Language and Coordination Games

Melody Lo[‡]
Hong Kong Baptist University

July, 2018

Abstract

I formalize the role of the self-signaling condition in guaranteeing coordination for complete information coordination games. I model a pre-existing common language by assuming that the Receiver either ignores or follows cheap talk recommendations, but never inverts them. This assumption creates asymmetry between messages, which captures the essence of a common language. It does not rule out any outcome at hand in that every equilibrium outcome of the original game remains an equilibrium outcome in this transformed game. However, applying iterative admissibility to this transformed game yields sharp predictions. If the stage game satisfies a certain self-signaling condition, then the Sender gets her Stackelberg payoff in every iteratively admissible outcome. On the other hand, if the stage game violates a weaker self-signalling condition, miscoordination can happen in an iteratively admissible outcome.

*I am grateful to Stephen Morris, Dino Gerardi and Ben Polak for their guidance and advice. I have benefitted greatly from the discussion with Roberto Serrano, Kim-Sau Chung and Jimmy Chan. I thank the participants of the Summer Student Workshop at Yale University, Theory Lunch at Brown University, Canadian Economic Theory Workshop and Workshop at University of Hong Kong. This paper benefitted from the basic research grant at Shanghai University of Finance and Economics.

†Address: Department of Economics, Hong Kong Baptist University, Hong Kong. Email: peiyulo2006@gmail.com.

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

Table 1: Investment game

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

Table 2: Battle-of-the-sexes game

1 Introduction

When can communication improve coordination? In a standard cheap talk model, communication cannot guarantee play of a Pareto efficient equilibrium. This happens because meaning is only endogenously defined in equilibrium, and thus babbling equilibria always exist.

Intuitively many messages have meanings beyond the game in question. Some papers link messages to actions by assuming that a credible message should be believed with various definitions of credibility. Farrell (1988) argues that a suggestion to play a Nash equilibrium is credible because, if the Sender believes that the Receiver will follow the suggestion, then the Sender has an incentive to do so as well. Aumann (1990) argues that a suggestion to play a Nash equilibrium must be *self-signaling* to be credible; that is, the Sender prefers the Receiver to follow the suggestion only if the Sender intends to do so as well.

Compare the investment game in Table 1 and the battle-of-the-sexes game in Table 2. There are two pure strategy Nash equilibria in both games. In the investment game, the suggestion to play the $(Invest, Invest)$ equilibrium is not self-signaling because the Sender prefers the Receiver to $Invest$ even if she plans to choose $Not\ Invest$. In the battle-of-the-sexes game, the suggestion

	"You should invest!"	"You should NOT invest!"
$Constant - Invest$	$Invest$	$Invest$
$Constant - Not\ Invest$	$Not\ Invest$	$Not\ Invest$
$Literal$	$Invest$	$Not\ Invest$
$Perverse$	$Not\ Invest$	$Invest$

Table 3: Receiver's strategiest in the standard one-sided cheap-talk extension of the investment game

	"You should invest!"	"You should NOT invest!"
$Constant - Invest$	$Invest$	$Invest$
$Constant - Not\ Invest$	Not	Not
$Literal$	$Invest$	Not

Table 4: Receiver's strategies in the language-talk extension of the investment game

to play the $(Opera, Opera)$ equilibrium is self-signaling because the Sender prefers the Receiver to go to the $Club$ if she plans to go to the $Club$ herself. For Farrell, in both games, the Receiver should follow any suggestion to play an equilibrium. Thus the Sender will suggest her favorite equilibrium to obtain her Stackelberg payoff. For Aumann, only in the battle-of-the-xexes game can communication guarantee coordination.

Unlike Farrell and Aumann, I place no restrictions on the Receiver's response to any message. Instead I place restrictions on how the Receiver's response to one message relates to that to another. Table 3 and 4 list all pure Receiver strategies in the standard cheap talk extension game and in my language talk game, respectively. I incorporate literal meaning by ruling out the $Perverse$ strategy. Note that $Literal$ strategy and $Perverse$ strategy each respond to one message with $Invest$ and the other with $Not\ Invest$, albeit in reverse labeling. I assume that the only strategic uncertainty lies in the set of actions the Sender may persuade the Receiver to take, but not in how messages are used to do so. Messages are symmetric in the standard cheap talk game (reversing the labeling does not change the game), but asymmetric in the language talk game. I believe this asymmetry captures

	"*Opera*"	"*Club*"
$Constant - Opera$	*Opera*	*Opera*
$Constant - Club$	*Club*	*Club*
Literal	*Opera*	*Club*

Table 5: Receiver's strategies in the language talk extension of the battle-of-the-sexes game

the essence of a common language.

Using the language talk game, I formalize the idea that self-signaling is necessary and sufficient to guarantee coordination by applying iterative admissibility, which deletes all weakly dominated strategies in each iteration.

In the investment game example, if the Sender believes with high probability that the Receiver will choose the $Constant - Not\ Invest$ strategy, then the Sender will choose $Not\ Invest$. In the language talk game, saying "You should NOT invest!" and choosing any action is weakly dominated by saying "You should invest!" and choosing that same action. Because the Sender has an incentive to say "You should invest!" regardless of her planned action, cheap talk cannot guarantee coordination. Self-signaling is necessary.

To illustrate sufficiency, consider the battle-of-the-sexes game. Table 5 lists all pure Receiver strategies in the corresponding language talk game. In this game, saying "You should go to the Opera!" but choosing *Club* is weakly dominated for the Sender by saying "You should go to the Club!" and choosing the same action *Club*. Therefore, if the Sender says "You should go to the Opera!," the Receiver can conclude that the Sender must intend to go to the *Opera*, and thus best respond with *Opera*. Thus the Sender can guarantee her Stackelberg payoff by saying "You should go to the *Opera*!"

I generalize the above insight to all finite complete information games where every Receiver action and the Sender's best response to it forms a Nash equilibrium. I call such games coordination games. To generalize the language model, I allow for iteratively more refined suggestions and assume that, at each level of refinement, the Receiver may ignore or follow the Sender's

			Receiver	
		X	Y	Z
	x	$12, 12$	$1, 0$	$2, 6$
Sender	y	$6, 0$	$10, 10$	$4, 7$
	z	$4, 0$	$6, 1$	$8, 8$

Table 6: A coordination game that is BM self-signaling

suggestion, but never inverts it.

The "no inversion" assumption is already implicitly made in the literature that incorporates literal meanings into cheap talk games. To link words with actions, this literature often assumes that, for every Receiver best response X to some conjecture, a message "X" exists that literally recommends action X. In assessing the credibility of message "X," one has to ask when the Sender wants the Receiver to take action X. The implicit assumption is that the Receiver responds to message "X" with action X and to another message with a different action. I formally model these alternative messages to literally mean "not X."

Whether the Sender wants the Receiver to take action X crucially depends on which alternative actions the Sender can induce. Consider the stage game given by Table 6. If the Sender intends to take action Y, she will try to induce action Y. But if action Z is the only alternative action to X, she will want the Receiver to take action X. Credibility of message "X" thus depends on whether the Sender believes she can surely induce action Y as an alternative to X. This is equivalent to asking whether there exists an alternative message to "X" which literally recommends Y and is credible conditional on the Receiver taking either action Y or Z. The iterative structure of my language model captures this kind of reasoning.

There are many ways to generalize Aumann's definition of self-signaling beyond the 2x2 games that he discusses. Following Baliga and Morris (2002), a coordination game is *BM self-signaling* if, for any action the Sender plans to take, she prefers the Receiver's best response to her planned action over any other Receiver action. I define a coordination game to be *weak self-signaling*

if, for any action the Sender plans to take, she prefers the Receiver's best response to her planned action over SOME Receiver action which gives the Sender a higher payoff in the associated stage game Nash equilibrium.

I use iterative admissibility as the solution concept. I show that the BM self-signaling condition is sufficient but not necessary to guarantee coordination on the Sender's most preferred stage game equilibrium, while the weak self-signaling condition is necessary but not sufficient to do so. If the Sender's preference over the Receiver actions is invariant with the Sender's own action, then every stage game outcome is itertatively admissble.

The combination of my language assumption and iterative admissibility is key to these results. Without the language assumption, every action profile is an iteratively admissible outcome. On the other hand, every Nash equilibrium outcome in the original cheap talk extension game remains a Nash equilibrium outcome in the language talk game.

By assuming that the Receiver does not invert messages, I assume away double bluff strategies.[1] As such, this paper can be seen as a step in the direction suggested by Rabin (1990): "S[ender] might reasonably doubt his ability to systematically fool R[eceiver]. Incorporating this more cautious thinking by S[ender] into a solution concept might be possible."

1.1 Related Work

This paper builds on the literature that incorporates literal meanings into cheap talk games. Among these studies, Aumann (90) and Farrell (88) consider cheap talk games about planned actions, while Farrell (1993) and Rabin (1990) study cheap talk games about private information. These papers first apply some rationality analysis to determine the credibility of messages, and then apply a solution concept assuming that the Receiver follows credible messages. In contrast, I apply iterative admissibility to reason about the

[1] Bluff is always ignored by the Receiver. If the Receiver believes that action X is not optimal when the Sender recommends it, then the Receiver will take the same action whether the Sender says "X" or "Not X".

credibility of messages and about the opponents' strategies at the same time. In a recent paper, Sobel (2017) also applies iterative admissibility to pre-play cheap talk games where players are restricted to playing "monotonic" strategies.

In another related work, Baliga and Morris (2002) formalize Aumann's argument by turning the complete information game into one with incomplete information. Even though the Sender will invest if his payoff is given by Table 1 and if he can convince the Receiver to invest, the Receiver worries that the Sender may actually have a higher investment cost such that $Not\ Invest$ is a dominant strategy. Thus statements about the Sender's investment costs (her private information) can be informative about the Sender's planned behavior. They show that the BM self-signaling condition ensures existence of a fully revealing equilibrium. But they cannot rule out babbling equilibria where communication is completely ineffective. In contrast, in this paper I keep the complete information structure of the stage game. I show that the BM self-signaling condition rules out babbling equilibria. Moreover, it ensures coordination in every solution.

Ellingsen and Östling (2010) study cheap talk about intentions using the level-k model with lexicographic preference for truthfulness. Self-signaling is irrelevant in their model because they assume that the unsophisticated Sender randomizes between all actions and tells the truth. Thus a level-1 Receiver best responds to the unsophisticated Sender and follows the Sender's recommendation. It follows that all players that are even more sophisticated coordinate on the Pareto efficient equilibrium.

A related line of research uses evolutionary dynamics or *curb* (closed under rational behavior) concept to capture a common language.[2] Self-signaling does not matter in these papers, which generally select the Pareto efficient equilibrium uniquely.[3] Consider the investment game. By evolutionary drift,

[2] See Blume (1998), Kim and Sobel (1995), Hurkens (1996) and Demichelis and Weibull (2008), just to name a few.

[3] An exception is Heller (2014), which modifies Demichelis and Weibull (2008) and show that an inefficient equilibrium can be evolutionarily stable if preferences incorporating

some messages will eventually develop the meaning for $Invest$ because it is optimal to respond to an unused message with any action. Then the Sender will send such a message and choose $Invest$ as well. Coordination on the Pareto dominant equilibrium $(Invest, Invest)$ is thus guaranteed in the long run.

In contrast, for coordination to be guaranteed in my one-shot setting, it is necessary that the message "$Invest$" induces the action $Invest$ in every (not just some) Receiver best response. For this to be the case, the Sender must use the message "$Invest$" only when she intends to choose $Invest$. This is where self-signaling comes in. In some sense, the difference between the two approaches is that curb is the smallest set of strategy profiles with the best response property while rationalizability is the largest set of strategy profiles with such property.

The remainder of this paper is structured as follows. Section 2 describes the language-talk game. Section 3 describes the solution concept. Section 4 presents the main results. Section 5 discusses implications of the iterative structure of the language model, while section 6 concludes.

2 The Language-Talk Game

Let g denote a finite complete information game between the Sender (S, she) and the Receiver (R, he). Each player i chooses simultaneously an action $a^i \in A^i$ and obtains a payoff $g^i\left(a^S, a^R\right)$. For simplicity, I assume that $g^i\left(a^S, .\right)$ and $g^i\left(., a^R\right)$ are injective, for all $a^S \in A^S, a^R \in A^R$, and $i, j \in \{S, R\}$. Denote the best response function in g by $b^i : A^j \to A^i$ for $i \neq j \in \{S, R\}$. I restrict attention to coordination games, where $b^i\left(b^{-i}\left(a^i\right)\right) = a^i$ for every action a^i where $i = S, R$.

Let G denote the standard one-sided cheap talk extension game where the Sender sends a message $m \in M$ to the receiver before they play the complete information game g. In the standard cheap talk game G, the Sender's pure

small lying costs are continuous.

strategy set consists of all pairs $\left(m, a^S\right) \in M \times A^S$, while the Receiver's pure strategy set consists of all functions $s^R : M \to A^R$. Player i's payoff from a pure strategy profile in the cheap talk extension game is $u^i\left(\left(m, a^S\right), s^R\right) = g^i\left(a^S, s^R(m)\right)$, for $i = S, R$.

I take messages to be recommendations; I indicate messages with quotation marks in the following discussion. By the coordination nature of the stage game, the claim about intended action $a^S \in A^S$ or the suggestion to play equilibrium $\left(a^S, b^R\left(a^S\right)\right)$ are both equivalent to the recommendation $b^R\left(a^S\right)$.[4] I assume that the common language contains an expression for every subset of the Receiver action set A^R and an expression for concatenation. Therefore, for every decreasing sequence $A_0, A_1, A_2, ..., A_n$ where $A_0 =: A^R \supset A_1 \supset ... \supset A_n$, the message space M contains a "hierarchical recommendation" denoted by "$A_0A_1A_2...A_n$". Consider for example $A^R = \{Stay\ Home, Opera, Club\}$. The message "Don't stay home tonight. If you want a more specific recommendation, I'd say go to the opera," is a hierarchical recommendation denoted by

$$\text{“}\{Opera, Club, Home\}\{Opera, Club\}\{Opera\}\text{”}.^5$$

Table 7 lists in the second row all hierarchical recommendations when $A^R = \{X, Y, Z\}$.

Messages that begin with the same sequence $A_0A_1...A_j$ share the first j levels of recommendations. Denote the set of all such messages by $M(A_0...A_j)$. If A_j is not a singleton, then $M(A_0...A_j)$ contains messages of various further refinements. Among them, $M(A_0...A_jA_{j+1})$ is related to $M(A_0...A_j(A_j \backslash A_{j+1}))$ by negation given that they express opposing refinements of "A_{j+1}" vs. "*Not* A_{j+1}". The union of these two sets is called a *component message block of*

[4] Ultimately, the Sender talks in order to influence the Receiver's action. Rabin (1990) holds a similar view.

[5] The message "Go to the opera tonight! At least go out!" is treated as the same message. Essentially, the hierarchical recommendation "$A_0A_1A_2...A_n$" is a partial order on Receiver actions that says some Receiver action in A_n is preferred to some in $A_{n-1} \backslash A_n$, and some action in A_{n-1} is preferred to some in $A_{n-2} \backslash A_{n-1}$, etc.

message block	E_X			E_Y			E_Z		
message /R strategy	$\{X\}$	$\{Y,Z\}$ $\{Y\}$	$\{Y,Z\}$ $\{Z\}$	$\{Y\}$	$\{X,Z\}$ $\{X\}$	$\{X,Z\}$ $\{Z\}$	$\{Z\}$	$\{X,Y\}$ $\{X\}$	$\{X,Y\}$ $\{Y\}$
s_Y^R	X	Y	Y	Z	Z	Z	Z	Z	Z
s_Z^R	X	Z	Z	Z	Z	Z	Z	Z	Z
s_{lit}^R	X	Y	Z	Z	Z	Z	Z	Z	Z

Table 7: Some language-based responses when the Receiver action set is X,Y,Z. The initial layer of recommendation is omitted given that it is common to all messages.

$M(A_0...A_j)$, or simply a *message block*. When A_j contains at least three elements, $M(A_0...A_j)$ contains multiple component message blocks that reflect different ways to partition A_j into two subsets.

I capture a common language by assuming that the Receiver either ignores or follows the opposing literal meanings of $M(A_0...A_jA_{j+1})$ vs. $M(A_0...A_j(A_j \backslash A_{j+1}))$. Formally, the language talk game G_L is the standard cheap talk game G with the restriction that the Receiver only plays *language-based responses*, defined as follows.[6]

Definition 1 *$s^R : M \to A^R$ is a language-based response if, for every decreasing sequence $A^R = A_0 \supsetneq A_1.... \supsetneq A_k \neq \emptyset$, either s^R is constant on the message block $M(A_0...A_{k-1}A_k) \cup M(A_0...A_{k-1}(A_{k-1} \backslash A_k))$ or s^R maps $M(A_0...A_{k-1}A_k)$ into A_k and $M(A_0...A_{k-1}(A_{k-1} \backslash A_k))$ into $A_{k-1} \backslash A_k$.*

The language assumption does not restrict the set of actions a message may induce, because the strategy that responds to all messages with action a^R, called $Constant - a^R$, is a language-based response, for any $a^R \in A^R$.

Instead, the language assumption restricts how the Receiver's response to one message relates to his response to another message. It creates asymmetry between messages related by negation: for any $m' \in M(A_1...A_kA_{k+1})$, $m'' \in M(A_1...A_k(A_k \backslash A_{k+1}))$ and any language-based response s^R, either

[6]The Receiver's mixed strategy set thus consists of all probability distritions over language-based responses.

$s^R(m') = s^R(m'')$ or $\left(s^R(m'), s^R(m'')\right) \in A_{k+1} \times (A_k \backslash A_{k+1})$. This contrasts with standard cheap talk games where $\left(s^R(m'), s^R(m'')\right)$ can be any pair in $A^R \times A^R$. This property allows the Sender to signal her preference over the Receiver's actions, which is informative of her intentions in self-signaling games.

The set of language-based responses retains some product structure. Let E be a union of component message blocks of $M(A_0...A_k)$. Say that s^R is language-based on E if s^R is a language-based response and s^R maps E into A_k. Say a mapping from E into A^R is language-based if it is the restriction of s^R to E where s^R is language-based on E. We observe that the set of language-based mappings on $M(A_0....A_k)$ is the product set of language-based mappings on each of its component message blocks. In addition, if we identify message $A_0...A_k...A_n$ when $A^R = A_0$ with message $A_k...A_n$ when $A^R = A_k$, then a mapping is language-based on $M(A_0...A_k)$ when $A^R = A_0$ if and only if it is language-based on $M(A_k)$ when $A^R = A_k$.

Consider for example $A^R = \{X, Y, Z\}$. There are three component message blocks of $M(\{X, Y, Z\})$ as shown in Table 7. By the language assumption, a language-based mapping on the message block E_X is either constant on E_X at $a^R \in \{X, Y, Z\}$ or is equal on E_X to one of the strategies in Table 7. This completes the description of all language-based responses by the product structure and symmetry between E_X, E_Y and E_Z.

The language assumption alone does not deliver the results of this paper. In fact, the set of equilibrium outcomes in the language-talk game is the same as that in the standard cheap talk game. To see this, note that for every Receiver strategy in the standard cheap talk game, there exists a language-based response with the same image. The aforementioned claim follows immediately because no restriction is made on the Sender's strategy space.

3 The Solution Concept

Let $X^i \subset S^i$ be a subset of player i's pure strategies. Denote by ΔX^i the set of probability distributions on X^i, and by $\Delta^+ X^i$ that whose support equals X^i. The following definitions are adapted from Brandenburger et al (2004), who provided an epistemic foundation for iterative admissibility.

Definition 2 *A strategy s^i is weakly dominated by $\hat{\sigma}^i$ with respect to X^{-i} if $u^i\left(\hat{\sigma}^i, s^{-i}\right) \geq u^i\left(s^i, s^{-i}\right)$ for every $s^{-i} \in X^{-i}$ and $u^i\left(\hat{\sigma}^i, \hat{s}^{-i}\right) > u^i\left(s^i, \hat{s}^{-i}\right)$ for some $\hat{s}^{-i} \in X^{-i}$. It is weakly dominated with respect to X if it is weakly dominated by some $\hat{\sigma}^i \in \Delta X^i$ w.r.t. X^{-i}.*

Definition 3 *Set $S^i(0) = S^i$ for $i \in \{S, R\}$ and iteratively define*

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

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.

Prior studies have shown that when there are only two players, a strategy is weakly dominated if and only if it is never a best response to any totally mixed conjecture.

Lemma 3.1 *A strategy $\hat{s}^i \in X^i$ is not weakly dominated w.r.t. X if and only if there exists $\hat{\sigma}^j \in \Delta^+ X^j$ to which $\hat{s}^i$ is a best response among X^i.*

Our assumption about the stage game g implies that every strategy profile is iteratively admissible in the standard cheap talk game. Because the language talk game differs from the standard cheap talk game only in the Receiver's strategy set, we have

$$S^S(2k+2) = S^S(2k+1) \text{ and } S^R(2k+1) = S^R(2k) \text{ for } k = 0, 1, 2, ... \tag{1}$$

The following lemma is useful to show that $S^R(n)$ contains strategies with certain properties. It follows from lemma 3.1 and continuity of the expected utility function.

Lemma 3.2 *For any $n = 0, 1, ...,$ at least one of player i's best responses to $\sigma^{-i} \in \Delta S^{-i}(n)$ belongs to $S^i(n+1)$.*

4 The role of Self-signaling

By Aumann (1990), the Sender wants a self-signaling recommendation to be followed only if the recommendation is optimal for the Receiver. There are many different ways to extend Aumann's definition of self-signaling beyond 2×2 games because whether the Sender wants a recommendation to be followed depends on what alternative actions the Sender can induce by communication. Baliga and Morris (2002) propose the following definition that I call *BM self-signaling*.

Definition 4 (*BM self-signaling*) *The stage game g is BM self-signaling if $g^S\left(a^S, b^R\left(a^S\right)\right) > g^S\left(a^S, a^R\right)$ for every $\left(a^S, a^R\right) \in A^S \times A^R$ where $a^R \neq b^R\left(a^S\right)$.*

In words, g is *BM* self-signaling if, out of all Receiver actions, the Sender prefers the Receiver to take his best response to her action, regardless of the action she plans to take. In that case, if the Sender can induce any Receiver action with surety while holding her own action fixed, she will induce the Receiver's best response to her own action, whichever it is.

If communication guarantees coordination on a set of stage game Nash equilibria, which one will the Sender choose? Because b^R and b^S are inverse functions of each other, we can associate each Sender and Receiver action with a unique Nash equilibrium. Sender actions and Receiver actions can thus be ranked according to the Sender's payoff in the corresponding stage game Nash equilibrium. I call a Sender action that ranks the highest the

Sender's Stackelberg action, and its associated Nash equilibrium payoff her Stackelberg payoff.

Given any Sender action, the Sender always prefers the Receiver's best response to her own action to any lower ranking Receiver action.[7] I define a weaker self-signaling condition that requires the Sender to prefer the Receiver's best response to her own action to at least one higher ranking Receiver action.

Definition 5 (*weak self-signaling*) *The stage game g is weak self-signaling if, for every a^S which is not a Stackelberg action, there exists $\tilde{a}^S$ where $g^S\left(\tilde{a}^S, b^R\left(\tilde{a}^S\right)\right) > g^S\left(a^S, b^R\left(a^S\right)\right)$ and $g^S\left(a^S, b^R\left(a^S\right)\right) > g^S\left(a^S, b^R\left(\tilde{a}^S\right)\right)$.*

I will show that the *BM* self-signaling condition is sufficient, but not necessary, to guarantee the Sender her Stackelberg payoff, while the weak self-signaling condition is necessary but not sufficient.

4.1 Sufficiency

Proposition 1 *If the coordination game g is BM self-signaling, then every iteratively admissible strategy profile $\left(\left(m, a^S\right), s^R\right)$ in the language talk game G_L gives the Sender her Stackelberg payoff.*

(All omitted proofs are in the Appendix.)

When can two language-based responses, $\tilde{s}_E^R$ and s^R, be pasted together to form another language-based response? Let E denote a union of component message blocks of $M\left(A_0 ... A_k\right)$. Define $T\left(\tilde{s}_E^R, s^R, E\right) : M \to A^R$ to equal $\tilde{s}_E^R$ on E and equal s^R elsewhere. Say that s^R is *free* on E if $T\left(\tilde{s}_E^R, s^R, E\right)$ is a language-based response for any $\tilde{s}_E^R$ language-based on E. By the language assumption, s^R is free on E if it is language-based on both

[7] Suppose a_2^R ranks lower than $b^R\left(a_1^S\right)$. Then $g\left(a_1^S, b^R\left(a_1^S\right)\right) > g^S\left(b^S\left(a_2^R\right), a_2^R\right) > g^S\left(a_1^S, a_2^R\right)$ where the first inequality comes from the ranking and the second inequality holds because $\left(b^S\left(a_2^R\right), a_2^R\right)$ is a Nash equilibrium by the coordination game property.

$M(A_0...A_{k-1}A_k)$ and $M(A_0...A_{k-1}(A_{k-1}\backslash A_k))$, or equivalently if it is non-constant on $M(A_0...A_{k-1}A_k) \cup M(A_0...(A_{k-1}\backslash A_k))$. If $k = 0$, then every language-based response is free on E.

To prove proposition 1, I show that as the iteration n increases, coordination happens in $S(n)$ conditional on a larger set of messages. To be more precise, let E be a component message block of $M(A_0...A_k)$ and X^R a set of language-based responses. Say that $\left(m, a^S\right)$ guarantees coordination conditional on E w.r.t. X^R if $s^R(m) = b^R\left(a^S\right)$ for every $s^R \in X^R$ *free* on E. Lemma 4.1 gives a sufficient condition for coordination to be conditionally guaranteed.

Lemma 4.1 *Let E be a message block. If there exists a language-based response $\hat{s}^R$ which is language-based on E and is a best response to every Sender strategy $\left(m, a^S\right) \in S^S(n-1)$ where $m \in E$, then every Sender strategy in $S^S(n-1)$ that uses a message in E guarantees coordination conditional on E w.r.t. $S^R(n)$.*

Proof. It follows by noting that if s^R is *free* on E, then $T\left(\hat{s}^R, s^R, E\right)$ is a language-based response, equals s^R outside of E, and is a best response to every Sender strategy $\left(m, a^S\right) \in S^S(n-1)$ where $m \in E$. ■

I will illustrate the proof for proposition 1 using the 3×3 coordination game in Table 6, where the Sender prefers Nash equilibrium (x, X) to (y, Y) to (z, Z). The formal proof is relegated to the Appendix.

By definition, a language-based response non-constant on the message block $M(\{X, Y, Z\}\{X\}) \cup M(\{X, Y, Z\}\{Y, Z\})$ responds to "$\{X, Y, Z\}\{X\}$" with action X and to any message in $M(\{X, Y, Z\}\{Y, Z\})$ with some action in $\{Y, Z\}$. By BM self-signaling, (m, x) is weakly dominated by ("$\{X, Y, Z\}\{X\}$", x) w.r.t. $S^R(0)$ for any $m \in M(\{X, Y, Z\}\{Y, Z\})$. Similarly, a language-based response non-constant on $\{$"$\{X, Y, Z\}\{Y, Z\}\{Y\}$"$\} \cup$ $\{$"$\{X, Y, Z\}\{Y, Z\}\{Z\}$"$\}$ responds to the former with Y and the latter with Z. So ("$\{X, Y, Z\}\{Y, Z\}\{Y\}$", z) is weakly dominated w.r.t. $S(0)$ by ("$\{X, Y, Z\}\{Y, Z\}\{Z\}$", z). Similarly, ("$\{X, Y, Z\}\{Y, Z\}\{Z\}$", y) is

weakly dominated w.r.t. $S^R(0)$ by $(\text{“}\{X,Y,Z\}\{Y,Z\}\{Y\}\text{”}, y)$. Therefore, any recommendation in $M(\{X,Y,Z\}\{Y,Z\})$ must be optimal for the Receiver. Define the *Literal* strategy to respond to message $m =$“$A_0...A_n$” with the action in the singleton set A_n, for every hierarchical recommendation m. We now have the following claim.

Claim 1 *The Literal strategy is a Receiver best response to every Sender strategy in $S^S(1)$ that uses a message in $M(\{X,Y,Z\}\{Y,Z\})$.*

But no language-based response is a best response to every Sender strategy in $S^S(1)$. For example, $(\text{“}\{X,Y,Z\}\{X\}\text{”}, x)$ and $(\text{“}\{X,Y,Z\}\{X\}\text{”}, y)$ both belong to $S^S(1)$. Here is why. If the Sender believes that with sufficiently high probability the Receiver plays $Constant - X$ $(Constant - Y)$, then the Receiver's best response must involve taking action x (y). If, in addition, the Sender believes that with the complementary probability the Receiver uses strategy s_Z^R, which responds to message “$\{X,Y,Z\}\{X\}$” with action X and to all other messages with action Z, then the Sender will send message “$\{X,Y,Z\}\{X\}$,” whether her planned action is x or y. Even though g is BM self-signaling, the Sender might have an incentive to recommend a suboptimal action if she believes that, when communication has an impact, she cannot induce the Receiver's best response to her intended action.

Now I show that $(\text{“}\{X,Y,Z\}\{X\}\text{”}, y)$ is weakly dominated w.r.t. $S^R(2)$.

I first show that a dominator exists. Consider the belief described above to which $(\text{“}\{X,Y,Z\}\{X\}\text{”}, y)$ is a best response. I modify this belief by substituting s_Z^R with $T\left(Literal, s_Z^R, M(\{X,Y,Z\}\{Y,Z\})\right)$ (denoted by s_{lit}^R in Table 7), which equals *Literal* on $M(\{X,Y,Z\}\{Y,Z\})$ and s_Z^R elsewhere. Because s_Z^R is free on $M(\{X,Y,Z\}\{Y,Z\})$, $T\left(Literal, s_Z^R, M(\{X,Y,Z\}\{Y,Z\})\right)$ is a language-based response. The Sender's best response to this modified belief is then $(\text{“}\{X,Y,Z\}\{Y,Z\}\{Y\}\text{”}, y)$. Therefore,

$$(\text{“}\{X,Y,Z\}\{Y,Z\}\{Y\}\text{”}, y) \text{ belongs to } S^S(1). \tag{2}$$

By (2), claim 1, lemma 4.1 and the language assumption, every $s^R \in S^R(2)$ either responds to “$\{X,Y,Z\}\{X\}$” with X and “$\{X,Y,Z\}\{Y,Z\}\{Y\}$” with

Y or it responds to both with the same action. Therefore, ("$\{X,Y,Z\}\{Y,Z\}\{Y\}$", y) fares at least as well as ("$\{X,Y,Z\}\{X\}$", y) against every strategy in $S^R(2)$.

I then show that ("$\{X,Y,Z\}\{Y,Z\}\{Y\}$", y) and ("$\{X,Y,Z\}\{X\}$", y) are not equivalent against every strategy in $S^R(2)$. A Receiver best response to the conjecture that places equal weights on ("$\{X,Y,Z\}\{X\}$", x) $\in S^S(1)$ and ("$\{X,Y,Z\}\{Y,Z\}\{Y\}$", y) $\in S^S(1)$ must respond to "$\{X,Y,Z\}\{X\}$" and "$\{X,Y,Z\}\{Y,Z\}\{Y\}$" with different actions. One such strategy must survive $S^R(2)$ by lemma 3.2.

Therefore, ("$\{X,Y,Z\}\{X\}$", y) is weakly dominated by ("$\{X,Y,Z\}\{Y,Z\}\{Y\}$", y) w.r.t. $S^R(2)$. So ("$\{X,Y,Z\}\{X\}$", y) does not belong to $S^S(3)$. Similarly we can show that ("$\{X,Y,Z\}\{X\}$", z) does not belong to $S^S(3)$.

So *Literal* strategy is a best response to every $\left(m, a^S\right) \in S^S(3)$ where $m \in M(\{X,Y,Z\}\{X\}) \cup M(\{X,Y,Z\}\{Y,Z\}) = E_X$. I next show that some message must guarantee coordination on (x,X) w.r.t. $S^R(4)$. If ("$\{X,Y,Z\}\{X\}$", x) $\in S^S(3)$, then by lemma 4.1, s^R("$\{X,Y,Z\}\{X\}$") $= X$ for all $s^R \in S^R(4)$ because every language-based response is free on E_X. If ("$\{X,Y,Z\}\{X\}$", x) $\notin S^S(3)$, then ("$\{X,Y,Z\}\{X\}$", x) must be weakly dominated by (m^*,x) w.r.t. $S^R(2)$ for some $m^* \in E_Y \cup E_Z$. Then, for every $s^R \in S^R(2) \supseteq S^R(4)$, we must have $s^R(m^*) = X$ because $T\left(Literal, s^R, E_X\right) \in S^R(2)$ responds to "$\{X,Y,Z\}\{X\}$" with X and to message m^* with $s^R(m^*)$ and because the Sender prefers Receiver action X given her intended action x.[8] Then (m^*,x) guarantees coordination w.r.t. $S^R(2) \supseteq S^R(4)$.

Therefore, in $S(5)$, every Sender strategy must give the Sender her Stackelberg payoff.

Weak dominance reasoning combined with the asymmetry between messages allow the Sender to convey information about her preference over A^R.

[8] $T\left(Literal, s^R, E_X\right)$ belongs to $S^R(2)$ because every language-based response is free on E_X.

			Receiver	
		X	Y	Z
	x	$12, 12$	$1, 1$	$2, 0$
Sender	y	$6, 1$	$10, 10$	$4, 0$
	z	$9, 0$	$6, 1$	$8, 8$

Table 8: The Sender gets her Stackelberg payoff in every iteratively admissible outcome of the language talk extension to this coordination game that is not BM self-signaling

This indirectly enables the Sender to convey information about her intention because her most preferred Receiver action changes with her intended action when the stage game g is BM self-signaling.

4.2 Necessity

Given the mechanism underlying the sufficiency result, it is intuitive that communication does not guarantee any information transmission if the Sender's preference over A^R is invariant with her own intention.

Proposition 2 *Suppose the Sender's preference over the Receiver's actions is independent of her own action, that is, for any* $a_1^R, a_2^R \in A^R$ *and* $a^S, \tilde{a}^S \in A^S$, $g^S\left(a^S, a_1^R\right) > g^S\left(a^S, a_2^R\right)$ *if* $g^S\left(\tilde{a}^S, a_1^R\right) > g^S\left(\tilde{a}^S, a_2^R\right)$. *Then, for every* $\left(a^S, a^R\right) \in A$, *an iteratively admissible strategy profile* $\left(\left(m, a^S\right), s^R\right)$ *exists where* $s^R(m) = a^R$.

The proof involves showing that, for any $\tilde{a}^S$,$\hat{a}^S \in A^S$ and $\hat{m} \in M$, $\left(\hat{m}, \hat{a}^S\right)$ is iteratively admissible if $\left(\hat{m}, \tilde{a}^S\right)$ is iteratively admissible. Thus, all conjectures about the Sender's action can be rationalized regardless of the message sent.

When can communication guarantee coordination on the Sender's best stage game Nash equilibrium? Proposition 2 illustrates that the Sender's preference over A^R has to vary with her own action. Example 1, however, shows that the BM self-signaling condition is not necessary.

Example 1 *In the game shown in Table 8, the Sender prefers Nash equilibrium (x, X), to (y, Y) to (z, Z). The stage game is not BM self-signaling because the Sender prefers Receiver action X to Z to Y whether her own action is x or z. I will show that the Sender nonetheless obtains her Stackelberg payoff in every iteratively admissible outcome.*

First I show that $\left(\text{“}\{X,Y,Z\}\{Y\}\text{”}, a^S\right)$ belongs to $S^S(1)$ if and only if $a^S = y$. If the Sender plans to take action y, then she prefers Receiver action Y to any action in $\{X, Z\}$. If the Sender plans to take action x or z, then she prefers any action in $\{X, Z\}$ to Y. Note also that, if a language-based response is non-constant on $E_Y = \{\text{“}\{X,Y,Z\}\{Y\}\text{”}\} \cup M(\{X,Y,Z\}\{X,Z\})$, it responds to message “$\{X,Y,Z\}\{Y\}$” with action Y and message $m \in M(\{X,Y,Z,\}\{X,Z\})$ with action X or Z. Therefore, $\left(\text{“}\{X,Y,Z\}\{Y\}\text{”}, a^S\right)$ is weakly dominated by $\left(\text{“}\{X,Y,Z\}\{X,Z\}\{X\}\text{”}, a^S\right)$ for $a^S = x, z$. This establishes the “only if” part. The “if” part follows because $(\text{“}\{X,Y,Z\}\{Y\}\text{”}, y)$ is the unique best response to the language-based response that responds to “$\{X,Y,Z\}\{Y\}$” with Y and all other messages with Z.

Therefore, a Receiver strategy $\hat{s}^R$ which does not respond to message “$\{X,Y,Z\}\{Y\}$” with $Y = b^R(y)$ is weakly dominated w.r.t. $S^S(1)$ by $\phi\left(\hat{s}^R\right)$ which responds to “$\{X,Y,Z\}\{Y\}$” with Y and equals $\hat{s}^R$ elsewhere.. Note that $\phi\left(\hat{s}^R\right)$ is a language-based response because $\hat{s}^R$ must be constant at X or Z on $M(\{X,Y,Z\}\{X,Z\})$. Hence, $s^R(\text{“}\{X,Y,Z\}\{Y\}\text{”}) = Y$ for every $s^R \in S^R(2)$.

Because playing action z gives the Sender at most $9 < g^S(y, Y)$, every Sender strategy that chooses action z is dominated by $(\text{“}\{X,Y,Z\}\{Y\}\text{”}, y)$ w.r.t. $S^R(2)$. Note that the payoff matrix restricted to $\{x, y\} \times \{X, Y\}$ is BM self-signaling. The conclusion then follows from arguments similar to those for the battle-of-the-sexes game.

The mechanism that guarantees the Sender her Stackelberg payoff in example 1 breaks down if the stage game g is not weak self-signaling. In that case, holding the Sender's action fixed at some $\hat{a}^S$, she has an incentive to

			Receiver	
		X	Y	Z
	x	$12, 12$	$1, 1$	$2, 0$
Sender	y	$6, 1$	$10, 10$	$4, 0$
	z	$11, 0$	$6, 1$	$8, 8$

Table 9: A weak self-signaling coordination game where miscoordination happens in an iteratively admissible outcome

message block	E_X			E_Y			E_Z		
Receiver strategy /message	$\{X\}$	$\{Y,Z\}$ $\{Y\}$	$\{Y,Z\}$ $\{Z\}$	$\{Y\}$	$\{X,Z\}$ $\{X\}$	$\{X,Z\}$ $\{Z\}$	$\{Z\}$	$\{X,Y\}$ $\{X\}$	$\{X,Y\}$ $\{Y\}$
$S^S(1)$	x,z,y	y	$\emptyset$	y	x,z	$\emptyset$	x,z	x,z	y
$S^R(2)$	X	Y	Z	Y	X	Z	Z	X	Y
	X	Y	Y	Y	X	X			
	Y	Y	Y	Y	Z	Z	X	X	X
	Z	Z	Z				Z	Z	Z
$S^S(3)$	x,z	$\emptyset$	$\emptyset$	y	x,z	$\emptyset$	$\emptyset$	x,z	$\emptyset$
$S^R(4)$	X	Y	Z	Y	X	Z	Z	X	Y
	X	Y	Y	Y	X	X	X	X	X
$\mathfrak{s}^R_{ZZ}$	Z	Z	Z	Y	Z	Z	Z	Z	Z
$S^S(5)$	x,z	$\emptyset$	$\emptyset$	y	x,z	$\emptyset$	$\emptyset$	x,z	$\emptyset$

Table 10: Iterative process for a weak self-signaling game that does not guarantee coordination

recommend any Receiver action with a higher ranking than what's optimal for the Receiver, i.e. $b^R\left(\hat{a}^S\right)$. Thus, after no iteration can the Sender guarantee coordination on a Nash equilibrium that gives a higher payoff than $\left(\hat{a}^S, b^R\left(\hat{a}^S\right)\right)$.

Proposition 3 *If a coordination game g is not weak self-signaling, then there exists an iteratively admissible strategy profile $\left(\left(m, a^S\right), s^R\right)$ where $s^R(m) \neq b^R\left(a^S\right)$.*

The weak self-signaling condition is, however, not sufficient to guarantee coordination, as example 2 shows.

Example 2 *In the coordination game shown in Table 9, the Sender prefers Nash equilibrium (x, X) to (y, Y) to (z, Z). The Sender prefers Receiver action X to Z to Y if her own action is x or z, while she prefers Receiver action Y to X to Z if her action is y. So this game is weak self-signaling but not BM self-signaling. In fact, for any Sender action a^S, the Sender has the same ordinal preference over Receiver actions as that in the stage game of example 1. I will show that not every iteratively admissible outcome in this game gives the Sender her Stackelberg payoff.*

We can infer all strategies surviving iteration 1-5 from Table 10 by (1).[9] *Sender action a^S is listed under column m and row $S^S(n)$ if and only if $\left(m, a^S\right)$ belongs to $S^S(n)$. Receiver strategy s^R belongs to $S^R(n)$ if and only if s^R restricted to each message block E_X, E_Y, E_Z is listed in one of the rows for $S^R(n)$.*[10] *For example, $S^R(4)$ contains the Receiver strategy $\mathfrak{s}^R_{Zlit}$ that equals Literal on E_Y and is constant at Z on $E_X \cup E_Z$; it also contains the Receiver strategy $\mathfrak{s}^R_{ZZ}$ that responds to "$\{X, Y, Z\}\{Y\}$" with Y and to all other messages with Z.*

As in example 1, in $S(2)$ the Sender can guarantee coordination on (y, Y) by using strategy ("$\{X, Y, Z\}\{Y\}$", y). But the analysis for the two examples diverges from here: if the Receiver plays action X with probability $\frac{3}{4}$ and Z with probability $\frac{1}{4}$, then z is the Sender's stage game best response, which gives her a payoff of $\frac{41}{4} > g^S(y, Y)$. Therefore, not all strategies using action z are dominated w.r.t. $S^R(2)$ by ("$\{X, Y, Z\}\{Y\}$", y) as in example 1. In fact, ("$\{X, Y, Z\}\{X, Z\}\{X\}$", z) is the unique best response to $\frac{3}{4}\mathfrak{s}^R_{Zlit} + \frac{1}{4}\mathfrak{s}^R_{ZZ} \in \Delta S^R(2)$.[11] *It thus belongs to $S^S(3)$.*

---

[9] If the Sender believes that with high probability the Receiver uses the strategy $Constant - a^R$, then the Sender will take action $b^S\left(a^R\right)$. Thus, $\left(\text{"}A_0..\left\{a^R\right\}\text{"}, a^S\right)$ belongs to $S^S(1)$ if and only if a^R is not the worst Receiver action holding Sender's action fixed at a^S. The omitted derivations of the rest of the table are in the online Appendix.

[10] This is because the set of language-based responses is the product set of language-based mappings on E_X, E_Y and E_Z.

[11] Here is why s^R_{Zlit} and s^R_{ZZ} belong to $S^R(2)$. Every best response to a conjecture uniform on $\{(\text{"}\{X\}\text{"}, z), (\text{"}\{X, Y\}\{X\}\text{"}, z), (\text{"}\{Y\}\text{"}, y), (\text{"}\{X, Z\}\{X\}\text{"}, x)\}$ must equal s^R_{Zlit} except on "$\{X, Z\}\{Z\}$" which is not used in $S^S(1)$. s^R_{ZZ} is the unique best response to the conjecture uniform on

By Table 10 and (1), $S(5) = S(4)$*. Thus* $S(4) = S(\infty)$*. Coordination is not guaranteed because* $\left(\left(\text{“}\{X,Y,Z\}\{XZ\}\{X\}\text{”}, x\right), \mathfrak{s}^R_{ZZ}\right)$ *is iteratively admissible and gives rise to outcome* (x, Z)*.*

In the two examples, the Receiver has the same payoff matrix, while the Sender has the same ordinal preference over Receiver actions for any intended action. In fact, the Sender's payoff at (z, X) is the only difference between the two games. However, the Sender is guaranteed her Stackelberg payoff in example 1, but not in example 2. Therefore, no condition on the Sender's ordinal preference over Receiver actions can be both necessary and sufficient to guarantee the Sender her Stackelberg payoff.

5 The Role of Hierarchical Recommendations

The nesting structure of the language-talk model can be thought of as a reasoning tool. If message "$\{X,Y,Z\}\{X\}$" is credible, then the Sender should be able to induce action X for sure by sending message "$\{X,Y,Z\}\{X\}$". As mentioned in the introduction, whether the Sender might want to send message "$\{X,Y,Z\}\{X\}$" when she plans to take action y depends on the set of all alternative actions the Sender can induce. Therefore, to reason about the credibility of message "$\{X,Y,Z\}\{X\}$", one asks: if the Sender wants the Receiver NOT to take action X (i.e. to take an action in $\{Y,Z\}$), can the Sender induce action Y for sure? Using the hierarchical structure of the language assumption, we can ask: suppose the Sender wants to induce an action in $\{Y,Z\}$ (i.e. not X), when would the Sender want to induce action Y instead of Z?

This paper employs the above reasoning to endogenously determine the set of alternative actions. In contrast, Farrell (1993) assumes that the set of alternative actions consists of all Receiver actions in the candidate equilibrium. Applying Farrell's approach presumes the stability of the very equilibrium whose stability is in question. This concern is often called the *Stiglitz*

$\{(\text{“}\{X\}\text{”}, z), (\text{“}\{X,Y\}\{X\}\text{”}, z), (\text{“}\{Y\}\text{”}, y), (\text{“}\{X,Z\}\{X\}\text{”}, z)\}$.

critique. Rabin (1990) assumes maximal communication to select among multiple solutions. Applying Rabin's approach assumes away the necessity of self-signaling by fiat, whereas the necessity of self-signaling is among the core issues we investigate.

The following two examples highlight the role of hierarchical recommendations.

5.1 Crude Language

Consider the language talk game in Section 2 with a smaller message space $M^{cr} = \left\{ \text{“}\{A_0\}\{A_1\}\text{”} : \emptyset \neq A_1 \subsetneq A_0 = A^R \right\}$. Call this game the crude-language game $G_{L^{cr}}$.

Consider the stage game in Table 6, which is used to illustrate the sufficiency result in section 4.1. Table 11 lists in the second row all messages in M^{cr}. The initial layer $\{X, Y, Z\}$ common to all messages is omitted to save space. By (1), all strategies surviving iteration 1-3 in the crude language game can be inferred from the table. Sender action a^S is listed at the intersection of the column for message m and the row for $S^S(n)$ if and only if $\left(m, a^S\right)$ belongs to $S^S(n)$. Receiver strategy s^R belongs to $S^R(n)$ if s^R restricted to each message block E_X, E_Y, E_Z is listed in one of the rows for $S^R(n)$. Then $S(\infty) = S(2)$ because $S(3) = S(2)$. Miscoordination is not ruled out because Sender strategies $(\text{“}\{X, Y, Z\}\{X\}\text{”}, y)$ and $(\text{“}\{X, Y, Z\}\{X\}\text{”}, x)$ are both iteratively admissible.

The strategy $(\text{“}\{X, Y, Z\}\{X\}\text{”}, y)$ survives the first iteration in the crude language game $G_{L^{cr}}$ for the same reason it does in the language talk game G_L. Why, then, does it survive the third iteration under the crude language model but not under the rich language model?[12]

Instead of having a smaller message set, we can think of the crude-language game as the language talk game with the additional restriction that the Receiver ignores all higher level recommendations. That is, under the crude language model, the Receiver must respond to “$\{X, Y, Z\}\{Y, Z\}\{Y\}$”

[12]Recall that $S^S(1) = S^S(2)$ by (1).

message block	E_X		E_Y		E_Z	
Receiver strategy /message	$\{X\}$	$\{Y,Z\}$	$\{Y\}$	$\{X,Z\}$	$\{Z\}$	$\{X,Y\}$
$S^S(1)$	x,y	y,z	y,z	x,z	x,z	x,y
$S^R(2)$	X	Y	Y	X	Z	X
	X	Z	Y	Z	Z	Y
	Y	Y	X	X	Z	Z
	Z	Z	Z	Z		
$S^S(3)$	x,y	y,z	y,z	x,z	x,z	x,y

Table 11: Iterative process for the crude language model.

and "$\{X,Y,Z\}\{Y,Z\}\{Z\}$" with the same action, whereas under the rich language model, the Receiver may respond with different actions, that is, he responds to "$\{X,Y,Z\}\{Y,Z\}\{Y\}$" with Y and "$\{X,Y,Z\}\{Y,Z\}\{Z\}$" with Z. The richer language thus allows the Sender to signal her preference between Receiver action Y and Z. This property together with the BM self-signaling condition implies that, in $S^S(1)$, the Sender uses the message "$\{X,Y,Z\}\{Y,Z\}\{Y\}$" only when she plans to take action y, and "$\{X,Y,Z\}\{Y,Z\}\{Z\}$" only when she plans to take action z. This continues the unraveling process that ends with a unique outcome. In contrast, in the crude language model, "$\{X,Y,Z\}\{Y,Z\}\{Y\}$" and "$\{X,Y,Z\}\{Y,Z\}\{Z\}$" are essentially the same message. Therefore, both may be sent in $S^S(1)$ whether the Sender plans to take action y or z. The unraveling process thus stops and ("$\{X,Y,Z\}\{X\}$", y) survives the third iteration in the crude language model.

We can strengthen the self-signaling condition to obtain the sufficiency result in the crude language game.

Definition 6 *Let x denote the Sender's Stackelberg action and $X = b^R(x)$. The stage game g is strongly self-signaling if $g^S\left(a^S, X\right) < g^S\left(a^S, a^R\right)$ for any $a^R \neq X$, $a^S \neq x$.*

If the stage game is strongly self-signaling, then the Sender will never

want to induce the Receiver's best response to her Stackelberg action unless it is indeed optimal for the Receiver.

Proposition 4 *Suppose g is strongly self-signaling. Then every iteratively admissible outcome in the crude language game gives the Sender her Stackelberg payoff.*

Proof. $\left(\text{“}A^R\left\{X\right\}\text{”}, a^S\right)$ is weakly dominated by $\left(\text{“}A^R\left(A^R\backslash\left\{X\right\}\right)\text{”}, a^S\right)$ by the strong self-signaling condition and because, if the Receiver responds to the two messages with different actions, then he responds to "$A^R\left\{X\right\}$" with X and to "$A^R\left(A^R\backslash\left\{X\right\}\right)$" with some $a^R \neq X$. Thus $s^R\left(\text{“}A^R\left\{X\right\}\text{”}\right) = X$ for every $s^R \in S^R(2)$ because $T\left(Constant - X, s^R, \text{“}A^R\left\{X\right\}\text{”}\right) \in S^R(0)$ for every $s^R \in S^R(0)$. ■

5.2 Simple Language

What if we allow only simple messages that recommend exactly one action? Suppose the message space is

$$M^{sim} = \left\{\text{“}\left\{X,Y,Z\right\}\left\{X\right\}\text{”}, \text{“}\left\{X,Y,Z\right\}\left\{Y\right\}\text{”}, \text{“}\left\{X,Y,Z\right\}\left\{Z\right\}\text{”}\right\}.$$

The set of language-based strategies should certainly include all constant mappings to avoid assuming information transmission by fiat. The question is, which non-constant mappings should be included? If we apply the spirit that recommendations are either ignored or followed, then the only non-constant strategy responds to message "$\left\{X,Y,Z\right\}\left\{a^R\right\}$" with action a^R. It is then straightforward that the Sender is guaranteed her Stackelberg payoff if the stage game is BM self-signaling. But this assumption precludes informative equilibria in the standard cheap talk game where the range of the Receiver strategy is a strict subset of A^R. If we allow Receiver strategies with range $\{Y,Z\}$, then it seems natural that the Receiver should respond to message "$\left\{X,Y,Z\right\}\left\{Y\right\}$" with action Y, and to message "$\left\{X,Y,Z\right\}\left\{Z\right\}$" with action Z, but what about the message "$\left\{X,Y,Z\right\}\left\{X\right\}$"? By symmetry, we should allow both mapping s^R_{YYZ} and s^R_{ZYZ} in Table 12. Applying this

idea, the set of language-based responses should be the set of all constant mappings plus all those listed in Table 12. I call this the simple talk game $G_{L^{sim}}$.

Consider again the stage game in Table 6. Strategies surviving each iteration can be inferred from Table 13.[13] Then $S(\infty) = S(3)$. Coordination is not guaranteed because $(\text{“}\{X,Y,Z\}\{Y\}\text{”}, y)$ and $(\text{“}\{X,Y,Z\}\{Y\}\text{”}, z)$ are both iteratively admissible.

Let's identify the message “$\{X,Y,Z\}\{Y\}$” in the simple talk game with the message “$\{X,Y,Z\}\{Y,Z\}\{Y\}$” in the language talk game. In the language talk game, Sender strategy $(\text{“}\{X,Y,Z\}\{Y,Z\}\{Y\}\text{”}, z)$ is weakly dominated by $(\text{“}\{X,Y,Z\}\{Y,Z\}\{Z\}\text{”}, z)$ w.r.t. $S(1)$ because, when the Receiver responds to these two with different actions, he responds to message “$\{X,Y,Z\}\{Y,Z\}\{Y\}$” with Y and “$\{X,Y,Z\}\{Y,Z\}\{Z\}$” with Z. This force is absent in the simple language game because the Receiver may respond to “$\{X,Y,Z\}\{Y\}$” with Y and to “$\{X,Y,Z\}\{Z\}$” with X (s^R_{XYX} in Table 12). Because $(\text{“}\{X,Y,Z\}\{Y\}\text{”}, z)$ survives the first iteration, s^R_{XZZ} survives the second iteration, which justifies $(\text{“}\{X,Y,Z\}\{X\}\text{”}, y)$ in the third iteration, while $(\text{“}\{X,Y,Z\}\{X\}\text{”}, y)$ does not survive the third iteration in the language-talk game because $s^R_{XZZ} \notin S^R(2)$.

6 Conclusion

This paper formalizes and extends Aumann's idea that credibility of a recommendation can be inferred through introspection if and only if it is self-signaling. Instead of modelling a common language by restricting the Receiver's responses to individual messages, I restrict the relationship between the Receiver's responses to different messages. The asymmetry between mes-

[13] $\left(\text{“}\{X,Y,Z\}\left\{a^R\right\}\text{”}, a^S\right)$ survives $S^S(1)$ as long as a^R is not the worst Receiver action for the Sender when her planned action is a^S. The rest of the table follows easily by noting that action Z is optimal against a conjecture that puts probability $\frac{1}{2}$ on both Sender action x and y. For example, $Constant - Z$ belongs to $S^R(2)$ because it is the Receiver's unique best response to the conjecture $\frac{1}{6}(\text{“}\{X,Y,Z\}\{X\}\text{”}, x) + \frac{1}{6}(\text{“}\{X,Y,Z\}\{X\}\text{”}, y) + \frac{1}{3}(\text{“}\{X,Y,Z\}\{Y\}\text{”}, z) + \frac{1}{3}(\text{“}\{X,Y,Z\}\{Z\}\text{”}, z)$.

Receiver strategy /message	$\{X\}$	$\{Y\}$	$\{Z\}$
	X	Y	Y
	X	Y	X
s^R_{YYZ}	Y	Y	Z
s^R_{ZYZ}	Z	Y	Z
s^R_{XZZ}	X	Z	Z
	X	X	Z
Literal	X	Y	Z

Table 12: Non-constant language-based responses in the simple-talk game

Receiver strategy /message	$\{X\}$	$\{Y\}$	$\{Z\}$
$S^S(1)$	x, y	y, z	x, z
$S^R(2)$	X	X	X
	Z	Z	Z
	X	Y	X
	Y	Y	Z
	Z	Y	Z
	X	Z	Z
	X	X	Z
	X	Y	Z
$S^S(3)$	x, y	y, z	x, z

Table 13: Iterative process for the simple talk game

sages created by this language model together with weak dominance reasoning allow inference about the Sender's intention without assuming effective communication by fiat. This framework makes assumption choices explicit: the set of alternative messages and, more importantly, the set of alternative actions. These assumptions are crucial to results on communication.

A Appendix

A.1 Preliminaries

Given any m ="$A_1...A_n$", define $X\left(m, a^S\right)$ to be the message block $M\left(A_0....A_iA_{i+1}\right) \cup M\left(A_0....A_i\left(A_i \backslash A_{i+1}\right)\right)$ where A_i is the smallest set in $\{A_0, A_1, ..., A_n\}$ that contains $b^R\left(a^S\right)$. It follows from the language assumption that s^R is a best response to $\left(m, a^S\right)$ if and only if s^R is constant on $X\left(m, a^S\right)$ at $b^R\left(a^S\right)$. If $X\left(m_1, a_1^S\right) \cap X\left(m_2, a_2^S\right) \neq \emptyset$ and $b^R\left(a_1^S\right) \neq b^R\left(a_2^S\right)$, then no language-based response is a best response to both $\left(m_1, a_1^S\right)$ and $\left(m_2, a_2^S\right)$.

Lemma A.1 *If no language-based response exists that is a best response to every $s^S \in X^S$, then there exists $\left(m_1, a_1^S\right), \left(m_2, a_2^S\right) \in X^S$ such that no language-based response is a best response to both $\left(m_1, a_1^S\right)$ and $\left(m_2, a_2^S\right)$.*

Proof. Suppose for every $\left(m_i, a_i^S\right), \left(m_j, a_j^S\right)$ in $X^S = \left\{\left(m_1, a_1^S\right), ..., \left(m_K, a_K^S\right)\right\}$, either $X\left(m_i, a_i^S\right) \cap X\left(m_j, a_j^S\right) = \emptyset$ or $b^R\left(a_i^S\right) = b^R\left(a_j^S\right)$. Define s_0^R to be the *Literal* strategy. Iteratively define s_k^R by substituting s_{k-1}^R with $Constant - b^R\left(a_k^S\right)$ on $X\left(m_k, a_k^S\right)$. It is easy to check that s_k^R is a language-based response and is a best response to $\left(m_i, a_i^S\right)$ for $i = 1, ..., k$. We then obtain a contradiction. ■

A.2 Omitted Proofs for Proposition 1

A.2.1 Preliminaries

Some notations are in order. Let N denote the number of Receiver actions in the stage game g. Given any bijective function $\phi : \{1, 2, ..., N\} \to A^R$, define

$A_{\phi,0} = A^R$ and $A_{\phi,k} =: \{\phi(k+1), ..., \phi(N)\}$ for $k = 1, .., N-1$. Define

$$\begin{aligned} M_\phi(k) &= : M(A_{\phi,1}...A_{\phi,k}); \\ m_{\phi,k} &= : \text{“}A_{\phi,1}...A_{\phi,k-1}\{\phi(k)\}\text{”}. \end{aligned}$$

Then $m_{\phi,k} \cup M_\phi(k)$ is a component message block of $M_\phi(k-1)$. In words, $M_\phi(k)$ is the set of messages that say “Do not take action $\phi(1)$. Among the rest, do not take action $\phi(2)$. ... Among $\{\phi(k), ..., \phi(N)\}$, do not take action $\phi(k)$.”

Throughout this section, all statements hold for every bijective function ϕ, and the stage game g is assumed to be BM self-signaling.

Lemma A.2 *Given* $\left(m, a^S\right) \in S^S(1)$ *where* $m \in M_\phi(k)$*, we have* $b^R\left(a^S\right) \in A_{\phi,k}$.

Proof. Consider $\hat{m} \in M_\phi(k)$ and $\hat{a}^S$ where $b^R\left(\hat{a}^S\right) = \phi(j)$ for some $j \leq k$. If $s^R \in S^R(0)$ responds to $m_{\phi,j}$ and $\hat{m}$ with different actions (e.g. $Literal \in S^R(0)$), then by definition 1, $s^R(m_{\phi,j}) = \phi(j)$ and $s^R(\hat{m}) \in \{\phi(j+1), ..., \phi(N)\}$. Thus, $\left(\hat{m}, \hat{a}^S\right)$ is weakly dominated by $\left(m_{\phi,j}, \hat{a}^S\right)$ w.r.t. $S^R(0)$ by the BM self-signaling condition. ■

For any component message block F of $M_\phi(j)$, denote by $\Sigma_F(n)$ the set of $\left(m, a^S\right) \in S^S(n-1) \cap F \times A^S$ that does NOT guarantee coordination conditional on F w.r.t. $S^R(n)$. Say that E is parallel to F if they are different component message blocks of $M_\phi(j)$.

The core idea of lemma A.3 below has an analogue in the standard cheap talk game. Consider a standard cheap talk game and message $m_E \neq m_F$. Because the Receiver's pure strategy set in the standard cheap talk game is the product set M^{A^R}, every strategy s^R is free on both m_E and m_F. Suppose $\left(m_E, a_E^S\right)$, $\left(m_F, a_F^S\right) \in S^S(n-1)$. Then $\left(m_F, a_F^S\right) \in \Sigma_F(n)$ if and only if there exists $\left(m_F, \hat{a}_F^S\right) \in S^S(n-1)$ where $b^R\left(a_F^S\right) \neq b^R\left(\hat{a}_F^S\right)$. It is easy to see that, given any $s^R \in S^R(n)$ in the standard cheap talk game, there exists $\psi_{\left(m_F, \hat{a}_F^S\right)}\left(s^R\right)$ that is a best response to $\left(m_E, a_E^S\right)$ and $\left(m_F, \hat{a}_F^S\right)$ (and hence not a best response to $\left(m_F, a_F^S\right)$) and equal to s^R outside of

$\{m_E, m_F\}$. Lemma A.3 is more complicated to state and prove because the set of all language-based responses is not the product set M^{A^R}.

Lemma A.3 *Fix $j = 1, ..., N-3$. Suppose $\Sigma_F(n) \neq \emptyset$ for every component message block F of $M_\phi(j)$. Then*

1. *for any $s^R \in S^R(n)$ non-constant on $m_{\phi,j} \cup M_\phi(j)$, any component message block E of $M_\phi(j)$, and any $\left(m_E, a_E^S\right) \in S^S(n-1)$ where $m_E \in E$, $\varphi_{\left(m_E, a_E^S\right)}\left(s^R\right) \in \Delta S^R(n)$ exists that equals s^R on $m \notin M_\phi(j)$, is a best response to $\left(m_E, a_E^S\right)$, but is not a best response to any $\left(m, a^S\right) \in \Sigma_F(n)$ where F is parallel to E;*

2. *if, in addition, every $s^R \in S^R(n)$ is constant on the component message block F of $M_\phi(j)$, and no language-based response exists that is a best response to both $\left(m_1, a_1^S\right) \in S^S(n+1)$ and $\left(m_2, a_2^S\right) \in S^S(n-1)$ where $m_1, m_2 \in F$, then there exists $\sigma^R_{\left(m_2, a_2^S\right)} \in \Delta S^R(n)$ to which every Sender best response uses a message in F and an action in $A^S \backslash \left\{a_2^S\right\}$.*

Proof. Fix a component message block E of $M_\phi(j)$, a Sender strategy $\left(m_E, a_E^S\right) \in S^S(n-1) \cap E \times A^S$ and Receiver strategy $s^R \in S^R(n)$ non-constant on $m_{\phi,j} \cup M_\phi(j)$. I will show that, for every message block F parallel to E, every $\left(m_F, a_F^S\right) \in \Sigma_F(n)$, $\psi^{\left(m_F, a_F^S\right)}\left(s^R\right)$ exists in $S^R(n)$ where

$$\psi^{\left(m_F, a_F^S\right)}\left(s^R\right)(m) \begin{cases} = s^R(m) & \text{if } m \notin M_\phi(j) \\ = b^R\left(a_E^S\right) & \text{if } m = m_E \\ \neq b^R\left(a_F^S\right) & \text{if } m = m_F \end{cases} \tag{3}$$

Statement 1 follows by defining $\varphi_{\left(m_E, a_E^S\right)}\left(s^R\right)$ to be uniformly distributed on

$$\left\{\psi^{\left(m_F, a_F^S\right)}\left(s^R\right) : \left(m_F, a_F^S\right) \in \Sigma_F(n),\ F \text{ parallel to } E\right\}.$$

Now I show existence of $\psi^{\left(m_F, a_F^S\right)}\left(s^R\right) \in S^R(n)$. Because $\left(m_F, a_F^S\right) \in \Sigma_F(n)$, there exists $s_F^R \in S^R(n)$ free on F that is not a best response to $\left(m_F, a_F^S\right)$. Because replacing s_F^R with $b^R\left(a_F^S\right)$ on F results in another

language-based response, which is a best response to $\left(m_F, a_F^S\right) \in S^S(n-1)$, there exists $\sigma^S_{\left(m_F, a_F^S\right)} \in \Delta S^S(n-1)$ with support on $F \times A^S$ to which no best response among strategies language-based on F is also a best response to $\left(m_F, a_F^S\right)$. Because $j \leq N-3$, $M_\phi(j)$ has at least three component message blocks. Thus we can pick F' parallel to both E and F, and any $\left(m_{F'}, a_{F'}^S\right) \in \Sigma_{F'}(n)$. Define $\sigma^S_{\left(m_{F'}, a_{F'}^S\right)}$ analogously. Define

$$\sigma_{F'}^S = \begin{cases} \left(m_{F'}, a_{F'}^S\right) & \text{if } b^R\left(a_{F'}^S\right) \neq b^R\left(a_E^S\right) \\ \sigma^S_{\left(m_{F'}, a_{F'}^S\right)} & \text{otherwise} \end{cases}.$$

By lemma 3.2, a best response s_*^R to

$$(1-\varepsilon)\left(m_E, a_E^S\right) + \varepsilon(1-\varepsilon)\,\sigma^S_{\left(m_F, a_F^S\right)} + \varepsilon^2(1-\varepsilon)\left(\sigma_{F'}^S\right) \in \Delta S^S(n-1) \quad (4)$$

exists in $S^R(n)$. For $\varepsilon > 0$ sufficiently small, s_*^R must be a best response to $\left(m_E, a_E^S\right)$; it must also be optimal against $\sigma^S_{\left(m_F, a_F^S\right)}$ among best responses to $\left(m_E, a_E^S\right)$, and optimal against $\sigma_{F'}^S$ among strategies with the aforementioned properties. Thus, $s_*^R(m_E) = b^R\left(a_E^S\right)$, which belongs to $A_{\phi,j}$ by lemma A.2. Because $m_E \in E \subset M_\phi(j)$, s_*^R must map $M_\phi(j)$ into $A_{\phi,j}$ by the language assumption. So s_*^R is language-based on F and F'. Then $s_*^R(m_F) \neq b^R\left(a_F^S\right)$ and $s_*^R(m_{F'}) \neq s_*^R(m_E)$ by the construction of $\sigma^S_{\left(m_F, a_F^S\right)}$ and $\sigma_{F'}^S$. Thus s_*^R is non-constant on $M_\phi(j)$.

I now show that $T\left(s_*^R, s^R; M_\phi(j)\right) \in S^R(n)$. Suppose to the contrary, then $T\left(s_*^R, s^R; M_\phi(j)\right)$ is weakly dominated w.r.t. $S^S(k)$ by some $\tilde{\sigma}^R \in \Delta S^R(k)$ for some $k \leq n-1$. Then $\tilde{\sigma}^R$ must be a best response to (4). Thus $\tilde{\sigma}^R$ contains only s^R non-constant on $M_\phi(j)$ because every best response to (4) must be non-constant on $M_\phi(j)$, thus free and language-based on $M_\phi(j)$. Define $T\left(s_*^R, \tilde{\sigma}^R; M_\phi(j)\right)$ to put probability $\tilde{\sigma}^R\left(s^{R\prime}\right)$ on strategy $T\left(s_*^R, s^{R\prime}; M_\phi(j)\right)$ for all $s^{R\prime} \in S^R$. Define $T\left(\tilde{\sigma}^R, s^R; M_\phi(j)\right)$ analogously. Then $T\left(s_*^R, \tilde{\sigma}^R; M_\phi(j)\right)$ and $T\left(\tilde{\sigma}^R, s^R; M_\phi(j)\right)$ both belong to $\Delta S^R(0)$. By the definition of weak dominance, $\tilde{\sigma}^R$ weakly dominates $T\left(s_*^R, s^R; M_\phi(j)\right)$ either w.r.t. $S^S(k) \cap M_\phi(j) \times A^S$, or w.r.t. $S^S(k) \cap (M \backslash M_\phi(j)) \times A^S$. If

the former is true, $\tilde{\sigma}^R$ weakly dominates s_*^R w.r.t. $S^S(k) \cap M_\phi(j) \times A^S$ and thus $T\left(\tilde{\sigma}^R, s_*^R; M_\phi(j)\right)$ weakly dominates s_*^R w.r.t. $S^S(k)$, contradiction to the construction that $s_*^R \in S^R(n) \subset S^R(k+1)$. A contradiction is obtained analogously if the latter is true. Define $\psi^{\left(m_F, a_F^S\right)}\left(s^R\right) =: T\left(s_*^R, s^R; M_\phi(j)\right)$. Then it belongs to $S^R(n)$ and has the desired properties in (3).

I now prove statement 2. By replacing $\sigma^S_{\left(m_F, a_F^S\right)}$ in (4) with $\left(m_i, a_i^S\right)$ for $i = 1, 2$, one can show that there exists $s_1^R, s_2^R \in S^R(n)$ that are equal to each other outside of F whereas $s^R(m_i) = b^R\left(a_i^S\right)$ for $i = 1, 2$. By hypothesis, s_i^R is constant on F at $b^R\left(a_i^S\right)$ for $i = 1, 2$. Because $\left(m_1, a_1^S\right) \in S^S(n+1)$, σ_1^R exists in $\Delta^+ S^R(n)$ to which $\left(m_1, a_1^S\right)$ is a best response. Define $\sigma_*^R \in \Delta^+ S^R(n)$ by moving half of the weight σ_1^R places on s_2^R to s_1^R. Then, for every $m \notin F$,

$$u^S\left(\left(m, a^S\right), \sigma_*^R\right) = u^S\left(\left(m, a^S\right), \sigma_1^R\right) \leq u^S\left(\left(m_1, a_1^S\right), \sigma_1^R\right);$$

for every $m \in F$,

$$\begin{aligned} u^S\left(\left(m, a_2^S\right), \sigma_*^R\right) &< u^S\left(\left(m, a_2^S\right), \sigma_1^R\right) \\ &\leq u^S\left(\left(m_1, a_1^S\right), \sigma_1^R\right) < u^S\left(\left(m_1, a_1^S\right), \sigma_*^R\right). \end{aligned}$$

Thus, a best response to σ_*^R must belong to $F \times A^S \backslash \left\{a_2^S\right\}$. ■

Given any $\sigma^R \in \Delta S^R(0)$, any a^S where $b^R\left(a^S\right) \in A_{\phi,j}$, define

$$\begin{aligned} &\chi\left(a^S; \sigma^R, m_{\phi,j} \cup M_\phi(j)\right) \\ :&= \sum_{s^R \text{ constant on } m_{\phi,j} \cup M_\phi(j)} \sigma^R\left(s^R\right) g^S\left(a^S, s^R(m_{\phi,j})\right) \\ &+ \sum_{s^R \text{ non-constant on } m_{\phi,j} \cup M_\phi(j)} \sigma^R\left(s^R\right) g^S\left(a^S, b^R\left(a^S\right)\right) \end{aligned}$$

where $\sigma^R\left(s^R\right)$ denotes the probability σ^R places on s^R. Pick $\hat{m}$ in any component message block E of $M_\phi(j)$ and any $\hat{a}^S$ where $b^R\left(\hat{a}^S\right) \in A_{\phi,j}$. If s^R is non-constant on $m_{\phi,j} \cup M_\phi(j)$, then $T\left(Constant - b^R\left(\hat{a}^S\right), s^R, E\right)$ is a language-based response. Derive $\phi\left(\sigma^R\right)$ by moving the weight σ^R puts on s^R

to $T\left(Constant - b^R\left(\hat{a}^S\right), s^R, E\right)$, for every s^R non-constant on $m_{\phi,j} \cup M_\phi(j)$. Then $\chi\left(\hat{a}^S; \sigma^R, m_{\phi,j} \cup M_\phi(j)\right) = u^S\left(\left(\hat{m}, \hat{a}^S\right), \phi\left(\sigma^R\right)\right)$. Note that

$$\chi\left(\hat{a}^S; \sigma^R, m_{\phi,j} \cup M_\phi(j)\right) \geq u^S\left(\left(\hat{m}, \hat{a}^S\right), \sigma^R\right),$$

where equality holds if and only if $\left(\hat{m}, \hat{a}^S\right)$ guarantees coordination conditional on E w.r.t. the support of σ^R. Define

$$B^S_{\phi,j}\left(\sigma^R\right) := \arg \max_{a^S : b^R(a^S) \in A_{\phi,j}} \chi\left(a^S; \sigma^R, m_{\phi,j} \cup M_\phi(j)\right).$$

Lemma A.4 *Consider $m_1 \in m_{\phi,j} \cup M_\phi(j)$. Suppose $\left(m_1, a_1^S\right) \in S^S(n+1)$ is a best response to $\sigma_1^R \in \Delta S^R(n)$, but $s^R(m_1) \neq b^R\left(a_1^S\right)$ for some s^R non-constant on $m_{\phi,j} \cup M_\phi(j)$ in the support of σ_1^R. Then for any component message block F of $M_\phi(j)$ where F does not contain m_1, no Sender strategy in $F \times \left(B^S_{\phi,j}\left(\hat{\sigma}^R\right) \cup \left\{a_1^S\right\}\right)$ guarantees coordination w.r.t. the support of $\hat{\sigma}^R$ conditional on F.*

Proof. Suppose to the contrary that $\left(\hat{m}, \hat{a}^S\right) \in F \times \left(B^S_{\phi,j}\left(\sigma_1^R\right) \cup \left\{a_1^S\right\}\right)$ does, for some component message block F of $M_\phi(j)$ not containing m_1. Then

$$\begin{aligned} & u^S\left(\left(\hat{m}, \hat{a}^S\right), \sigma_1^R\right) \\ = \; & \chi\left(\hat{a}^S; \sigma_1^R, m_{\phi,j} \cup M_\phi(j)\right) \\ \geq \; & \chi\left(a_1^S; \sigma_1^R, m_{\phi,j} \cup M_\phi(j)\right) \\ > \; & u^S\left(\left(m_1, a_1^S\right), \sigma_1^R\right), \end{aligned}$$

because $\hat{a}^S \in B^S_{\phi,j}\left(\sigma_1^R\right) \cup \left\{a_1^S\right\}$. Contradiction. ■

A.2.2 The Main Part of the Proof

Loosely speaking, lemma A.5 shows existence in any given iteration of a potential dominator to a Sender strategy that causes miscoordination.

Lemma A.5 *Consider any $j \leq N-3$. Suppose there exists $\left(m_1, a_1^S\right), \left(m_2, a_2^S\right)$ in $S^S(n+1)$ where $m_1, m_2 \in m_{\phi,j} \cup M_\phi(j)$, $b^R\left(a_1^S\right) \neq b^R\left(a_2^S\right)$ and $X\left(m_2, a_2^S\right) \subset X\left(m_1, a_1^S\right)$. Then the following statements must hold:*

1. *for every component message block* E *of* $M_\phi(j)$ *where* E *does not containing* m_1*, and every* $\hat{\sigma}^R \in \Delta S^R(n-1)$ *to which* $\left(m_1, a_1^S\right)$ *is a best response, there exists* $\left(m_E, a_E^S\right) \in S^S(n)$ *where* $m_E \in E$ *and* $a_E^S \in B_{\phi,j}^S\left(\hat{\sigma}^R\right)$*;*

2. $s_1^R \in S^R(n)$ *non-constant on* $m_{\phi,j} \cup M_\phi(j)$ *exists where* $s_1^R(m_1) \neq b^R\left(a_1^S\right)$*.*

Proof. First note that m_1 does not end with $\left\{b^R\left(a_1^S\right)\right\}$. Otherwise, $m_2 \in X\left(m_2, a_2^S\right) \subset X\left(m_1, a_1^S\right) = \{m_1\}$. Then m_2 must end with $\left\{b^R\left(a_2^S\right)\right\}$ and thus $b^R\left(a_2^S\right) = b^R\left(a_1^S\right)$, contradiction.

Define $S^i(-1) = S^i(0)$. Consider $n = 0$. Pick any $\hat{a}^S \in B_{\phi,j}^S\left(\hat{\sigma}^R\right)$ and any component message block E of $M_\phi(j)$ not containing m_1. We can write $E = M\left(A_{\phi,0}...A_{\phi,j}A_{j+1}\right) \cup M\left(A_{\phi,0}...A_{\phi,j}\left(A_{\phi,j} \backslash A_{j+1}\right)\right)$. We can assume w.l.o.g. that $b^R\left(\hat{a}^S\right) \in A_{j+1}$ because $B_{\phi,j}^S\left(\hat{\sigma}^R\right) \subset A_{\phi,j}$ by definition. Statement 1 holds because E contains message "$A_{\phi,0}...A_{\phi,j}A_{j+1}\left\{b^R\left(\hat{a}^S\right)\right\}$" and every Sender strategy belongs to $S^S(0)$. Statement 2 holds because $Literal \in S^R(0)$ and because message m_1 does not end with $\left\{b^R\left(a_1^S\right)\right\}$.

Suppose the statements are true for $n = k \geq 0$ for all $j \leq N-3$. Consider $n = k+1$. Pick any $\hat{\sigma}^R \in \Delta S^R(k)$ to which $\left(m_1, a_1^S\right) \in S^S(k+2)$ is a best response. By the induction hypothesis, $s_{1,k}^R$ non-constant on $m_{\phi,j} \cup M_\phi(j)$ exists in $S^R(k)$ where $s_{1,k}^R(m_1) \neq b^R\left(a_1^S\right)$.

Suppose $\hat{\sigma}^R$ puts positive probability on $s_{1,k}^R$. For any component message block F of $M_\phi(j)$ that does not contain m_1, we have

$$\begin{aligned} \emptyset &\neq S^S(k) \cap \left(F \times B_{\phi,j}^S\left(\hat{\sigma}^R\right)\right) \text{ (by the induction hypothesis)} \\ &\subset \Sigma_F(k) \text{ (by lemma } A.4). \end{aligned} \tag{5}$$

Pick one such component message block E and any $\left(m_E, a_E^S\right) \in S^S(k) \cap \left(E \times B_{\phi,j}^S\left(\hat{\sigma}^R\right)\right)$. Then, for every $s^R \in S^R(k)$ non-constant on $m_{\phi,j} \cup M_\phi(j)$, $\varphi_{\left(m_E, a_E^S\right)}\left(s^R\right)$ as defined in lemma A.3 exists in $\Delta S^R(k)$. Define $\varphi_{\left(m_E, a_E^S\right)}\left(s^R\right) := s^R$ if s^R is constant on $m_{\phi,j} \cup M_\phi(j)$. Extend the definition

to $\sigma^R \in \Delta S^R(k)$ by putting probability $\sigma^R\left(s^R\right)$ on $\varphi_{\left(m_E,a_E^S\right)}\left(s^R\right)$. It follows that

$$\begin{aligned} u^S\left(\left(m_E,a_E^S\right),\varphi_{\left(m_E,a_E^S\right)}\left(\hat{\sigma}^R\right)\right) &= \chi\left(a_E^S;\hat{\sigma}^R,m_{\phi,j}\cup M_\phi(j)\right) \\ &\geq \chi\left(a_1^S;\hat{\sigma}^R,m_{\phi,j}\cup M_\phi(j)\right) \\ &> u^S\left(\left(m_1,a_1^S\right),\hat{\sigma}^R\right) \end{aligned}$$

where the equality holds by the construction of $\varphi_{\left(m_E,a_E^S\right)}\left(\hat{\sigma}^R\right)$, the first inequality holds because $a_E^S \in B_{\phi,j}^S\left(\hat{\sigma}^R\right)$ and the second inequality holds because $\hat{\sigma}^R$ puts positive probability on $s_{1,k}^R$, which is non-constant on $m_{\phi,j}\cup M_\phi(j)$ with $s_{1,k}^R(m_1)\neq b^R\left(a_1^S\right)$. For $m\notin M_\phi(j)$, we have

$$\begin{aligned} & u^S\left(\left(m_1,a_1^S\right),\hat{\sigma}^R\right) \\ \geq\ & u^S\left(\left(m,a^S\right),\hat{\sigma}^R\right) \\ =\ & u^S\left(\left(m,a^S\right),\varphi_{\left(m_E,a_E^S\right)}\left(\hat{\sigma}^R\right)\right). \end{aligned}$$

For $\left(m,a^S\right)\in S^S(k)\cap M_\phi(j)\times A^S$, we have $b^R\left(a^S\right)\subset A_{\phi,j}$ by lemma A.2 and thus

$$\begin{aligned} & \chi\left(a_E^S;\hat{\sigma}^R,m_{\phi,j}\cup M_\phi(j)\right) \\ \geq\ & \chi\left(a^S;\hat{\sigma}^R,m_{\phi,j}\cup M_\phi(j)\right) \qquad (6) \\ \geq\ & u^S\left(\left(m,a^S\right),\varphi_{\left(m_E,a_E^S\right)}\left(\hat{\sigma}^R\right)\right). \qquad (7) \end{aligned}$$

Moreover, (6) holds strictly if $a^S\notin B_{\phi,j}^S\left(\hat{\sigma}^R\right)$ while (7) holds strictly if $\left(m,a^S\right)\in\left(M_\phi(j)\setminus F\right)\times B_{\phi,j}^S\left(\hat{\sigma}^R\right)$ by construction of $\varphi_{\left(m_E,a_E^S\right)}$. It follows that every best response in $S^S(k)$ to $\varphi_{\left(m_E,a_E^S\right)}\left(\hat{\sigma}^R\right)\in\Delta S^R(k)$ belongs to $E\times B_{\phi,j}^S\left(\hat{\sigma}^R\right)$. Thus, $S^S(k+1)\cap\left(E\times B_{\phi,j}^S\left(\hat{\sigma}^R\right)\right)\neq\emptyset$.

Suppose $\hat{\sigma}^R$ places zero weight on $s_{1,k}^R$. Because $\left(m_1,a_1^S\right)\in S^S(k+2)\subset S^S(k+1)$ and $s_{1,k}^R\in S^R(k)$, for ε sufficiently small, $\left(m_1,a_1^S\right)$ is a best response to $\hat{\sigma}_\varepsilon^R =: (1-\varepsilon)\hat{\sigma}^R+\varepsilon s_{1,k}^R$. Thus, $S^S(k+1)\cap\left(E\times B_{\phi,j}^S\left(\hat{\sigma}_\varepsilon^R\right)\right)\neq\emptyset$ by the previous discussion. Because $B_{\phi,j}^S(.)$ is upper-hemicontinuous and $A_{\phi,j}$ finite, $B_{\phi,j}^S\left(\hat{\sigma}_\varepsilon^R\right)=B_{\phi,j}^S\left(\hat{\sigma}^R\right)$ for all ε sufficiently small. It follows that statement 1 holds for $n=k+1$.

I now show that there exists $s^R_{1,k+1} \in S^R(k+1)$ non-constant on $m_{\phi,j} \cup M_\phi(j)$ that is not a best response to $\left(m_1, a_1^S\right)$.

Case 1 $\left(m_1, m_2 \in M_\phi(j)\right)$ *By lemma A.2,* $X\left(m_i, a_i^S\right) \subset M_\phi(j)$ *for* $i = 1, 2$. *Then* m_1 *and* m_2 *must belong to the same component message block of* $M_\phi(j)$, *denoted by* E. *If there exists* $\left(m_F, a_F^S\right) \in S^S(k)$ *where* $m_F \in M_\phi(j) \setminus E$ *and* $b^R\left(a_F^S\right) \neq b^R\left(a_2^S\right)$, *then some best response to* $\frac{1}{2}\left(m_F, a_F^S\right) + \frac{1}{2}\left(m_2, a_2^S\right)$ *has the desired property. Otherwise, every Sender strategy in* $S^S(k)$ *that sends a message in* $M_\phi(j) \setminus E$ *uses action* a_2^S. *Then Constant* $- b^R\left(a_2^S\right)$ *is a best response to every Sender strategy in* $S^S(k)$ *that uses a message in* $M_\phi(j) \setminus E$. *By lemma 4.1 and because statement 1 holds for* $n = k+1$, *for every message block* F *parallel to* E, *there exists* $m_F \in F$ *such that* $\left(m_F, a_2^S\right) \in S^S(k+1)$ *guarantees coordination conditional on* F *w.r.t.* $S^R(k+1)$. *At least one best response to* $\frac{1}{2}\left(m_1, a_1^S\right) + \frac{1}{2}\left(m_F, a_2^S\right)$ *survive* $S^R(k+2) \subset S^R(k+1)$, *and it must be non-constant on* $M_\phi(j)$ *and must not respond to* m_2 *with* $b^R\left(a_2^S\right)$. *Contradiction to the construction that* $\left(m_2, a_2^S\right) \in S^S(k+2)$ *by lemma A.4.*

Case 2 $\left(m_1 = m_{\phi,j}\right)$ *It suffices to show that there exists* $s^R \in S^R(k+1)$ *non-constant on* $m_{\phi,j} \cup M_\phi(j)$ *because such a strategy must respond to* $m_{\phi,j}$ *with* $\phi(j)$, *and* $\phi(j) \neq b^R\left(a_1^S\right)$ *because* m_1 *does not end with* $\left\{b^R\left(a_1^S\right)\right\}$. *Because statement 1 holds for* $n = k+1$, *we have*

$$S^S(k+1) \cap M_\phi(j) \times B^S_{\phi,j}\left(\sigma^R_{(\hat{m},\hat{a}^S)}\right) \neq \emptyset. \tag{8}$$

If $\left(m_3, a_3^S\right)$ *and* $\left(m_4, a_4^S\right)$ *exist in* $S^S(k+1)$ *where* $m_3, m_4 \in M_\phi(j)$ *and* $b^R\left(a_3^S\right) \neq b^R\left(a_4^S\right)$, *then either* $X\left(m_3, a_3^S\right) \cap X\left(m_4, a_4^S\right) \neq \emptyset$ *and then we are done using arguments for the previous case, or* $X\left(m_3, a_3^S\right) \cap X\left(m_4, a_4^S\right) = \emptyset$ *and then at least one best response to* $\frac{1}{2}\left(m_3, a_3^S\right) + \frac{1}{2}\left(m_4, a_4^S\right)$ *have the desired property. Otherwise, every strategy in* $S^S(k+1)$ *that sends a message in* $M_\phi(j)$ *uses the same action, denoted by* $\hat{a}^S$. *Let* $\left(\hat{m}, \hat{a}^S\right)$ *be one such strategy. (Existence is established by (8).) Suppose to the contrary that every Receiver strategy in* $S^R(k+1)$ *is constant on* $m_{\phi,j} \cup M_\phi(j)$. *Then by statement 2 of lemma A.3, there exists* $\sigma^R_{(\hat{m},\hat{a}^S)} \in \Delta S^R(k+1)$ *to which every best response*

sends a message in the message block $m_{\phi,j} \cup M_\phi(j)$ *and uses an action in* $A^S \backslash \{\hat{a}^S\}$. *Let* $(\tilde{m}, \tilde{a}^S)$ *be a best response to* $\sigma^R_{(\hat{m},\hat{a}^S)}$ *in* $S^S(k+2)$. *Then* $\tilde{m} = m_{\phi,j}$. *Suppose* $b^R(\tilde{a}^S) \in A_{\phi,j}$. *Then* $X(m_2, a_2^S) \subset X(\tilde{m}, \tilde{a}^S)$. *By (8), we have* $\hat{a}^S \in B^S_{\phi,j}\left(\sigma^R_{(\hat{m},\hat{a}^S)}\right)$. *Then*

$$\begin{aligned} & u^S\left((\hat{m}, \hat{a}^S), \sigma^R_{(\hat{m},\hat{a}^S)}\right) \\ = \; & \chi\left(\hat{a}^S, \sigma^R_{(\hat{m},\hat{a}^S)}, m_{\phi,j} \cup M_\phi(j)\right) \\ \geq \; & \chi\left(\tilde{a}^S, \sigma^R_{(\hat{m},\hat{a}^S)}, m_{\phi,j} \cup M_\phi(j)\right) \\ = \; & u^S\left((\tilde{m}, \tilde{a}^S), \sigma^R_{(\hat{m},\hat{a}^S)}\right), \end{aligned}$$

where equalities hold because every s^R *in the support of* $\sigma^R_{(\hat{m},\hat{a}^S)} \in \Delta S^R(k+1)$ *is constant on* $m_{\phi,j} \cup M_\phi(j)$ *by hypothesis. By definition,* $(\tilde{m}, \tilde{a}^S)$ *is a best response to* $\sigma^R_{(\hat{m},\hat{a}^S)}$. *It follows that* $(\hat{m}, \hat{a}^S)$ *is also a best response to* $\sigma^R_{(\hat{m},\hat{a}^S)}$, *contradiction to the construction of* $\sigma^R_{(\hat{m},\hat{a}^S)}$. *So* $b^R(\tilde{a}^S) \notin A_{\phi,j}$ *and thus* $b^R(\tilde{a}^S) = \phi(j)$ *by lemma A.2. But at least one best response to* $\frac{1}{2}(\tilde{m}, \tilde{a}^S) + \frac{1}{2}(\hat{m}, \hat{a}^S)$ *have the desired property, contradiction.*

These are the only two cases. If $m_2 = m_{\phi,j}$ and $m_1 \in M_\phi(j)$, then $X(m_1, a_1^S) \subset M_\phi(j)$ by lemma A.2. Then either $b^R(a_2^S) = \phi(j)$ and thus $X(m_2, a_2^S) = \{m_{\phi,j}\}$ or $b^R(a_2^S) \in A_{\phi,j}$ and thus $X(m_2, a_2^S) = m_{\phi,j} \cup M_\phi(j)$. Both contradict the assumption that $X(m_2, a_2^S) \subset X(m_1, a_1^S)$. ■

Lemma A.6 establishes that after sufficiently many iterations, coordination is guaranteed on the entire message space. Note that we can see $M(A_0)$ as $m_{\phi,0} \cup M_{\phi,0}$.

Lemma A.6 $\forall n = 1, ..., N$, *for every bijective function* $\phi : \{1, .., N\} \to A^R$, *every* $(m, a^S) \in S^S(2n)$ *guarantees coordination conditional on message block* $m_{\phi,N-n} \cup M_\phi(N-n)$ *w.r.t.* $S^R(2n)$.

Proof. By lemma A.2, $(m_{\phi,N}, a^S) \in S^S(1)$ only if $b^R(a^S) = \phi(N)$. Define ϕ' by swapping the last two elements of ϕ. Then $m_{\phi,N-1} = m_{\phi',N}$

and thus $\left(m_{\phi,N-1}, a^S\right) \in S^S(1)$ only if $b^R\left(a^S\right) = \phi(N-1)$. So *Literal* is a best response to every $\left(m, a^S\right) \in S^S(1)$ where $m \in \{m_{\phi,N-1}, m_{\phi,N}\} = m_{\phi,N-1} \cup M_\phi(N-1)$. Then the statement holds for $n = 1$ by lemma 4.1.

I now show that

$$\left(m_{\phi,N-1}, \hat{a}^S\right) \in S^S(2) \tag{9}$$

where $b^R\left(\hat{a}^S\right) = \phi(N-1)$. Define $\hat{s}^R$ as follows. For every decreasing sequence $A_0...A_{k-1}$ where $A_{k-1} \supsetneq \{\phi(N-1)\}$ and "$A_0...A_{k-1}\{\phi(N-1)\}$"$\neq m_{\phi,N-1}$, define $\hat{s}^R$ to be constant at some $a^R \in A_{k-1} \setminus \{\phi(N-1)\}$ on the message block $M(A_0....A_{k-1}\{\phi(N-1)\}) \cup M(A_0...A_{k-1}(A_{k-1} \setminus \{\phi(N-1)\}))$. Define $\hat{s}^R$ to equal *Literal* for messages on which $\hat{s}^R$ is not yet defined. Then $\hat{s}^R$ is language-based and responds with action $\phi(N-1)$ to and only to message $m_{\phi,N-1}$. Then $\left(m_{\phi,N-1}, \hat{a}^S\right)$ is the unique best response to $(1-\varepsilon) * Constant - \phi(N-1) + \varepsilon * \hat{s}^R$, for $\varepsilon > 0$ sufficiently small. (9) then follows by lemma 3.2 and by (1). Similarly, $\left(m_{\phi,N}, \tilde{a}^S\right) \in S^S(1)$ where $b^R\left(\tilde{a}^S\right) = \phi(N)$. Then

$$S^R(2) \text{ contains strategies non-constant on } M_\phi(N-2) = \{m_{\phi,N-1}, m_{\phi,N}\} \tag{10}$$

because a best response to $\frac{1}{2}\left(m_{\phi,N-1}, \hat{a}^S\right) + \frac{1}{2}\left(m_{\phi,N}, \tilde{a}^S\right) \in \Delta S^S(1)$ must be non-constant on $M_\phi(N-2)$. By (9) and (10) and because the statement holds for $n = 1$, $\left(m_{\phi,N-2}, \hat{a}^S\right)$ is weakly dominated by $\left(m_{\phi,N-1}, \hat{a}^S\right)$ w.r.t. $S^R(2)$. Similarly $\left(m_{\phi,N-2}, a^S\right) \notin S^S(3)$ if $b^R\left(a^S\right) = \phi(N)$. Then $\left(m_{\phi,N-2}, a^S\right) \in S^S(3)$ only if $b^R\left(a^S\right) = \phi(N-2)$ by lemma A.2. So *Literal* is a best response to every $\left(m, a^S\right) \in S^S(3)$ where $m \in m_{\phi,N-2} \cup M_\phi(N-2)$. By lemma 4.1, the statement holds for $n = 2$.

Assume that the statement holds for $n = k \geq 2$.

Suppose the statement does not hold for $n = k+1$ for some permutation ϕ. By lemma A.1 and 4.1, there exists $\left(m_1, a_1^S\right), \left(m_2, a_2^S\right) \in S^S(2k+1)$ where $m_1, m_2 \in m_{\phi,N-k-1} \cup M_\phi(N-k-1)$, $X\left(m_1, a_1^S\right) \cap X\left(m_2, a_2^S\right) \neq \emptyset$ and $b^R\left(a_1^S\right) \neq b^R\left(a_2^S\right)$. Because message blocks are either disjoint or ordered by set inclusion, we can assume w.l.o.g. that $X\left(m_1, a_1^S\right) \supset X\left(m_2, a_2^S\right)$. $\left(m_1, a_1^S\right)$ must be a best response to some $\hat{\sigma}^R \in \Delta^+ S^R(2k)$. By lemma A.5,

there exists $\hat{s}^R \in S^R(2k)$ non-constant on $m_{\phi,N-k-1} \cup M_\phi(N-k-1)$ where $\hat{s}^R(m_1) \neq b^R\left(a_1^S\right)$. By the induction hypothesis, $m_1 \notin m_{\phi,N-k} \cup M_\phi(N-k)$, which is a component message block of $M_\phi(N-k-1)$. Thus, by lemma A.5, $\left(\tilde{m}, \tilde{a}^S\right) \in S^S(2k)$ exists where $\tilde{m} \in m_{\phi,N-k} \cup M_\phi(N-k)$ and $\tilde{a}^S \in B^S_{\phi,k+1}\left(\hat{\sigma}^R\right)$. By the induction hypothesis, $\left(\tilde{m}, \tilde{a}^S\right)$ guarantees coordination conditional on $m_{\phi,N-k} \cup M_\phi(N-k)$ w.r.t. $S^R(2k)$. Contradiction by lemma A.4. ■

Now I can finish the proof for Proposition 1. Consider a Sender strategy $\left(\hat{m}, a_*^S\right) \in S^S(k)$ for some $k \geq 0$ where a_*^S is the Sender's Stackelberg action. One Receiver best response to $\left(\hat{m}, a_*^S\right)$ must belong to $S^R(k+1)$. Denote it by s^R_{k+1}. It must respond to $\hat{m}$ with $b^R\left(a_*^S\right)$. By the coordination nature and because a_*^S is the Sender's Stackelberg action, $\left(\hat{m}, a_*^S\right)$ is a best response against s^R_{k+1}. So a Sender strategy using her Stackelberg action a_*^S can not be weakly dominated by one using any other action. Thus, some Sender strategy in $S^S(2N)$ must use her Stackelberg action a_*^S. By lemma A.6, the Sender must obtain her Stackelberg payoff in every strategy profile in $S(2N+1)$.

A.3 Proof for Proposition 2

Proof. Say that $a_1^R \geq a_2^R$ if $g^S\left(a^S, a_1^R\right) \geq g^S\left(a^S, a_2^R\right)$. The definition does not depend on a^S by the assumption on g. Define M^* to be the set of messages $m =$"$A_0...A_n$" (for any integer $n \in [1, N]$) such that, for every $j = 1, .., n$, there exists $\bar{a}_j(m) \in A_j$ and $\underline{a}_j(m) \in A_{j-1} \backslash A_j$ where $\bar{a}_j(m) > \underline{a}_j(m)$. I will show that $S^S(1) = M^* \times A^S$.

I first show that $M^* \times A^S \subset S^S(1)$. Fix $\hat{m} =$"$\hat{A}_0...\hat{A}_n$"$\in M^*$. For $k = 1, ..., n$, iteratively define $s^R_{\hat{m},k}$ to respond with action $\bar{a}_k(\hat{m})$ to messages in $M\left(\hat{A}_0...\hat{A}_{k-1}\hat{A}_k\right)$, to respond with action $\underline{a}_k(\hat{m})$ to the remaining messages in $M\left(\hat{A}_0...\hat{A}_{k-1}\right)$, and to equal $s^R_{\hat{m},k-1}$ on all other messages. Given the conjecture $s^R_{\hat{m},1}$, the Sender prefers messages in $M\left(\hat{A}_0\hat{A}_1\right)$ to those outside. For $k = 1, ..., n-1$, within $M\left(\hat{A}_0...\hat{A}_k\right)$, the conjecture $s^R_{\hat{m},k+1}$ renders messages in $M\left(\hat{A}_0...\hat{A}_k\hat{A}_{k+1}\right)$ preferable to messages outside of $M\left(\hat{A}_0...\hat{A}_{k+1}\right)$.

Define

$$\hat{\sigma}^R_{\hat{m},\varepsilon} = \sum_{j=1}^{n-1} \varepsilon^{j-1}(1-\varepsilon)\, s^R_{\hat{m},j} + \varepsilon^n s^R_{\hat{m},n}.$$

Then for $\varepsilon > 0$ sufficiently small, $\left(\hat{m}, a^S\right)$ is the unique best response to the belief $(1-\varepsilon) * Constant - b^R\left(a^S\right) + \varepsilon * \hat{\sigma}^R_{\hat{m},\varepsilon}$, for any a^S.

Next I show that $S^S(1) \subset M^* \times A^S$. If $\hat{m} = \text{“}\hat{A}_0 ... \hat{A}_n\text{”} \notin M^*$, then $\exists j \in \{1, ..., n\}$ such that $\max \hat{A}_j < \min\left(\hat{A}_{j-1} \backslash \hat{A}_j\right)$. Consider any message m' in $M\left(\hat{A}_1 ... \hat{A}_{j-1}\left(\hat{A}_{j-1} \backslash \hat{A}_j\right)\right)$. If a language-based response s^R responds to $\hat{m}$ and m' with different actions, then $s^R(\hat{m}) \leq \max \hat{A}_j < \min\left(\hat{A}_{j-1} \backslash \hat{A}_j\right) \leq s^R(m')$. Thus, for any a^S, $\left(\hat{m}, a^S\right)$ is weakly dominated by $\left(m', a^S\right)$.

I now show that $S^R(2) = S^R(0)$. Pick any $\hat{s}^R \in S^R(0)$. Every best response to the conjecture that is uniform on

$$\left\{\left(m, b^S\left(\hat{s}^R(m)\right)\right) : m \in M^*\right\} \subset S^R(1)$$

must equal $\hat{s}^R$ on M^* because $b^R = \left(b^S\right)^{-1}$. Because no message outside of M^* is used in $S^S(1)$, all such best responses give the same payoff to the Receiver against any $\left(m, a^S\right) \in S^S(1)$. Thus, $\hat{s}^R \in S^R(2)$. By (1), $S(2) = S(1)$. By induction, $S(\infty) = \left(M^* \times A^S, S^R(0)\right)$. ■

A.4 Proof for Proposition 3

Order the Sender actions so that $\left(a^S_i, b^R\left(a^S_i\right)\right)$ gives the Sender the i^{th} highest payoff among all stage game Nash equilibria. For $i = 1, 2, ...N$, and $n = 1, 2, ...$, define $M^n_i =: \left\{m \in M : \left(m, a^S_i\right) \in S^S(n)\right\}$.

Because g is not weak self-signaling, there exists $j \in \{2, .., N\}$ such that $g^S\left(a^S_j, b^R\left(a^S_j\right)\right) < g^S\left(a^S_j, b^R\left(a^S_i\right)\right)$ for every $i = 1, ..., j-1$.

Lemma A.7 *For every $n = 0, 1, 2, ...$,*

1. *$M^n_j \neq \emptyset$*
2. *$\exists s^R_n \in S^R(n)$ where $s^R_n(m) = b^R\left(a^S_j\right)$ for all $m \in M^{n-1}_j$ and $s^R_n(m) \neq b^R\left(a^S_i\right)$ for any $i = 1, ..., j-1$ and any m,*

3. *for any* $i = 1, ..., j-1$, *any* $m \in M_i^n$, $\exists m' \in M_j^n$ *such that there exists no language-based best response to both* $\left(m, a_i^S\right)$ *and* $\left(m', a_j^S\right)$.

Proof. For $n = 0, 1, ...$, at least one best response to $\frac{1}{\#M_j^n} \sum_{m \in M_j^n} \left(m, a_j^S\right) \in \Delta S^S(n)$ belongs to $S^R(n+1)$ by lemma 3.2. Denote it by s_{n+1}^R. Then $s_{n+1}^R(m) = b^R\left(a_j^S\right)$ for all $m \in M_j^n$.

The statements are true for $n = 0$ because $M_i^0 = M$ for all $i = 1, ..., n$ and because $Constant - b^R\left(a_j^S\right) \in S^R(0)$.

Suppose the states are true for $n = k$. Suppose

$$M_i' =: \left\{m : s_{k+1}^R(m) = b^R\left(a_i^S\right)\right\} \neq \emptyset$$

for some $i = 1, ...j-1$. Let q be the smallest such i. A Sender best response to $s_{k+1}^R \in S^R(k+1)$ exists in $S^S(k)$, which must be $\left(m', a_q^S\right)$ for some $m' \in M_q'$. Then the language-based response s_{k+1}^R is a best response to $\left(m', a_q^S\right)$ and to every strategy in $M_j^k \times \left\{a_j^S\right\}$, contradiction to the assumption that statement 3 holds for $n = k$. Thus $M_i' = \emptyset$ for all $i = 1, ..., j-1$. This established statement 2. Statement 1 follows because every best response to s_{k+1}^R must take action a_j^S.

Suppose to the contrary of statement 3 that for some $i = 1, ..., j-1$, a language-based response exists that is a best response to some $\left(\hat{m}, a_i^S\right) \in S^S(k+1)$ and to every strategy in $M_j^{k+1} \times \left\{a_j^S\right\}$. Then a best response $s_{d,k+1}^R \in S^R(k+1)$ to

$$\frac{1}{\#M_j^{k+1}+1}\left(\left(\hat{m}, a_i^S\right) + \sum_{m \in M_j^{k+1}} \left(m, a_j^S\right) \in \Delta S^S(k)\right) \tag{11}$$

exists by lemma 3.2, which responds to $\hat{m}$ with $b^R\left(a_i^S\right)$ and to every $m \in M_j^{k+1}$ with $b^R\left(a_j^S\right)$. For $\varepsilon > 0$ sufficiently small, a best response to $(1-\varepsilon)\, s_k^R + \varepsilon s_{d,k+1}^R \in \Delta S^S(k)$ must be optimal against $s_{d,k+1}^R$ among all best responses to s_k^R, and thus must use action a_j^S and a message in $\left(s_k^R\right)^{-1}\left(b^R\left(a_j^S\right)\right)$. By lemma 3.2, at least one best response belongs to $S^S(k+1)$. Denote it by $\left(m^*, a_j^S\right)$. Because statement 3 holds for $n = k$ by hypothesis, we have

$s_k^R(\hat{m}) = b^R(a_j^S)$. By the construction of $s_{d,k+1}^R$, $u^S\left(\left(\hat{m}, a_j^S\right), s_{d,k+1}^R\right) = g^S\left(a_j^S, b^R\left(a_i^S\right)\right) > g^S\left(a_j^S, b^R\left(a_j^S\right)\right) = u^S\left(\left(m, a_j^S\right), s_{d,k+1}^R\right)$ for any $m \in M_j^{k+1}$. Thus $m^* \notin M_j^{k+1}$, contradiction to the definition of M_j^{k+1}. ■

Now we can prove Proposition 3. By finiteness of the strategy space, $S(\infty) = S(K)$ for some K. Because a strategy using the Stackelberg action a_1^S cannot be weakly dominated by one that uses any $a^S \neq a_1^S$, $\exists m^* \in M_1^K$. The proposition follows immediately from statement 2 of lemma A.7 because miscoordination happens at $\left(\left(m^*, a_1^S\right), s_K^R\right) \in S^R(\infty)$.

References

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

[2] Baliga, S., Morris, S.: Co-ordination, Spillovers, and Cheap Talk. *J Econ. Theory* **105**, 450-468 (2002)

[3] Brandenburger, A., Friedenberg, A., Keisler, H.J.: Admissibility in Games. *Econometrica* **76**, 307-352 (2008)

[4] Blume, A.: Communication, Risk, and Efficiency in Games. *Games and Economic Behavior* **22**, 171–202 (1998)

[5] Demichelis, S., Weibull, J.: Language, Meaning and Games: A Model of Communication, Coordination and Evolution. *American Economic Review* **98**, 1292-1311 (2008)

[6] Ellingsen, T., Östling, R.: When Does Communication Improve Coordination. *American Economic Review* **100**, 1695-1724 (2010)

[7] Farrell, J.: Communication, Coordination and Nash Equilibrium. *Econ. Lett.* ***27***, 209-214 *(*1988)

[8] Farell, J.: Meaning and Credibility in Cheap-Talk Games. *Games and Economic Behavior* **5**, 514-531 (1993)

[9] Heller, Y.: Language, Meaning, and Games: A Model of Communication, Coordination, and Evolution: Comment. *American Economic Review* **104**, 1857-1863 (2014)

[10] Hurkens, S.: Multi-sided Pre-play Communication by Burning Money. *Journal of Economic Theory* **69**, 186-197 (1996)

[11] Kim, Y.-G., Sobel, J.: An Evolutionary Approach to Pre-Play Communication. Econometrica **63**, 1181-1193 (1995)

[12] Rabin, M.: Communication between Rational Agents. *Journal of Economic Theory* **51**, 144-170 (1990)

[13] Sobel, J.: A Note on Pre-play Communication. *Games and Economic Behavior* **102**, 477-486 (2017)