left(\left(\hat{m}, \hat{a}^S\right), \hat{\sigma}^R\right) \\ &< \chi\left(\hat{\sigma}^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, \hat{a}^S\right) \\ &\leq \chi\left(\hat{\sigma}^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a_1^S\right) \\ &= u^S\left(\left(m_1, a_1^S\right), T_d\left(\hat{\sigma}^R\right)\right). \end{aligned}$$

If $m \in E \cup F_1 \cup F_2$, then

$$\begin{aligned} & u^S\left(\left(m, a^S\right), T_d\left(\hat{\sigma}^R\right)\right) \\ \leq\ & \chi\left(\hat{\sigma}^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a^S\right) \quad (7) \\ \leq\ & \chi\left(\hat{\sigma}^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a_1^S\right) \quad (8) \\ =\ & u^S\left(\left(m_1, a_1^S\right), T_d\left(\hat{\sigma}^R\right)\right). \end{aligned}$$

If $a^S \notin A^{\max}$, then strictly inequality holds in line 8. If $a^S \in A^{\max}$, $m \in E \cup F_2$, and $\left(m, a^S\right) \in S^S\left(\bar{k}\right)$, then strict inequality holds in line 7 because of inequality 6 and the construction of $\hat{\sigma}^R$ and T_d such that $T_d\left(\hat{\sigma}^R\right)$ puts positive probability on Receiver strategies non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. If $m \in F_1$ but $a^S \notin A^{\max}$, then it follows that

$$u^S\left(\left(m, a^S\right), T_d\left(\hat{\sigma}^R\right)\right) < u^S\left(\left(m_1, a_1^S\right), T_d\left(\hat{\sigma}^R\right)\right).$$

Therefore, if $\left(m_*, a_*^S\right)$ is a best response in $S^S\left(\bar{k}\right)$ to $T_d\left(\hat{\sigma}^R\right) \in \Delta S^R\left(\bar{k}\right)$, it has to be the case that message $m_* \in F_1$ and Sender action $a_*^S \in A^{\max}$. Since at least one Sender best response to $T_d\left(\hat{\sigma}^R\right)$ survives the $\left(\bar{k}+1\right)^{th}$ iteration, we are done.

6.3 Proof for lemma 5

Both statements are true for $k = 1$, for all $N - l \neq q$. Since $M_{\phi(N-3)}$ is intention clear given $S^S(3)$, we are concerned only if $l \geq 3$. Suppose both statements are true for every $N - l \neq q$, for $k = 1, ..., \bar{k}$. Let $\left(m_{\phi(N-l)}, a_{\phi(q)}^S\right)$ be a Sender strategy that survives the $\left(\bar{k}+1\right)^{th}$ round of deletion, where $b^R\left(a_{\phi(q)}^S\right) = a_{\phi(q)}^R$. Observation 4.1.2 implies that $q \geq N - l$. Suppose $\left(m_{\phi(N-l)}, a_{\phi(q)}^S\right)$ satisfies both the *Unclear Intention* condition and the *No Sincere Recommendation* condition, . This implies that $q \neq N - l$.

Let σ_q^R be a Receiver strategy in $\Delta S^R\left(\bar{k}\right)$ to which $\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right)$ is a best response w.r.t. $S\left(\bar{k}\right)$. By assumption, there exists $\left(\hat{m}, \hat{a}^S\right) \in S^S\left(\bar{k}\right)$ where $\hat{m} \in M_{\phi(N-l)}$ and

$$\begin{aligned} & \chi\left(\sigma_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, \hat{a}^S\right) \\ \geq & \chi\left(\sigma_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a^S_{\phi(q)}\right), \end{aligned}$$

and there exists $\hat{\sigma}_q^R \in \Psi_{m_{\phi(N-l)} \cup M_{\phi(N-l)}}\left(\sigma_q^R\right)$ to which $\left(\hat{m}, \hat{a}^S\right)$ is a best response where $s^R\left(\hat{m}\right) = b^R\left(\hat{a}^S\right)$ for every s^R non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ that receives positive weights under $\hat{\sigma}_q^R$. If $S^R\left(\bar{k}\right)$ contains only Receiver strategies constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, then for σ_q^R totally mixed in $\Delta S^R\left(\bar{k}\right)$,

$$\begin{aligned} & U^S\left(\left(\hat{m}, \hat{a}^S\right), \sigma_q^R\right) \\ = & \chi\left(\sigma_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, \hat{a}^S\right) \\ \geq & \chi\left(\sigma_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a^S_{\phi(q)}\right) \\ = & U^S\left(\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right), \sigma_q^R\right). \end{aligned}$$

Thus $\left(\hat{m}, \hat{a}^S\right)$ belongs to $S^S\left(\bar{k}+1\right)$ and we are done.

Suppose $S^R\left(\bar{k}\right)$ contains Receiver strategies non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. Then there exists $s_w^R \in S^R\left(\bar{k}\right)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ such that $s_w^R\left(\hat{m}\right) \neq b^R\left(\hat{a}^S\right)$. Otherwise, for σ_q^R totally mixed,

$$\begin{aligned} & U^S\left(\left(\hat{m}, \hat{a}^S\right), \sigma_q^R\right) \\ = & \chi\left(\sigma_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, \hat{a}^S\right) \\ \geq & \chi\left(\sigma_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a^S_{\phi(q)}\right) \\ > & U^S\left(\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right), \sigma_q^R\right), \end{aligned}$$

because $s^R\left(m_{\phi(N-l)}\right) \neq b^R\left(a^S_{\phi(q)}\right)$ for every s^R non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. This contradicts the construction of σ_q^R.

Let E be the message bundle which $\hat{m}$ belongs to. By lemma 6, there exists $\left(m_1, a_1^S\right)$, $\left(m_2, a_2^S\right) \in S^S\left(\bar{k}\right)$ where each m_i belongs to a different message bundle F_i parallel to E,and a_i^S belongs to

$$\arg \max_{a'^S \in (b^R)^{-1}\left(a^R_{\phi(j)}\right) \text{ for } j > N-l} \chi\left(\hat{\sigma}_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a'^S\right),$$

for $i = 1, 2$. Therefore, for $i = 1, 2$,

$$\begin{aligned}
& \chi\left(\hat{\sigma}_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a_i^S\right) \\
\geq\ & \chi\left(\hat{\sigma}_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, \hat{a}^S\right) \\
=\ & \chi\left(\sigma_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, \hat{a}^S\right) \\
\geq\ & \chi\left(\sigma_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a_{\phi(q)}^S\right).
\end{aligned}$$

Similarly, there must exist $s_{wi}^R \in S^R\left(\bar{k}\right)$ such that $s_{wi}^R\left(m_i\right) \neq b^R\left(a_i^S\right)$. If $b^R\left(a_i^S\right) \neq b^R\left(\hat{a}^S\right)$ for some $i \in \{1, 2\}$, then we use lemma 2, otherwise, we use lemma 7 to establish existence of a mapping $\psi : S^R\left(k\right) \to S^R\left(k\right)$ that belongs to $\Psi_{m_{\phi(N-l)} \cup M_{\phi(N-l)}}$ such that $\psi\left(s^R\right)\left(\hat{m}\right) = b^R\left(\hat{a}^S\right)$, for every $s^R \in S^R\left(k\right)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. It follows that

$$\begin{aligned}
& u^S\left(\left(\hat{m}, \hat{a}^S\right), \psi\left(\sigma_q^R\right)\right) \\
=\ & \chi\left(\sigma_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, \hat{a}^S\right) \\
\geq\ & \chi\left(\sigma_q^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a_{\phi(q)}^S\right) \\
>\ & u^S\left(\left(m_{\phi(N-l)}, a_{\phi(q)}^S\right), \psi\left(\hat{\sigma}^R\right)\right) \\
=\ & u^S\left(\left(m_{\phi(N-l)}, a_{\phi(q)}^S\right), \hat{\sigma}^R\right) \qquad (9) \\
\geq\ & u^S\left(\left(m, a^S\right), \hat{\sigma}^R\right) \\
=\ & u^S\left(\left(m, a^S\right), \psi_d\left(\hat{\sigma}^R\right)\right)
\end{aligned}$$

for every $\left(m, a^S\right)$ where $m \notin M_{\phi(N-l)}$. Equality holds in line 9 because $s^R\left(m_{\phi(N-l)}\right) = a_{\phi(N-l)}^R$ for every s^R non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. Therefore, any best response to $\psi\left(\sigma_q^R\right)$ uses a message in $M_{\phi(N-l)}$. Since at least one of the Sender best responses to $\psi\left(\sigma_q^R\right)$ survives the $\left(\bar{k}+1\right)^{th}$ iteration, we are done.

References

[1] R. Aumann, "Nash Equilibria are not Self-Enforcing," *Economics Decision-Making: Games, Econometrics and Optimization (J.J. Gabszewicz, J.-F. Richard, and L. A. Wolsey, Eds.), 1990.*

[2] Brandenburger, A., Friedenberg, A., and H.J. Keisler., "Admissibility in Games," mimeo, 2004. Available at www.stern.nyu.edu/~abranden.

[3] S. Baliga and S. Morris, "Coordination, Spillover, and Cheap Talk," Journal of Economic Theory, 2002, 450-468.

[4] J. Farrell, "Communication, Coordination and Nash Equilibrium," Economic Letter, 1988, 209-214.