Common Knowledge of Language and Iterative Admissibility in Cheap Talk Games

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Pei-yu Lo

Dissertation Director: Professor Stephen Morris

December 2006

Contents

List of Tables

List of Figures

ABSTRACT

Common Knowledge of Language and Iterative Admissibility in Cheap Talk Games

Pei-yu Lo

2006

This dissertation investigates the implications of common knowledge of language on cheap talk games. A general framework is proposed where language is modeled as a direct restriction on players' strategies, and the predictions under iterative admissibility (IA) are characterized.

In the first two chapters, we apply this framework to sender-receiver games a la Crawford and Sobel (1982), where the Receiver takes a one-dimensional action. We incorporate two observations about natural language into the game: 1) literal meaning — there always exists a natural expression to induce a certain action, if that action is indeed inducible by some message, 2) convexity — messages that are more different from each other induce actions that are weakly more different. It is assumed to be common knowledge that the Receiver plays only language-based strategies. Typically, there is a severe multiplicity issue in CS games. This procedure, however, eliminates outcomes where only a small amount of information is transmitted. Under certain regularity conditions, all equilibrium outcomes are eliminated except the most informative one. However, with an example, we point out that the normal form procedure does not take care of sequential rationality. To address this issue, we propose an extensive form procedure and characterize the solution in the second

chapter.

In the third chapter we apply this framework to coordination games with complete information to formalize the debate over the criterion that guarantees coordination. We define a similarity relation between messages in this class of games and then apply the literal meaning and convexity conditions described earlier. We show that self-committing alone is not sufficient to guarantee coordinated play, while the self-committing condition combined with the self-signaling condition are sufficient.

Acknowledgments

I am indebted to Stephen Morris, Dino Gerardi and Benjamin Polak for their invaluable guidance and support. I am also grateful to Dirk Bergemann and Itzhak Gilboa for useful suggestions. I would like to thank seminar participants at Yale University and Brown University. Finally I would like to thank Daniel Monte, Rebecca Sawyer, Amalavoyal Chari, Greg Regan, Christopher Ksoll, Ulrich Wagner, Siddharth Sharma, Dmitry Shapiro for their help and support throughout the process.

Chapter 1

Introduction

Common sense suggests that speaking the same language helps with cooperation and efficiency, as long as there is room for cooperation. However, this phenomenon is not quite captured in economic analyses of communication. Game theoretic predictions do not depend on whether or not the players speak the same language. This is not surprising, since the notion of language is absent from standard models of cheap talk games. In the standard cheap talk analysis, all messages are treated symmetrically, in that the exact labeling does not matter. That is, two messages can have their names swapped with each other without changing the strategy set or the equilibrium outcome. However, if players speak the same language, convention offers a way of interpreting messages, suggesting that labeling does indeed matter. For example, suppose a man and a woman are both native English speakers and they have to choose between going to the opera or to a boxing match simultaneously. Suppose before they leave for the venue, only the woman can leave the man a voicemail. It is natural that two messages, "Opera" and "Boxing", are either taken literally, or ignored for strategic reasons. It is counter-intuitive that the message "Opera" would indicate going

to the boxing match while the message "Boxing" would indicate going to the opera.

Language manifests itself in the asymmetry among messages. This paper attempts to formalize the notion of language in terms of players' strategy sets. We propose the following general framework to incorporate language. First, we model language as a direct restriction on players' strategies. The restriction does not by itself shrink the set of communication outcomes. It eliminates only strategies that are replicas of other strategies up to the name change. We call this new game "the language game." Second, we characterize the predictions of the game with language under iterative admissibility, i.e., iterative deletion of weakly dominated strategies. Applying the language assumption alone or iterative admissibility alone does not shrink the set of outcomes, but the combination can give a sharp prediction.

The first two chapters of this dissertation apply the aforementioned framework to Sender-Receiver games first modeled by Crawford and Sobel (1982) and gives conditions under which this framework guarantees information transmission. Chapter three applies this framework to cheap talk games about intended action.

Chapter 2

Sender-Receiver Game — Normal Form

2.1 Introduction

This chapter applies the language framework to a classic sender-receiver game as in Crawford and Sobel (1982)(CS). The simple structure of CS games provides a straightforward implication for the language assumption, which we will describe below. In a game, the Sender (she) is the only player with private information, which is called the Sender's type, and is assumed to be one-dimensional. The Receiver (he), upon receiving the message, takes a one-dimensional action, which affects the utility of both. The Sender always prefers a different action from the Receiver. Since the Sender communicates in an attempt to influence the behavior of the Receiver, messages can be mapped to recommendations. Equating the message space with the action space allows us to linearly order messages because the action space is on the real line. Two observations of natural language usage are imposed as assumptions: (i) there always exists a natural expression to induce a certain action, if that

action is indeed inducible by some message; (ii) messages that are more different from each other induce actions that are weakly more different, i.e., if two messages induce the same action, any message in between the two will induce the same action. The second assumption exploits the linear order on the action space. It gives more structure to language and is important for our characterizations.

We first take the normal form approach to this multi-stage game described above. It seems natural as language is a normal form restriction, and sequential rationality is not an issue in standard cheap talk games, since all messages can get used with positive probability. We find that if the players' interests are sufficiently aligned, this procedure eliminates outcomes where only a small amount of information is transmitted. Under certain regularity conditions, all equilibrium outcomes, except the most informative one, are eliminated.

However, we find that the normal form approach might eliminate the most informative equilibrium of the game without language. We show an example where our procedure yields a unique outcome where some types receive different actions, in contrast with the original game where babbling is the unique equilibrium and thus the most informative equilibrium in this game. This example illustrates how normal form procedure might allow the Receiver to take a sub-optimal action after receiving some messages, though it requires strategies to be ex ante optimal for the Receiver with respect to his belief. This is because modeling language as a direct restriction on the strategy sets gives language the highest priority, overriding rationality at times. We then illustrate the tension between language, iterative deletion of weakly dominated strategies and sequential rationality.

Our approach falls into the tradition of trying to incorporate literal meanings into cheap

talk games. Farrell (1993), Rabin (1990) and Zapater (1997) share the assumption that the literal meaning of a message is believed if it is credible, but they propose different credibility criteria. Farrell (1993) uses the concept of credible literal meaning to restrict off-equilibrium-path beliefs held by the Receiver and proposes neologism-proofness as an approach to equilibrium refinement. However, this suffers from the Stiglitz Critique, because in establishing credibility, the Sender is assumed to be guaranteed her equilibrium payoff, even if the equilibrium in question is not stable. Additionally, it might result in an empty prediction. In particular, no equilibrium in a nontrivial CS game is neologism-proof.

Rabin (1990) and Zapater (1997) both use rationalizability to establish credibility. To begin the unraveling in rationalizability, they restrict the Sender's strategies and ask whether that restriction is consistent with rationality and common knowledge of the restriction. In making the restriction, certain communication outcomes are ruled out a priori. Credibility assures that if it is common knowledge that this information will be transmitted, the eliminated outcomes will not be realized. Rabin's "credible message rationalizability" represents the minimal amount of information the Sender can credibly transmit, while a "credible proposal," as defined by Zapater, represents the maximum amount. Credible message rationalizability always yields a non-empty (if sometimes weak) prediction, while a credible proposal is not guaranteed to exist. In particular, every CS equilibrium is credible message rationalizable.

Our approach is closely related to Rabin's and Zapater's, since in a two-player setting, rationalizability is equivalent to iterative deletion of strictly dominated strategies. Our approach differs from the literature in two key aspects. First, we make restrictions on the

Receiver's strategy set instead of on the Sender's strategy set, and hence avoid ruling out babbling or any equilibrium outcome a priori. Our definition of language applies without modification to the entire class of CS games, in contrast to Rabin and Zapater's definitions, which are not independent of the specifics of the game, such as the utility functions and the prior, since restrictions have to be credible and credibility differs with games. Second, looking at any message in isolation, we make no assumption about the actions the Receiver will take. Instead, all our assumptions concern the relation between messages in terms of the induced actions. On the other hand, Rabin and Zapater assume that the Receiver believes credible messages and carries out credible recommendations, while the relation between messages is roughly determined by the model. We argue that, in reality, messages have relative meanings in addition to absolute meanings. For example, when the audience says "good job," they might sincerely mean that they appreciate the performance, but they might just be polite. However, for the receiver of the comment, it is probably weakly better than if they say "horrible." We share with the literature the view on absolute meanings, but stress the asymmetry among messages as an important implication of language. In addition to equilibrium selection, our prediction then reflects the effect of the properties of language.

Rabin (1990) and Zapater (1997) both use rationalizability to establish credibility. To get the unraveling going in rationalizability, they restrict the Sender's strategies and ask whether that restriction is consistent with rationality and common knowledge of the restriction. In making the restriction, certain communication outcomes are ruled out a priori. Credibility assures that if it is common knowledge that this information is go-

ing to be transmitted, the eliminated outcomes will not be realized. Rabin's "credible message rationalizability" represents the minimal amount of information the Sender can credibly transmit, while a "credible proposal" as defined by Zapater represents the maximal amount. Credible message rationalizability always yields a non-empty (if sometimes weak) prediction, while a credible proposal is not guaranteed to exist. In particular, every CS equilibrium is credible message rationalizable.

Our approach is closely related to Rabin's and Zapater's, since in a two-player setting, rationalizability is equivalent to iterative deletion of dominated strategies. Our approach differs from the literature in two key aspects. First, we make restrictions on the Receiver's strategy set instead of on the sender's strategy set, and hence avoid ruling out babbling or any equilibrium outcome a priori. Our definition of language applies without modification to the entire class of CS games, in contrast to Rabin and Zapater's definitions which are not independent of the specifics of the game, such as the utility functions and the prior, since restrictions have to be credible and credibility differs with games. Second, looking at any message in isolation, we make no assumption about the actions the Receiver will take. Instead, all our assumptions concern the relation between messages in terms of the induced actions. On the other hand, Rabin and Zapater assume that the Receiver believes credible messages and carries out credible recommendations, while the relation between messages is roughly determined by the model. We argue that, in reality, messages have relative meanings in addition to absolute meanings. For example, when the audience says "good job," they might sincerely mean that they appreciate the performance, but they might just be polite. However, for the receiver of the comment, it is probably weakly better than if

they say "horrible." We share with the literature the view on absolute meanings, but stress the asymmetry among messages as an important implication of language. In addition to equilibrium selection, our prediction then reflects the effect of the properties of language.

The rest of the chapter is structured as follows. Section 2.2 provides a simple example to motivate our approach. Section 2.3 discusses the solution concept in use and the language assumptions. Section 2.4 outlines the setup of the game. Section 2.5 presents the results using normal form concept and highlights the conflict with sequential rationality. Section 2.6 concludes.

2.2 Motivating Example

Consider a two-player game with one-sided pre-play communication. Rob the pirate is planning to set sail for the treasure island. He does not know whether it is on the West sea or the East sea. He only knows that with probability $\frac{2}{3}$, the treasure island is on the West sea. The prior is common knowledge. Sally the witch, however, knows where the treasure island is. Rob asks Sally in which direction he should go and commits to giving Sally a commission if he finds the treasure. Their payoff matrix is as in table 2.1. The row indicates whether the treasure island is on the West Sea or East Sea. The column indicates the direction Rob chooses. W stands for west and E stands for east. The number on the left is Sally's payoff and the number on the right is Rob's payoff. The game goes like this: Sally tells Rob which direction to take, either west or east, and Rob chooses one direction and sets sail. If he finds the treasure, he has to give Sally a payoff of 2. If he does not, neither of them loses anything.

		a	
location of treasure		W	E
	West	2,1	0,0
	East	0,0	2,1

Table 2.1: Treasure Hunt Game

	"*west*"	"*east*"
Stubborn W	W	W
Stubborn E	E	E
Literal	W	E
Opposite	E	W

Table 2.2: Receiver Strategies in the Treasure Hunt Game

Given the true location of the treasure island, t, Sally chooses a message $s^S(t)$: either "west" or "east." Her strategy is therefore $s^S = \left(s^S(West), s^S(East)\right)$. A strategy for Rob, denoted by s^R, is a function from the message space M to the set of actions $A = \{W, E\}$. Table 2.2 lists all of Rob's possible strategies. Both the *Stubborn W* and the *Stubborn E* strategies completely ignore Sally's recommendation. *Literal* strategy and *Opposite* strategy are essentially the same strategy up to relabeling. This is because they both react to one message with the action W and the other with the action E.

This game has two equilibrium outcomes. One is the so-called "babbling" equilibrium, in which Rob always chooses W and Sally "babbles". The other equilibrium is what we call the informative equilibrium, in which Rob's decision changes with Sally's recommendation and Sally's recommendation depends nontrivially on the true state. There is an innocuous multiplicity here in terms of relabeling the two messages. Actually, if we relabel the messages, we will end up with the same strategy set. The symmetry between messages suggests that language does not play a role in standard analysis. Game theoretic predictions for an English speaking Rob and an English speaking Sally would be the same as the

predictions for an English speaking Rob and an alien Sally.

However, suppose Rob and Sally do share a common first tongue, say English. In the language English, "west" means the direction where the sun falls and "east" means the direction where the sun rises. It seems absurd that Rob and Sally would coordinate in such a way that the message "west" induces Rob to go east and the message "east" induces Rob to go west, if they are going to play the informative equilibrium. The issue is not credibility: if Rob does not believe that Sally's recommendation conveys information, Rob would ignore the message and take the same action regardless. If Rob's action depends nontrivially on Sally's message, then it seems more natural that he would go west upon hearing the suggestion "west" and he would go east upon hearing the suggestion "east."

Suppose it is common knowledge that Rob follows the convention of language and does not use the *Opposite* strategy. That is to say, in the game with language G_L, the set of strategies for Rob is $S_L^R \equiv \{Stubborn\ W,\ Stubborn\ E,\ Literal\}$. Then when the true state is $West$, for Sally, sending the message "east" is weakly dominated by sending the message "west". To see this, notice that both messages yield the same payoff if Rob plays either the *Stubborn W* strategy or the *Stubborn E* strategy. Sally's choice of message matters only if Rob plays the *Literal* strategy. In that case, message "west" induces the action W, which is strictly preferred by Sally when the true state is $West$. Similarly when the true state is $East$, the message "west" is weakly dominated for Sally. In conclusion, if Sally does not play weakly dominated strategies, then she says "west" when the true state is $West$ and "east" when the true state is $East$.

If the Receiver knows that Sally does not play weakly dominated strategies, then when

	Equilibrium	**IA**	**ID**
No Language	babbling,informative	everything	everything
Language	babbling,informative	informative	everything

Table 2.3: Predictions of Treasure Hunt Game

he receives the recommendation "west", he knows that the true location must be $West$, and when he receives the recommendation "east", he knows that the true location must be $East$. The optimal strategy then is to follow Sally's advice and play the strategy $Literal$. We therefore end up with a unique prediction that Rob and Sally play the informative equilibrium outcome, which is what we would "expect".

Eliminating the *Opposite* strategy by way of the language assumption is a key step in getting the unique prediction. In the game *without* language, both strategies *Literal* and *Opposite* belong to Rob's strategy set. When the true state is $West$ ($East$), sending message "*east*" ("*west*") performs better for Sally than sending message "*west*" ("*east*") if Rob plays the strategy *Opposite*, while sending "*west*" ("*east*") performs better if Rob plays the strategy *Literal*. In short, none of Sally's strategies are weakly dominated. Eliminating the strategy *Opposite* from Rob's strategy set gets the unraveling process going.

However, language alone does not do the trick. It is language combined with iterative deletion of weakly dominated strategies that sharpens the predictions. Table 2.3 summarizes the predictions under different combinations of solution concepts and language assumption.

IA means iterative admissibility, i.e., iterative deletion of weakly dominated strategies. ID stands for iterative deletion of strictly dominated strategies. "Everything" means every pair of strategies in the game except those where Rob plays *Stubborn E*. As long as Rob is rational, he will not play *Stubborn E* because going east blindly is worse ex ante than going

west blindly. Here the language assumption alone does not change the set of equilibrium outcomes. It only eliminates the innocuous equilibrium multiplicity where meanings are reversed. Comparing the prediction using IA and ID suggests that weak dominance is key in getting rid of the babbling outcome. This is not surprising since messages are costless, and therefore Sally does not have a strict preference for any message if she believes that Rob will ignore it.

2.3 General Framework

The example of section 2 suggests modeling language as a direct restriction on players' strategies. Let Γ denote a cheap talk game where a one-shot game is preceded by a communication stage. Let I denote the set of players, and T^i denote the set of types for player i. A strategy for player i, denoted by $s^i \in S^i$, is a mapping from player i's type space T^i to his action plans. Write player i's ex ante expected utility function as $U^i : \left(S^i\right)_{i \in E} \to R$. That is, U^i is a mapping from the set of strategy profiles to the real line. We can represent Γ in the strategic form $G = \left(I, \left(S^i\right)_{i \in I}, \left(U^i\right)_{i \in I}\right)$. Language transforms the game into $G_L = \left(I, \left(S^i_L\right)_{i \in I}, \left(U^i\right)_{i \in I}\right)$, which we call "the game with language". To make predictions about cheap talk games with language, we need to know two things: (1) the implications of "language," that is, which strategies belong to S^i_L for each $i \in I$, and (2) given $\left(S^i_L\right)_{i \in I}$, the solution to the game G_L. This is a clean way to incorporate language since all assumptions about language are embodied in $\left(S^i_L\right)_{i \in I}$. By altering the assumptions, we can understand the implications of specific properties of language. This section first discusses the solution concept employed and then motivates a specific way to

model language.

2.3.1 Solution Concept

The solution concept employed here is iterative admissibility (IA) when the normal form is used and a variation when the extensive form is used. The discussion of the variation for the extensive form analysis is deferred to section 3. Here we recall the definition of iterative admissibility and discuss the choice of this solution concept over others.

The definitions below follow Brandenburger et al (2004).

Definition 2.1. *Fix* $\left(X^j\right)_{j\in I} \subseteq \left(S^j\right)_{j\in I}$. *A strategy* s^i *is weakly dominated with respect to* X^{-i} *if there exists* $\hat{\sigma}^i \in \Delta X^i$ *such that* $U^i\left(\hat{\sigma}^i, s^{-i}\right) \geq U^i\left(s^i, s^{-i}\right)$ *for every* $s^{-i} \in X^{-i}$ *and that* $U^i\left(\hat{\sigma}^i, \hat{s}^{-i}\right) > U^i\left(s^i, \hat{s}^{-i}\right)$ *for some* $\hat{s}^{-i} \in X^{-i}$. *Otherwise, say that* s^i *is admissible with respect to* $\left(X^j\right)_{j\in I}$. *If* s^i *is admissible w.r.t.* $\left(S^j\right)_{j\in I}$, *simply say that* s^i *is admissible.*

Definition 2.2. *Set* $S^i(0) = S^i$ *for* $i \in I$ *and iteratively define*

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

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

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

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

as Lemma 2.1 below. Note that this result does not hold if there are more than two players unless players can play correlated strategies.

Lemma 2.1 (Brandenburger et al (2004)). *A strategy $\hat{s}^R \in X^R$ is admissible with respect to $X^S \times X^R$ if and only if there exists $\hat{\sigma}^S \in \Delta^+ S^S$ such that $U^R\left(\hat{\sigma}^S, \hat{s}^R\right) \geq U^R\left(\hat{\sigma}^S, s^R\right)$ for every $s^R \in X^R$.*

As our analysis of the Treasure Hunt game revealed, weak dominance is crucial to sharpening the prediction. As noted earlier, this is not surprising since messages are costless, and therefore senders are indifferent between messages. It is this indifference that causes severe multiplicity. In the evolutionary approach, it is important that any strategy that is weakly better than the current strategies gets used with strictly positive probability and gets taken into account by opponents. This corresponds to weak dominance in the iterative procedure instead of strong dominance, since a strategy that survives weak dominance is a best response to a belief that puts strictly positive weight on every surviving opponent strategy.

One reason we choose iterative admissibility over other non-equilibrium concepts employing weak dominance is its epistemic foundation. Brandenburger et al (2004) provide a sufficient epistemic condition under which the predicted strategy profiles are characterized by IA. More specifically, they show that if there is rationality and n-th order assumption of rationality, where n is higher than the number of iterations needed to arrive at IA, then players play strategies in IA. However, we are not incorporating language into the epistemic framework provided by Brandenburger et al (2004). We simply take the solution concept as given and apply it directly to the game transformed by our language assumption.

By representing the original game as $G = (I, \left(S^i\right)_{i \in E}, U)$, we implicitly assumed that players make their decisions at the initial node before nature makes her move. This ex ante interpretation implies that each player believes that different types of his opponent hold the same belief about his behavior. Alternatively, we can think of different types as representing different "individuals," chosen to appear by nature, and thus assume that players make their decisions after nature makes her move.[1] This interim interpretation implies that each player believes that different types of his opponent may hold different beliefs about his behavior. Let's rewrite the set of players, I, as $I^m \equiv \cup_{i \in I} T^i$. Every player q in I^m can then be written as $t^i \in T^i$ for some $i \in I$. Let $\tilde{S}^q = S^{t^i}$ be the set of action plans available to type t^i of player i. Define $\tilde{U}^q \equiv U^i \forall q \in I^m$. Then iterative admissibility in the game $(I, \left(S_L^i\right)_{i \in I}, U)$ under the interim interpretation is equivalent to iterative admissibility in the game $(I^m, \left(\tilde{S}_L^q\right)_{q \in I^m}, \tilde{U})$.

In equilibrium concepts, it does not matter whether players make their decisions before or after nature makes her move, because in equilibrium, every type of player i holds the correct belief about the behavior of their opponents, and thus every type of player i holds the same belief. However, the two interpretations make a difference in nonequilibrium solution concepts, since players are not assumed to hold the "correct" belief about the behavior of the opponents. The interim interpretation is more appealing if we think of private information as some hard-wired characteristics of the players. However, the ex ante interpretation is more closely related to the equilibrium concept in that it is as if players of different types hold the same belief about the opponents. Analysis is conducted

[1] See p.226 in [8].

under both interpretations. In general, it is easier to include strategies under the interim interpretation, while it is easier to exclude strategies under the ex ante interpretation.

Lastly, Lemma 2.2 shows that the equivalence between weak dominance and never best response to a totally mixed belief holds under the interim interpretation with only two players. This characterization, instead of weak dominance, is directly used in practice. To simplify the analysis, we assume that there is only one-sided incomplete information. Player S holds private information while player R does not. It should easily generalize to cases with two-sided incomplete information. Under the interim interpretation, each type of S is considered an individual player, so Lemma 2.1 does not directly apply. Let $X^S \equiv \Pi_t X^S(t)$. The proof for equivalence is similar to that in Pearce (1984).

Lemma 2.2. *s^R is weakly dominated w.r.t. $\left(\Pi_t X^S(t)\right) \times X^R$ if and only if there does not exist a $\sigma^S(t) \in \Delta^+ X^S(t)$ for every t such that*

$$s^R \in \arg\max_{s' \in X^R} U^R\left(\left(\sigma^S(t)\right)_{t \in T^S}, s'\right).$$

2.3.2 Incorporating Language

Recall that language here is simply a subset of players' strategies resembling conventional language usage. This paper focuses on sender-receiver games where only the Sender (S) possesses private information and only the Receiver (R) has a non-trivial one-dimensional action space A. (In arbitrary communication games, our notion of language remains valid, although a different set of restrictions may be appropriate.) The relative simplicity of the communication protocol and the complete linear order on the action space A give language more structure and generate the assumptions discussed below.

Before talking about the implications of language for players' strategy sets, we need to discuss the message space. It is assumed throughout the paper that the message space M has the same number of elements as the action space A. With that assumption in the background, we argue that (1) language should restrict only the Receiver's strategy set, (2) the message space M can be identified with the action space A, i.e., $M = A$, and (3) every Receiver strategy in language should satisfy the *literal meaning* condition and the *convexity* condition. Lastly, we discuss the implications of these restrictions.

It is desirable that our definition of language itself does not restrict the set of communication outcomes in a given game, while eliminating the "innocuous" multiplicities in terms of how messages are used. An outcome of a sender-receiver game dictates which action (or probability distribution over actions) each type of sender induces. Messages are only means to implement a possibly nontrivial outcome since they are costless. As pointed out in the example in section 2.2, relabeling of messages might produce the same outcome. Language is a restriction only insofar as these relabelings are concerned.

In particular, our definition of language itself should not rule out the babbling outcome. This would imply that language does not place any restriction on the Sender's strategy set. Notice that in the babbling equilibrium, the Receiver ignores all messages and the Sender sends every message with equal probability[2]. In other words, language should not force the Sender to convey information, nor should it force the Receiver to react differently to different messages. Thus, language includes any Receiver strategy that is a constant on the message space. This implies that when looking at every message in isolation, any action is

[2] We focus on the babbling equilibrium strategy profile where every message is used with strictly positive probability, so that Bayesian update can be performed on every message.

possible. For a given type of Sender, every message belongs to her message space if she can play the strategy that puts equal probability on every message. If we do not take it as a literal assumption that there is one Sender at the initial node before nature decides on the true state, the message sent by a given type should not be physically linked to the message sent by another type. Therefore, the set of pure Sender strategies in language as mappings from the type space to the message space should be the product space of M. We conclude that language does not place any restriction on the Sender's strategies.

To justify the simplification that the message space M is equivalent to the Receiver's action space A, notice that the sender talks in an attempt to induce a certain behavior from the Receiver. Say that an action a can be induced in language S_L^R if there exists a Receiver strategy $s^R \in S_L^R$ and a message $m \in M$ such that $s^R(m) = a$. Every action $a \in A$ can be induced in language since language should contain all constant s^R. If the language is rich enough, there is usually a conventional way to express the literal meaning of a. For example, in the Treasure Hunt game, if Sally can get Rob to go east in some way, she can successfully do so simply by saying "Go east!"

Formally, this implies that for every action $\hat{a}$, there is at least one message $\hat{m}$ that invariably induces $\hat{a}$ whenever the Receiver is going to take $\hat{a}$ after some message. Call such $\hat{m}$ a message with literal "$\hat{a}$"-meaning. Whether the Receiver is going to take action $\hat{a}$ after any message is up to strategic considerations, but there is no ambiguity about the literal meanings of messages. If the language is rich enough, there exists a canonical message for every action, that is, for every action a in A, there exists a message m in M with literal "a" $-$ *meaning*. It is easy to show that a message cannot have different

literal meanings. Given the assumption that the number of messages in M is the same as the number of actions in A, we can label the message with literal "a" $-$ *meaning* by "a." Therefore, we can simply assume that $M = A$, given the assumption that $|M| = |A|$.

We can then compare messages since $M = A$. It is intuitive that "similar" messages should induce "similar" actions. For example, when a friend tells you that restaurant A is "faaabulous" instead of telling you that it is "so-so," it would appear that she means that restaurant A is drastically better than average. If you know that your friend have a tendency to exaggerate, and you wouldn't go to restaurant A even if she told you it was "faaabulous," then it is unlikely that you would go to restaurant B if she told you it was "good". Messages that lie on the two extremes should convey weakly more information than messages that lie in between them.

The preceding discussion leads us naturally to define language as follows.

Definition 2.3. *A mapping $s^R : M \to A$ is a language-based Receiver strategy, denoted by $s^R \in S_L^R$, if and only if*

1. *(literal meaning) $s^R(\hat{a}) = \hat{a}$ if there exists a message $\hat{m} \in M$ such that $s^R(\hat{m}) = \hat{a}$;*
2. *(convexity) If $s^R(m_1) = s^R(m_2)$ where $m_1 < m_2$, then $s^R(m_1) = s^R(m)$ for all m such that $m_1 \leq m \leq m_2$.*

Definition 2.4. *The language game G_L inherits all parameters from the original game G, except that the Receiver's pure strategy space is restricted to S_L^R.*

The following lemma characterizes language-based Receiver strategies. The relative meaning property says that a Receiver strategy consistent with language must be weakly

increasing, that is, a higher message induces a weakly higher action. It follows that, whenever a Receiver strategy consistent with language responds to two different messages with different actions, the action taken after receiving the high message is strictly higher. Moreover, the absolute meaning property gives a lower bound to the high action and a higher bound to the low action and says that the higher action has to be higher than the absolute value of the low message, and the lower action has to be lower than the absolute value of the high message.

Lemma 2.3 (Property of strategies in language). *If s^R belongs to language, then*

1. *(relative meaning) s^R is weakly increasing on M. That is, $\forall m_1 < m_2$, $s^R(m_1) \leq s^R(m_2)$,*
2. *(absolute meaning) if $m_1 < m_2$ and $s^R(\ _1) \neq s^R(m_2)$, then $s^R(m_1) < m_2$ and $s^R(m_2) > m_1$.*

Proof. We first show the relative meaning property. Suppose s^R is consistent with language and $s^R(m_2) < s^R(m_1)$ where $m_1 < m_2$. Write $s^R(m_2) = a_2$ and $s^R(m_1) = a_1$. Suppose $s^R(m_2) < m_2$. By the literal meaning condition, $s^R(a_2) = a_2$ and thus s^R responds to the two different messages, message a_2 and message m_2 with the same action. By the convexity condition, s^R responds to every message in $[a_2, m_2]$ with the same action a_2. Figure 2.1 shows the Receiver strategy s^R as a function from the horizontal axis of recommendation to the vertical axis of action. If $s^R(m_1) \leq m_2$, then in particular, s^R responds to message a_1 with action a_2. But by the literal meaning condition, s^R responds to message a_1 with action $a_1 > a_2$. Contradiction! Otherwise, $a_1 > m_2$. By the literal meaning and the

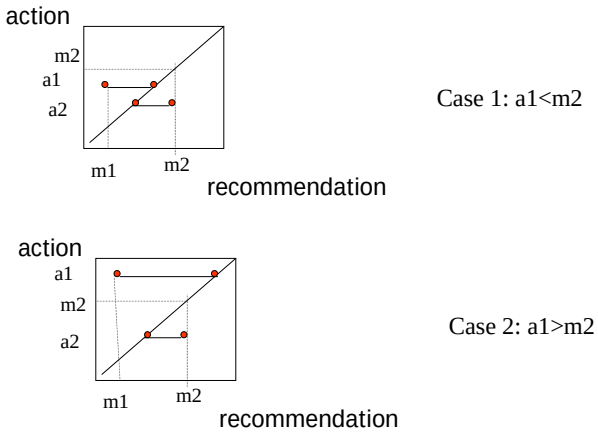


Figure 2.1: Language Property — Relative Meaning

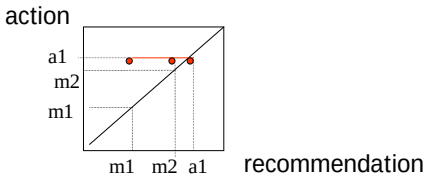


Figure 2.2: Language Property — Absolute Meaning

convexity condition, s^R responds to every message in $[m_1, a_1]$ with the same action a_1. In particular, s^R responds to message m_2 with action a_1. Contradiction! The case where $a_2 \geq m_2$ can be shown analogously.

Now we will prove the absolute meaning property. Suppose s^R is consistent with language and $s^R(m_1) \neq s^R(m_2)$ where $m_1 < m_2$. Suppose to the contrary that $s^R(m_1) = a_1 > m_2$. This is shown in figure 2.2. By the literal meaning condition, s^R responds to message a_1 with action a_1. By the convexity condition, s^R responds to every message in $[m_1, a_1]$ with the same action a_1. In particular, s^R responds to message m_2 with action a_1. So $s^R(m_1) = s^R(m_2)$. Contradiction. We can show analogously that $s^R(m_2) > m_1$. □

The first property reflects the relative difference in messages: a higher message induces a weakly higher action. A deadline of tomorrow signals a more urgent deadline than one 10 days later, if they convey any information at all. The second property reflects the absolute difference in messages: if "excellent" means something different from "good," then "excellent" means something at least as good as the absolute quality of being good.

Given s^R and $Q \subset M$, define

$$s^R(Q) \equiv \left\{a | \exists m \in Q \text{ s.t. } s^R(m) = a\right\}.$$

That is, $s^R(Q)$ is the set of actions induced by a message m in Q under the Receiver strategy s^R.

We began the discussion by arguing that it is desirable that a definition of language does not a priori rule out any outcomes in the original game. Lemma 2.4 stated below confirms that the specific way of modeling language given by definition 2.3 satisfies this condition.

Lemma 2.4 (Completeness of Language). *For all $B \subset A$, there exists a $s^R \in S_L^R$ such that $s^R(M) = B$.*

Proof. To see this, we simply need to construct a Receiver strategy, s^R, taking exactly the actions in a given $B \subset A$. We can linearly order the elements in B and write $B = \{a_1, a_2, ..., a_n\}$ where $a_j < a_{j+1}$ for every j. We can construct the Receiver strategy s^R by defining

$$\hat{s}^R(m) \equiv \begin{cases} a_1 & m \in [0, a_1] \\ a_j & m \in (a_{j-1}, a_j], j = 2, ..., n \\ a_n & m \in [a_n, 1] \end{cases}.$$

It is easy to check that $\hat{s}^R$ satisfies definition 2.3 and $\hat{s}^R(M) = B$. □

Corollary 2.1. *Every equilibrium outcome in the game without language is also an equilibrium outcome in the game with language.*

2.4 The Setup

We apply this general framework to a discretized version of sender-receiver games as in Crawford and Sobel (1982). There are two players, a Sender (S) and a Receiver (R). Only the Sender has private information, represented by her type $t \in T$. The common prior on T is $\pi \in \Delta T$. The Sender sends a message $m \in M$, and the Receiver takes an action a in A after receiving the message m. It is helpful to think of $T = A = M = \{0, \Delta, 2\Delta, ..., 1\}$, though all we need is that they are all finite spaces, and that $A = M$. Both players have Von Neumann-Morgenstern utility function $u^i(t, a)$, $i = S, R$. Though the type space and the action space are both discrete, we assume that u^i can be extended to a function from $[0, 1] \times [0, 1]$ to the real line. It is assumed that u^i is twice continuously differentiable.

As in Crawford and Sobel (1984), it is assumed throughout the paper that $\frac{\partial^2}{\partial a^2}u^i < 0$ and $\frac{\partial^2}{\partial t \partial a}u^i > 0$ for $i = S, R$. Define

$$y^i(t) := \arg\max_{a \in A} u^i(t, a).$$

From the conditions on u^i, $y^i(t)$ is weakly increasing in t for both $i = S, R$. Since A is discretized, $\arg\max_{a \in A} u_i(t, a)$ might not be a singleton. For simplicity, assume that $y^i(t_S)$ is a singleton for all t and both $i = S, R$. The bias is represented by

$$b := \min_{t \in T} \left\{ y^S(t) - y^R(t) \right\}.$$

To simplify the analysis, we also assume that $y^R(t) = t$. Let $E([t_1, t_2])$ denote the optimal action for the Receiver if he only knows that the Sender's type lies in the interval $[t_1, t_2]$. That is, for any $t_1 < t_2$,

$$E([t_1, t_2]) \equiv \arg\max_a \sum_{\substack{t \in T, \\ t_1 \leq t \leq t_2}} u^R(t, a)\, \pi(t).$$

A pure strategy of the Receiver (s^R) is a function from the message space M to the action space A which belongs to the language, that is, $s^R \in S_L^R$. Denote by σ^R a mixed strategy of the Receiver. Under the interim interpretation, a pure strategy of the type t Sender, $s^S(t)$, is an element in the message space M. Write $s^S \equiv \left(s^S(t)\right)_{t \in T}$. Let $\sigma^S(t) \in \Delta M$ denote a mixed strategy of type T Sender. In ex ante interpretation, a pure strategy of the Sender, s^S, is a function from the type space T to the message space M. Denote a pure Sender strategy by s^S and a mixed Sender strategy by σ^S. With some abuse

of notation, write $\left(\sigma^S(t)\right)_{t\in T}$ as σ^S.

For ease of exposition, we restate the related CS results here. In their paper, both the type space and the action space are the unit interval. That is, $T = A = [0,1]$. They showed that every equilibrium is characterized by a finite partition of the type space, $\{t_0, t_1, ..., t_N\}$, where $t_0 = 0$, $t_N = 1$, and type t_i is indifferent between being pooled with the immediately lower step and getting the action $E([t_{i-1}, t_i])$ and being pooled with the immediately higher step and getting the action $E([t_i, t_{i+1}])$. They proved that there exists a finite upper bound $N(b)$ on the maximum number of steps in an equilibrium, and that for every $1 \leq n \leq N(b)$, there exists an equilibrium with n steps.

They used a monotonicity condition to conduct comparative statics. Call a sequence $\tau \equiv \{\tau_0; \tau_1; ...; \tau_N\}$ a forward solution if type τ_i is indifferent between action $E([\tau_{i-1}, \tau_i])$ and action $E([\tau_i, \tau_{i+1}])$ for $i = 1, ..., N-1$. Call N the size of the forward solution τ. Say that τ is a size-N forward solution on $[\tau_0, \tau_N]$ and that $[\underline{\tau}, \bar{\tau}]$ has a forward solution of size N if there exists a forward solution $\{\tau_0; \tau_1; ...; \tau_N\}$ where $\tau_0 = \underline{\tau}$ and $\tau_N = \bar{\tau}$. With abuse of notation, we define

$$t_j^N([[\underline{\tau}, \bar{\tau}]]) \equiv \tau_j,\ j = 1, ..., N-1$$

where $\{\tau_0; \tau_1; ...; \tau_N\}$ is a forward solution on $[\underline{\tau}, \bar{\tau}]$. Write $\alpha_j^N([\underline{\tau}, \bar{\tau}]) \equiv E([\tau_{j-1}, \tau_j])$.

(M) If $\hat{\tau}$ and $\tilde{\tau}$ are two forward solutions with $\hat{\tau}_0 = \tilde{\tau}_0$ and $\hat{\tau}_1 > \tilde{\tau}_1$, then $\hat{\tau}_i > \tilde{\tau}_i$ for all $i \geq 2$.

CS proved that condition (M) implies that ex ante, the Receiver always prefers an equilibrium with more steps. Therefore, the most informative equilibrium, i.e. the equilibrium with the largest number of steps, gives the Receiver the highest ex ante utility. This

condition will play an important role in some of our results.

2.5 Normal Form Iterative Admissibility

Section 2.5.1 characterizes the solution to $NIAL$, which is simply iterative admissibility of the game with language. Section 2.5.2 compares $NIAL$ with equilibria of the game without language and discusses the caveats of $NIAL$.

2.5.1 Characterizations

The notation here implies the use of the interim interpretation. However, the main results hold under both interpretations. Recall that $NIAL$ is simply iterative admissibility in the game with language. By the equivalence of weak dominance and never best response to a totally mixed belief in two player incomplete information games, we rewrite the procedure of $NIAL$ as the following:

Definition 2.5. $S^R(0) = S_L^R$. $S^S(0;t) = M\ \forall t$. *Defined iteratively:*

$$S^R(k+1) = \left\{ \begin{array}{c} s^R \in S^R(k) \mid \\ \text{there exists } \sigma^S(t) \in \Delta^+ S^S(k;t) \text{ for every } t \text{ such that} \\ U^R\left(\left(\sigma^S(t)\right)_{t\in T}, s^R\right) \geq U^R\left(\left(\sigma^S(t)\right)_{t\in T}, s'\right) \text{ for all } s' \in S^R(k) \end{array} \right\}$$

and

$$s^S(k+1;t) = \left\{ \begin{array}{c} m \in S^S(k;t) \mid \\ \text{there exists } \sigma^R \in \Delta^+ S^R(k) \text{ such that} \\ u^S\left(t, \sigma^R(m)\right) \geq u^S\left(t, \sigma^R(m')\right) \text{ for all } m' \in S^S(k;t) \end{array} \right\}$$

where $u^S\left(t, \sigma^R(m)\right) \equiv \sum_{s^R \in S^R} \sigma^R\left(s^R\right) u^S\left(t, s^R(m)\right)$. *Write* $\cap_{k=0}^{\infty} S^R(k) = S^R(\infty)$ *and* $\cap_{k=0}^{\infty} S^S(k;t) = S^S(\infty;t)$. *That is,* $S^R(\infty)$ *and* $S^S(\infty)$ *are the limiting set of strategies*

for the Receiver and the Sender respectively under this normal form iterative procedure.

We need some more notations here.

Notation 1. $l(k;t) \equiv \min S^S(k;t)$;

2. $g(k;t) \equiv \max S^S(k;t)$;

3. $l^{-1}(k;m) \equiv \max\{t | l(k;t) \leq m\}$;

4. $g^{-1}(k;t) \equiv \min\{t | g(k;t) \geq m\}$.

$l(k;t)$ and $g(k;t)$ are respectively the smallest and the largest message that a type t Sender might send in round k. $l^{-1}(k;m)$ is the highest type t that might send a message smaller than or equal to m in round k, while $g^{-1}(k;m)$ represents the lowest type t that might send a message greater than or equal to m in round k. Given k, if $l(k;t)$ and $g(k;t)$ as functions from T to M are bijective when the range is restricted to $l(k;T)$ and $g(k;T)$ respectively, then $l(k;t)$ and $l^{-1}(k;m)$ are inverse functions to each other, while $g(k;t)$ and $g^{-1}(k;m)$ are inverse functions to each other.

Before characterizing the solution to cases where $b > 0$, let's look at the benchmark case where players' interests are aligned, that is, where $y^S(t) = y^R(t)$ for all t. Proposition 2.1 characterizes the $NIAL$ solution. It confirms conventional wisdom that players should be able to coordinate on the efficient outcome if the interests are aligned and they can communicate before playing the game.

We need the following observation for the proof. It says that if a message m is used by some type in round k, and if every message no greater than m can only come from types smaller than or equal to m, then any Receiver strategy that takes a higher-than-recommended action after receiving message m is weakly dominated.

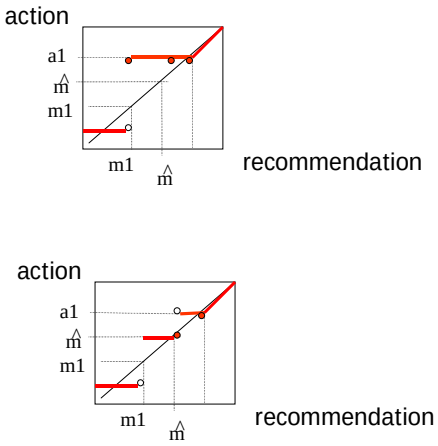


Figure 2.3: Illustration for Observation 2.5.1

Observation If $l^{-1}(k;m) \leq m$, and $m \in M(k)$, then $s^R(m) \leq m$ for all $s^R \in S^R(k+1)$.

Proof. Suppose $\hat{s}^R(\hat{m}) > \hat{m}$ and $l^{-1}(k;\hat{m}) \leq \hat{m}$. Let m_1 be the smallest message m such that $\hat{s}^R(m) = \hat{s}^R(\hat{m})$. Such a Receiver strategy is shown in the upper graph in figure 2.3. From the assumption that $l^{-1}(k;\hat{m}) \leq \hat{m}$, according to any belief on $\Delta S^S(k)$, any message in $[m_1, \hat{m}]$ can only come from types smaller than or equal to $\hat{m}$. Then $\hat{s}^R$ can be improved upon by lowering the value on $[m_1, \hat{m}]$ down to $\hat{m}$. This Receiver strategy is shown in the lower half of figure 2.3. This does not violate the language conditions. So $\hat{s}^R$ is weakly dominated w.r.t. $S(k)$ and does not belong to $S^R(k+1)$. □

Proposition 2.1. *If the Sender and the Receiver's most preferred Receiver actions are the same for every type, that is, $y^S(t^S) = y^R(t^S)$ for all t^S, then there is full communication in $S(\infty)$, i.e. $S(\infty) = \{s^R_{id}\}$ where $s^R_{id}(a) = a$ for all a.*

Proof. We first show that after the first round of deletion of weakly dominated strategies, the

extreme types, type 0 and type 1, sends only recommendations equal to their types. Type 0 prefers a lower action to a higher one. In addition, she prefers action 0 the most. Recall that from the discretization, Δ is the second lowest message. From the relative property of language (see Lemma 2.3 in Section 2.3.2), a lower message always induces a weakly lower action. Therefore, in the first round, every message $m \geq \Delta$ is weakly dominated by message 0 with respect to $S(0)$ for type 0. Similarly, every message $m \leq 1-\Delta$ is weakly dominated by message 1 with respect to $S(0)$ for the highest type Sender. Therefore, after the first round of deletion, type 0 sends only message 0 and type 1 sends only message 1.

Now we'll show that after the second round of deletion of weakly dominated strategies, the Receiver never takes an extreme action after receiving any non-extreme message. After the first round of deletion, the Receiver knows that any message between Δ and $1-\Delta$ can only come from types in $[\Delta, 1-\Delta]$. Suppose to the contrary that there exists $\hat{s}^R \in S^R(2)$ such that $\hat{s}^R(\hat{m}) = 0$ for some $\hat{m} > 0$ and $\hat{s}^R(\hat{m}+\Delta) \neq 0$. From the supermodularity condition of u^R, action Δ is better than action 0 for the Receiver whatever belief he has. If we change $\hat{s}^R$ by changing the action taken on $[\Delta, \hat{m}]$ from 0 to Δ, we strictly improve the Receiver's utility with respect to any belief in $\Delta S^S(1)$. Therefore, such Receiver strategy does not belong to $S^R(2)$. Similarly, any Receiver strategy s^R where $s^R(m) = 1$ for some $m \neq 1$ is weakly dominated w.r.t. $S(1)$ and does not belong to $S^R(2)$. Therefore, given any message in $[\Delta, 1-\Delta]$, the extreme actions the message may induce are Δ and $1-\Delta$, and the extreme types that may send messages in $[\Delta, 1-\Delta]$ are type Δ and type $1-\Delta$. Since on the interval $[\Delta, 1-\Delta]$, type Δ prefers a lower action to a higher one, every message m greater than Δ is weakly dominated for type Δ by message Δ with respect to

$S(2)$ because message Δ induces a weakly lower action in $[\Delta, 1-\Delta]$. The same holds here for type $1-\Delta$. We have thus shown that, restricting attention to messages in $[\Delta, 1-\Delta]$, the extreme types, type Δ and $1-\Delta$, may send only the extreme messages, message Δ and $1-\Delta$.

Likewise for all s^R in $S^R(4)$, s^R responds to every non-extreme message in $[\Delta, 1-\Delta]$, i.e, messages in $[2\Delta, 1-2\Delta]$ with only actions in $[2\Delta, 1-\Delta]$. Repeating the process iteratively, we obtain the result that $s^R(m) = m$ for all $s^R \in S^R(\infty)$ and $s^S(t) = t$ for all type $t \in T$. □

The finiteness assumption imposed on the type space T is crucial to the proof above. In the iterative process, we first showed that the lowest type of the Sender does not send any message other than the lowest one, because she prefers the lowest action (action 0) to a higher action, and message 0 induces a weakly lower action than any other message greater than 0. In response to that, the Receiver does not take the lowest action unless he receives the lowest message (message 0). Therefore, the lowest action a Sender will get by sending a message higher than 0 is the action preferred by the second lowest type of the Sender. Hence, the second lowest type Sender does not send any message higher than her most preferred action, which is equal to her type. However, if T is dense, the second lowest type does not exist. Therefore, this argument does not carry through. Nonetheless, the full communication result itself does not necessarily rely on the finiteness assumption. As a corollary to Proposition 2.4 to be stated later, when the monotonicity condition (M) holds (defined in section 2.4), full communication is the unique outcome in the limiting set even without the finiteness assumption. However, without the finiteness assumption, we

do not know at this stage about convergence.

From now on, it is assumed that $b > 0$. It implies that $b \geq \Delta$ since y^i is defined on the discretized space. The case that $b < 0$ is done in the same way.

NIAL gives a nontrivial upper bound and lower bound on the amount of information transmitted in a given game. Before stating the results, we need to define how we measure the amount of information transmitted. Let Q be a subset of the message space M. Say that an action a is inducible on Q under a Receiver strategy s^R if there exists a message m in Q such that $s^R(m) = a$. For the Receiver to be willing to use a strategy s^R that has many different inducible actions, he has to believe that the Sender can credibly transmit significant amounts of fine-tuned information. Let s^R be in $S^R(\infty)$. Define $M(\infty)$ to be the set of messages used by some type t under some strategy in $S^S(\infty; t)$. Say that a is inducible under s^R if a is inducible on $M(\infty)$ under s^R. When measuring the number of different inducible actions taken by s^R, we confine the attention to the message subset $M(\infty)$ since messages outside of $M(\infty)$ are never used by any type of the Sender, and hence actions taken by the Receiver outside of $M(\infty)$ are irrelevant. Proposition 2.2 stated below implies a limited number of different values for s^R. This is intuitive because as the interests of the Sender and the Receiver diverge, it becomes more difficult to transmit fine-tuned information credibly. Proposition 2.3 stated below says that if given a bias b sufficiently small, the number of inducible actions on $M(\infty)$ under s^R is at least $L \geq 2$, where L varies with the bias. In addition, L does not depend on how finely we discretize the action space, as long as it is not greater than the bias. These two results hold whether the interim interpretation or the ex ante interpretation is used in this incomplete information

game.

Lemma 2.5 is the building block for these two results. It states that no type of the Sender ever recommends an action that is smaller than what is most preferred by the Receiver. We say that two messages are equivalent if they receive the same action under any s^R in $S^R(\infty)$. Lemma 2.5 sets forth, more precisely, that the Sender always gives a recommendation at least as high as an equivalent recommendation of her most preferred action.

Lemma 2.5 (Sender Weak Inflation). *Given any type t, the lowest message she may send in a iterative admissible Sender strategy is at least as high as her type., that is, $l(\infty;t) \geq t$ for all $t \in T$. Moreover, type t Sender either sends a message at least as high as her preferred Receiver action, i.e., $l(\infty;t) \geq y^S(t)$ or the lowest message she sends is equivalent to the recommendation equal to her preferred Receiver action, i.e., $s^R(l(\infty;t)) = s^R\left(y^S(t)\right)$ for all $s^R \in S^R(\infty)$.*

Proof. The main idea is most easily understood through an example. Suppose the Sender prefers the Receiver action which is higher than her type by 0.05, and the players' utility functions are given by quadratic loss functions. That is,

$$\begin{aligned} u^S\left(t^S, a^R\right) &= -\left(t^S + 0.5 - a^R\right)^2 \\ u^R\left(t^S, a^R\right) &= -\left(t^S - a^R\right)^2. \end{aligned}$$

We first show that, after the first round of deletion of weakly dominated strategies, the highest group of Sender types always weakly inflate. This is illustrated by figure 2.4. The horizontal axis represents messages, and the vertical axis represents actions if we are

looking at Receiver strategies, and types if we are looking at Sender strategies. The dot represents the lowest message for type 0.86 surviving the first round of deletion. By the assumption that $b = 0.5 > 0$, every type above 0.95 prefers the highest action, action 1, the most and prefers a higher action to a lower one. By the relative meaning property, a higher message never induces a lower action, for every type above 0.95, message 1, the highest recommendation, weakly dominates every other message. Therefore, every type above 0.95 weakly inflates. Moreover, for every type $t > 0.9$, a message lower than her own type is weakly dominated for her by the message equal to her own type. To see this, let's look at type 0.91. If a message $\tilde{m}$ lower than her own type induces an action different from that induced by the message 0.91, then by the language properties, the action induced by the lower message $\tilde{m}$ is lower than 0.91, while the highest action that the message 0.91 can induce is the highest action 1. Since type 0.91 prefers action 1 to any action lower than 0.91, by concavity, she prefers any action between 0.91 and 1 to any action lower than 0.91. Therefore, the lower message $\tilde{m}$ is weakly dominated by the message 0.91 for type 0.91. We have thus shown that the highest group of Sender types, i.e., types weakly above 0.91, all weakly inflate after the first round of deletion.

Now we will show that, after the second round of deletion of weakly dominated strategies, the Receiver never takes an above-recommendation action after receiving any message weakly above 0.91. After the first round of deletion of weakly dominated strategies, every type above 0.91 weakly inflates. This implies that every message above 0.91 comes only from types below the value of the message. Observation 2.5.1 implies that a Receiver strategy that takes an above-recommendation action after receiving a message above 0.91 is

weakly dominated. Moreover, after the second round of deletion, he Receiver never takes an action above 0.91 after receiving a message below 0.91.

Now we will take a look at the strategies of the medium-high group of Sender types and show that, after the third round of deletion of weakly dominated strateiges, the medium-high group of Sender types all weakly inflate as well. Though every type of the Sender prefers a Receiver action strictly above their type, and the relative meaning property of language says that a higher message induces a weakly higher action, a type t Sender might want to send a recommendation smaller than her most preferred action for fear of being pooled with types that are too high and receiving too high an action. Figure 2.4 shows that the lowest message sent by type 0.86 after the first round of deletion is lower than 0.86. To see this, note that message 0.85 is a strict best response for type 0.86 w.r.t. to the Receiver strategy that takes the recommended action after receiving every message weakly below 0.85 and action 1 after receiving every message above 0.85. It can be easily checked that type 0.86 prefers action 0.85 to action 1, and action 0.85 to every action lower than that. Thus, message 0.85 is a strict best response for type 0.86, and hence is not weakly dominated. However, after the second round of deletion of weakly dominated strategies, the Receiver never takes an action above 0.91 after receiving message 0.86. Since 0.91 is the Receiver action most preferred by type 0.86, she prefers a higher action to a lower one if the higher action is below 0.91, and thus message 0.85 neve does better than message 0.86 for type 0.85. Thus, type 0.86 always weakly inflate after the third round of deletion. Similarly, there exists a medium high group of Sender that always weakly inflate after the third round of deletion. After iteration, we find that every type gives a recommendation at

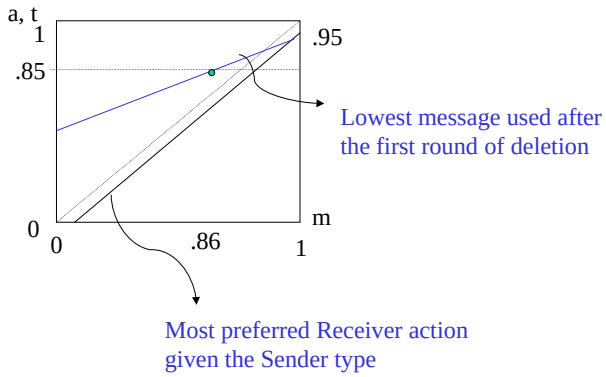


Figure 2.4: Sender Weak Inflation Round 1 Deletion

least as high as her most preferred action. Therefore, it is never optimal for the Receiver to take an action higher than the recommendation.

This sketch is a heuristic proof because an important step is omitted when we say that message 0.85 is weakly dominated by message 0.86 for type 0.85 in the third round — we didn't show that there exists a Receiver strategy surviving the second round of deletion that takes different actions after receiving messages 0.85 and 0.86. The real proof is more complicated because we need to construct such a Receiver strategy. □

We need some notations here. Write $M(k) = \cup_{t \in T} S^S(k;t)$. $M(k)$ is the set of messages used by some type up to round k. Then $M(\infty) = \cup_{t \in T} S^S(\infty;t)$. Define

$$\left(s^R\right)^{-1}(a) \equiv \left\{m \in M | s^R(m) = a\right\}.$$

It is the set of messages in M that induces the action a under the Receiver strategy s^R.

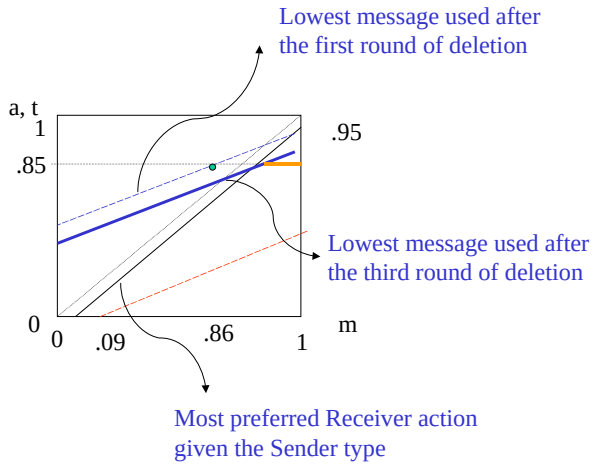


Figure 2.5: Sender Weak Inflation Round 3 Deletion

Now we state the two main results for $NIAL$.

Proposition 2.2 (Coarseness). *Given any* $s^R \in S^R(\infty)$, *suppose* $a_1 < a_2 < a_3$ *are adjacent actions in the range of* s^R, *and* $\left(s^R\right)^{-1}(a_2) \cap M(\infty) \neq \emptyset$, *that is,* a_2 *might be received by some type* t *Sender under some Sender strategy in* $S^S(\infty;t)$. *Then the following inequality holds:*

$$a_3 - a_1 > a_3 - y_S^{-1}(a_3) \geq b,$$

where $y_3^{-1}(a_3)$ *is the type that prefers action* a_3, *or the lowest type that prefers* a_3 *to any action lower than* a_3.

Proof. From the definition of S_L^R, $s^R(a_j) = a_j, j = 1,2,3$. Suppose to the contrary that there exists $\hat{s}^R \in S^R(\infty)$ where $a_1 < a_2 < a_3$ are adjacent actions taken in $\hat{s}^R$ and $a_3 - a_1 \leq b$. Let $[m_2, \bar{m}_2]$ be the interval on which $\hat{s}^R(m) = a_2$. That is, $\left(s^R\right)^{-1}(a_2) = [m_2, \bar{m}_2]$. From Lemma 2.5, types that prefer action a_3 to any lower one will send a message no smaller

than a message which is equivalent to message a_3. (Recall the definition that a message m' is equivalent to message m if $s^R(m) = s^R(m') \forall s^R \in S^R(\infty)$.) $\bar{m}_2$ is not equivalent to message a_3 because $\hat{s}^R(a_3) \neq \hat{s}^R(\bar{m}_2)$ and $\hat{s}^R$ belongs to $S^R(\infty)$. From the definition of b, every type no smaller than $a_3 - b$ prefers action a_3 to any lower one. Therefore, messages in $[m_2, \bar{m}_2]$ can only come from types smaller than $a_3 - b$. By assumption, $a_3 - b \leq a_1$. Then if $[m_2, \bar{m}_2] \cap M(\infty) \neq \emptyset$, $\hat{s}^R$ can be improved upon by changing the action taken on $[m_2, \bar{m}_2]$ from action a_2 to action a_1. Hence $\hat{s}^R$ is weakly dominated and should not belong to $S^R(\infty)$. We need the qualifier that $\left(\hat{s}^R\right)^{-1}(a_2)$ is used by at least one type under some strategy because otherwise the Receiver would not care what he does on the interval $\left(\hat{s}^R\right)^{-1}(a_2)$. □

Remark 2.1. *Proposition 2.2 shows that communication cannot be perfectly informative as long as the bias is greater than* 2Δ. *If the bias is large, we can be sure that very little information will be transmitted. Proposition 2.2 is parallel to Lemma 1 in Crawford and Sobel (1984), stating that there exists* $\varepsilon > 0$ *such that any two actions induced on the equilibrium path differ by at least* ε.

Remark 2.2. *The number of inducible actions under Receiver strategies in* $S^R(\infty)$ *is less than or equal to* $\frac{2}{b}$.

Proposition 2.3. *There exists* $L > 0$ *such that the number of inducible actions on* $M(\infty)$ *under any* $s^R \in S^R(\infty)$ *is at least* L. L *increases as the bias* b *decreases.*

Proof. Lemma 2.5 shows that $l(\infty; t) \geq t$ for all t. From observation 2.5.1, $s^R(m) \leq m$ for all $m \in M(\infty)$. We can also show that the maximum action taken by a strategy $s^R \in S^R(\infty)$ must be greater than or equal to $E([0,1])$. If b is small enough, then

$g(\infty;0) < E([0,1])$, which implies that there will exist some types that will never elicit the highest action because they always send lower messages. Therefore, every s^R in $S^R(\infty)$ must partition $M(\infty)$ into at least 2 subintervals. Let m_q be the lowest message that takes on $\max_m \hat{s}^R(m)$ where $\hat{s}^R \in S^R(\infty)$. Then $E([0,1]) \leq \hat{s}^R(m_q) = m_q$. Let m_{q-1} be the smallest message that takes on $\max_{m<m_q} \hat{s}^R(m)$, then by the same argument, $m_{q-1} = \hat{s}^R(m_{q-1}) \geq E\left(\left[0, g^{-1}(\infty; m_q)\right]\right)$. If b is small, then $g^{-1}\left(\infty; E\left(\left[0, g^{-1}(\infty; m_q)\right]\right)\right) > 0$, that is, there will be types sending only low messages which never elicit an action higher than the second highest one. If we stop after L steps, then we know that every Receiver strategy s^R in $S^R(\infty)$ partitions $M(\infty)$ into at least L intervals. □

CS showed that under the monotonicity condition (M) restated here in section 2.4, the Receiver prefers the most informative equilibrium. They argued that focusing on the most informative equilibrium would be natural. It is natural to ask whether $NIAL$ provides grounds for doing so. To relate $NIAL$ to the equilibrium concept, we will use ex ante interpretation, which is equivalent to assuming that different Sender types hold the same belief about the behavior of the Receiver. Proposition 2.4 states that every equilibrium which is not as informative as the largest equilibrium will be eliminated.

Discretization compels us to make certain assumptions. When $T = [0,1]$, continuity insures that boundary types, which are indifferent between two equilibrium actions, are of measure zero. We assume that this condition holds in the discrete case.

Assumption Given any $\underline{\tau} < \bar{\tau} \in T$, every equilibrium in the game restricted to the subset $[\underline{\tau}, \bar{\tau}] \cap T$ is such that no boundary types are indifferent between two equilibrium actions.

This assumption implies that every forward solution $\{\tau_0; \tau_1; ...; \tau_n\}$ is such that type $\tau_i - \Delta$ prefers action $E([\tau_{i-1}, \tau_i - \Delta])$ to action $E([\tau_i, \tau_{i+1} - \Delta])$, while type τ_i prefers action $E([\tau_i, \tau_{i+1} - \Delta])$ to action $E([\tau_{i-1}, \tau_i - \Delta])$. This assumption will be carried throughout the paper.

Proposition 2.4. *Under condition (M), every Receiver strategy satisfying NIAL takes at least as many different actions on $M(\infty)$ as the most informative equilibrium in the game without language. That is, $L \geq N(b)$.*

We will first show that the babbling outcome is not iteratively admissible in the language game if the original game has an informative equilibrium. Then we will show that, if there is an equilibrium in the original game with three different actions, every iteratively admissible Receiver strategy takes at least three different actions on $M(\infty)$. With induction, we finish the proof for games with bigger-size equilibria.

We need the following lemmas for the proof.

Lemma 2.6. *If there exists a forward solution of size 2 on $[\underline{\tau}, \bar{\tau}]$, and if the monotonicity condition (M) holds, then type $\underline{\tau}$ prefers action $\underline{\tau}$ to action $E([\underline{\tau}, \bar{\tau}])$.*

Proof. Let $\tau_1 = \underline{\tau}$. Since $\tau_1 < t_1^2([\underline{\tau}, \bar{\tau}])$, by the monotonicity condition there exists a forward solution of size two, $\{\underline{\tau}, \tau_1, \tau_2\}$. It has to be the case that $\tau_2 \geq \bar{\tau}$ because type τ_1 prefers action $E([\tau_1, \bar{\tau}])$ to action $E([\underline{\tau}, \tau_1]) = \underline{\tau}$, and by the definition of forward solution, type τ_1 is indifferent between action $E([\tau_1, \tau_2])$ to action $E([\underline{\tau}, \tau_1])$. This contradicts the monotonicity condition because $\tau_1 < t_1^2([\underline{\tau}, \bar{\tau}])$ while $\tau_2 > t_2^2([\underline{\tau}, \bar{\tau}])$. □

Crawford and Sobel showed that if there exists a forward solution of size n on $[\underline{\tau}, \bar{\tau}]$, then

there exists a forward solution of size q on $[\underline{\tau}, \bar{\tau}]$ for every natural number $q \leq n$. Recall that $\{t_0^q([\underline{\tau}, \bar{\tau}]), t_1^q([\underline{\tau}, \bar{\tau}]), ...t_q^q([\underline{\tau}, \bar{\tau}])\}$ denote a forward solution of size q on $[\underline{\tau}, \bar{\tau}]$.

Lemma 2.7. *If there exists a forward solution of size $n+1$ on $[\underline{\tau}, \bar{\tau}]$, and if the monotonicity condition (M) holds, then there exists a forward solution of size n on $[\underline{\tau}, t']$ for every $t' \in \left[t_{n-1}^n([\underline{\tau}, \bar{\tau}]), \bar{\tau}\right]$.*

Proof. Let $\tau_1 = \underline{\tau}$. We can find a forward solution of size n $\{\underline{\tau}, \tau_1, ..., \tau_n\}$. If we can show that $\tau_n \leq t_{n-1}^n([\underline{\tau}, \bar{\tau}])$, then from continuity of the utility functions, for every $t' \geq t_{n-1}^n([\underline{\tau}, \bar{\tau}])$, we can find $\tau_1(t') \geq \underline{\tau}$ such that the size-n forward solution starting with $\tau_1(t')$ ends with t'.

Now we will show that the forward solution $\{\underline{\tau}, \tau_1, ..., \tau_n\}$ must be such that $\tau_n \leq t_{n-1}^n([\underline{\tau}, \bar{\tau}])$. Suppose to the contrary that $\tau_n > t_{n-1}^n([\underline{\tau}, \bar{\tau}])$. By the definition of forward solution, $\{\tau_1, ..., \tau_n\}$ is a size-$(n-1)$ forward solution, and $\left\{\underline{\tau}, t_1^n([\underline{\tau}, \bar{\tau}]), ..., t_{n-1}^n([\underline{\tau}, \bar{\tau}])\right\}$ is also a size-$(n-1)$ forward solution. Since $\tau_1 = \underline{\tau}$ and $\tau_n > t_{n-1}^n([\underline{\tau}, \bar{\tau}])$, the monotonicity condition (M) implies that $\tau_2 > t_1^n([\underline{\tau}, \bar{\tau}])$. We can find τ_{n+1} such that $\{\tau_1, ..., \tau_n, \tau_{n+1}\}$ is a size-n forward induction. Then condition (M) implies that $\tau_{n+1} > \bar{\tau}$. Since $\{\underline{\tau}, \tau_1, ...\tau_n, \tau_{n+1}\}$ and $\left\{\underline{\tau}, t_1^n([\underline{\tau}, \bar{\tau}]), ..., t_{n-1}^n([\underline{\tau}, \bar{\tau}]), \bar{\tau}\right\}$ are both forward solutions of size $(n+1)$. Then $\tau_1 > t_1^n([\underline{\tau}, \bar{\tau}])$ because $\tau_{n+1} > \bar{\tau}$. Contradiction! □

Lemma 2.8. *If there exists a forward solution of size $n+1$ on $[\underline{\tau}, \bar{\tau}]$, and type $\hat{t}$ prefers action $E\left(\left[\hat{t}, \bar{\tau}\right]\right)$ to action $\alpha_n^n\left(\left[\underline{\tau}, \hat{t} - \Delta\right]\right)$, then $\hat{t} \geq t_n^{n+1}([\underline{\tau}, \bar{\tau}])$.*

Proof. Type $\hat{t}$ prefers action $E\left(\left[\hat{t}, \bar{\tau}\right]\right)$ to action $\alpha_n^n\left(\left[\underline{\tau}, \hat{t} - \Delta\right]\right)$, then there exists $\bar{\tau}' > \bar{\tau}$ such that $\left\{\underline{\tau}, t_1^n\left(\left[\underline{\tau}, \hat{t} - \Delta\right]\right), ..., \hat{t}, \bar{\tau}'\right\}$ is a forward solution of size $n+1$. Since $\left\{\underline{\tau}, t_1^{n+1}([\underline{\tau}, \bar{\tau}]), ..., t_n^{n+1}([\underline{\tau}, \bar{\tau}]\right.$

is also a forward solution of size $n+1$, the monotonicity condition (M) implies that $\hat{t} > t_n^{n+1}([\underline{\tau}, \bar{\tau}])$. □

Lemma 2.9. *If $\hat{s}^R \in S^R(\infty)$, a_2 is in the range of $\hat{s}^R$, and a_1 is the largest action below a_2 in the range of $\hat{s}^R$, then the smallest type that may send messages weakly above message a_2, i.e. type $g^{-1}(\infty, a_2)$ must prefer action a_2 to action $E\left(\left[0, g^{-1}(\infty; a_2)\right] - \Delta\right)$.*

Proof. Let $\hat{m}$ be the largest message lower than message a_2 that are iteratively admissible for some type. If $\hat{m} = a_2 - \Delta$, then from observation 2.5.1 and the lemma that every type of the Sender weakly inflates, $\hat{s}^R(a_2 - \Delta) = a_1$, and $s^R(a_2 - \Delta) < a_2$ for every iteratively admissible Receiver strategy s^R which responds to message a_2 and message $a_2 - \Delta$ with different actions. By definition of g^{-1}, every type below $g^{-1}(\infty, a_2)$ sends messages strictly below a_2. Therefore, $s^R \in S^R(\infty)$ and $s^R(a_2) \neq s^R(a_2 - \Delta)$, then $s^R(a_2 - \Delta) \geq E\left(\left[0, g^{-1}(\infty, a_2) - \Delta\right]\right)$ and $s^R(a_2) \geq a_2$. For type $g^{-1}(\infty, a_2)$ to be willing to send messages above a_2, there has to exist $s^R \in S^R(\infty)$ such that type $g^{-1}(\infty, a_2)$ prefers $s^R(a_2)$ to $s^R(a_2 - \Delta)$. It follows that type $g^{-1}(\infty, a_2)$ has to prefer action a_2 to action $E\left(\left[0, g^{-1}(\infty, a_2) - \Delta\right]\right)$.

Now we will discuss the case where $\hat{m} < a_2 - \Delta$. If $s^R(a_2) = a_2$ for every iteratively admissible Receiver strategy s^R that responds differently to message a_2 and message $\hat{m}$, then we are back to the previous case. Suppose to the contrary that type $g^{-1}(\infty, a_2)$ prefers action $E\left(\left[0, g^{-1}(\infty, a_2) - \Delta\right]\right)$ to action a_2, then it has to be the case that she prefers $\tilde{s}^R(a_2) < a_2$ to $\tilde{s}^R(\hat{m}) \neq \tilde{s}^R(a_2)$ for some $\tilde{s}^R \in S^R(\infty)$, and that she prefers action a_1 to action a_2. We will show that $\hat{s}'^R(m)$ which is equal to $\hat{s}^R$ except on messages in $[\hat{m} + \Delta, a_2 - \Delta]$ and takes on action a_1 on messages in $[\hat{m} + \Delta, a_2 - \Delta]$ belongs to $S^R(\infty)$,

and there exists a message in $\left[\hat{m}+\Delta, \tilde{s}^R(a_2)-\Delta\right]$ which is iteratively admissible for type $g^{-1}(\infty, a_2)$. We have thus reached a contradiction and the proof is done.

We know that $\hat{s}'^R$ is consistent with language. Since $\tilde{s}^R \in S^R(\infty)$, there exists a Sender strategy $\hat{\sigma}^S \in \Delta S^S(\infty)$ to which $\hat{s}^R$ is a best response. Since no types of Sender ever sends messages in $\left[\hat{m}+\Delta, \tilde{s}^R(a_2)-\Delta\right]$, $\hat{s}'^R$ is also a best response to $\hat{\sigma}^S$. From our assumption that type $g^{-1}(\infty, a_2)$ prefers action $\tilde{s}^R(a_2)$ to action $\tilde{s}^R(\hat{m})$ and that she prefers action $a_1 = \hat{s}'^R(a_2-\Delta)$ to action $a_2 = \hat{s}'^R(a_2)$, any Sender best response $\hat{s}'^S$ to the Receiver strategy $(1-\varepsilon)\hat{s}'^R + \varepsilon\tilde{s}^R$ must be such that $\hat{s}'^S\left(g^{-1}(\infty, a_2)\right) \in [\hat{m}+\Delta, a_2-\Delta]$. Since the Sender prefers a higher action than does the Receiver, given the type of the Sender, if she prefers the lower action, action a_1, to the higher action, ction a_2, then so does the Receiver. Therefore, any best response of the Receiver to $(1-\varepsilon)\hat{\sigma}^S + \varepsilon\hat{s}'^S$ where ε is sufficiently small must be takes an action lower than a_2 after receiving some message in $[\hat{m}+\Delta, a_2-\Delta]$, and therefore, by the relative meaning property, it must take an action lower than a_2 after receiving message $\hat{m}+\Delta$. So there exists a Receiver strategy $\hat{s}_2^R$ which is equal to $\hat{s}^R$ outside of $[\hat{m}+\Delta, a_2-\Delta]$, and takes an action lower than a_2 after receiving message $\hat{m}+\Delta$. By induction, some message in $[\hat{m}+\Delta, a_2-\Delta]$ is iteratively admissible for type $g^{-1}(\infty, a_2)$, a contradiction! We have thus completed the proof. □

Claim 2.1. *The Babbling outcome is eliminated if there exists an informative equilibrium in the original game.*

Proof. We will first show that there exists a group of low types that never send a message above $E([0,1])$. Recall that $E([0,1])$ is the best action for the Receiver if he knows only the prior. If the Receiver's utility function is a qudratic loss function, $E([0,1])$ is the average

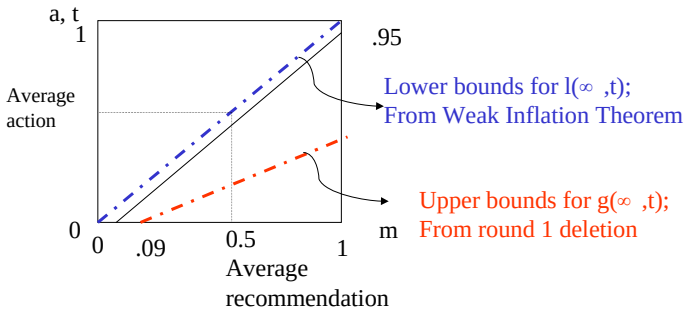


Figure 2.6: Bounds for Iteratively Admissible Messages

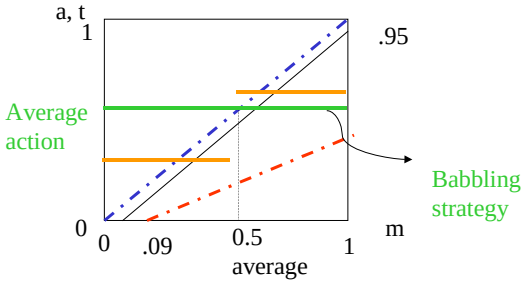


Figure 2.7: Babbling Receiver strategy is Strongly Dominated

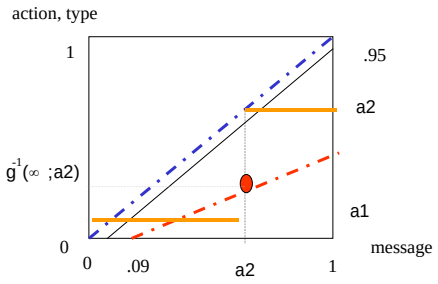


Figure 2.8: A Receiver Strategy with Two Steps

of Sender types. For ease of exposition, we call $E([0,1])$ the average.

By definition, if there exists a size-2 equilibrium in the original game, there exists a size-2 foward solution on $[0,1]$. By lemma 2.6, the lowest type of the Sender, type 0, must prefer action 0 to the average action, action $E([0,1])$. Therefore, the average recommendation, message $E([0,1])$, is weakly dominated in the first round of deletion for type 0 by a slightly below-average recommendation, message $E([0,1]) - \Delta$. To see this, notice that every Receiver strategy consistent with language belongs to $S^R(0)$. Therefore, there are Receiver strategies that takes different actions after receiving message $E([0,1])$ and message $E([0,1]) - \Delta$. By the absolute meaning property, if a Receiver strategy responds to the two messages with different actions, the action taken after receiving message $E([0,1]) - \Delta$ is lower than $E([0,1])$ and the action taken after receiving message $E([0,1])$ is weakly above $E([0,1])$. Type 0 prefers action 0 to the average action $E([0,1])$. By concavity, type 0 Sender prefers any action between 0 and $E([0,1]) - \Delta$ to any action weakly above $E([0,1])$. Therefore, the average recommendation is weakly dominated for type 0. Sim-

ilarly, every above-average recommendation is weakly dominated for type 0. The highest message for type 0 surviving the first round of deletion is lower than average. By continuity of utility functions, types that are close to 0 does not send above-average recommendations either. Therefore, the highest message for the bottom group surviving the iterative process is strictly below average. This is shown by the graph of $g(\infty; t)$ in figure 2.6.

Now we will show that the Receiver strategy that takes the same action after receiving every message in $M(\infty)$ is strongly dominated. Lemma 2.5 shows that every type of the Sender weakly inflates. In particular, every above-average Sender sends an above-average recommendation. We just showed that there exists a bottom group that never sends any message above $E([0,1])$. Therefore, the babbling strategy, the Receiver strategy that responds to every message with action $E([0,1])$, is strongly dominated by a different action that takes a slightly higher action after receiving every message above $E([0,1])$, and a slightly lower action after receiving every message below $E([0,1])$, as illustrated by the Receiver strategy with two steps in figure 2.7. Moreover, the highest action taken after receiving an iteratively admissible message by some Receiver action $s^R \in S^R(\infty)$ must be higher than action $E([0,1])$, the "average" action. □

Claim 2.2. *Receiver strategies with only two different actions in the range are eliminated if there is an equilibrium of size-3 in the original game.*

Proof. Take any iteratively admissible Receiver strategy s^R, write the highest action taken by s^R on $M(\infty)$ as a_2. Figure 2.8 illustrates one such strategy. We now know that $a_2 \geq E([0,1])$. From lemma 2.9, we know that type $g^{-1}(\infty; a_2)$, the smallest type that may send messages weakly above a_2, must prefer action a_2 to pooling with types below

herself and getting action $E\left(\left[0, g^{-1}(\infty; a_2) - \Delta\right]\right)$. Since a_2 is the highest action taken by s^R, and no types smaller than $g^{-1}(\infty, a_2)$ sends messages weakly above message a_2, it has to be the case that

$$a_2 \geq E\left(\left[g^{-1}(\infty; a_2), 1\right]\right).$$

By concavity, type $g^{-1}(\infty, a_2)$ prefers being pooled with types above herself and receiving action $E\left(\left[g^{-1}(\infty, a_2), 1\right]\right)$ to being pooled with types strictly below herself and receiving action $E\left(\left[0, g^{-1}(\infty; a_2) - \Delta\right]\right)$. By lemma 2.8, type $g^{-1}(\infty, a_2)$ must be higher than $t_1^2([0,1])$. From lemma 2.7 and the assumption that there exists a size-3 forward solution on $[0,1]$, there exists a size-2 forward solution on $\left[0, g^{-1}(\infty, a_2) - \Delta\right]$. Then type 0 must prefer action 0 to action $E\left(\left[0, g^{-1}(\infty; a_2) - \Delta\right]\right)$, and thus will not send any message above $E\left(\left[0, g^{-1}(\infty; a_2) - \Delta\right]\right)$ after the first round of deletion. This observation combined with Sender weak inflation shows that if s^R is a constant on messages in $\left[0, g^{-1}(\infty; a_2) - \Delta\right]$, s^R is strongly dominated w.r.t. $S^S(\infty)$. This is the same argument as the one used to show that babbling is eliminated whenever there is a size-2 forward solution on $[0,1]$. Similarly, a_1, the highest action taken on $M(\infty)$ by s^R must be weakly above $\alpha_2^2\left(\left[0, g^{-1}(\infty; a_2) - \Delta\right]\right)$. That is, $s^R(m) \geq \alpha_2^2\left(\left[0, g^{-1}(\infty; a_2) - \Delta\right]\right)$ for every $s^R \in S^R(\infty)$ and $s^R(m)$ is the highest action taken by s^R below a_2. Then for type $g^{-1}(\infty; a_2)$ to be willing to send a message above a_2, it has to be the case that she prefers action $\alpha_2^2\left(\left[0, g^{-1}(\infty; a_2) - \Delta\right]\right)$ to pooling with types above her. Lemma 2.8 shows that type $g^{-1}(\infty; a_2)$ must be weakly above $t_2^3([0,1])$. We have thus shown that the range of any iteratively admissible Receiver strategy must contain at least three different actions in the range w.r.t. to the domain $M(\infty)$, and that the highest action must be weakly above the highest action taken in the

size-three equilibrium. □

For general cases, we employ proof by induction. The proposition follows immediately from the following claim.

Claim 2.3. *For any* $\hat{\alpha} \in \hat{s}^R(M(\infty))$ *where* $\hat{s}^R \in S^R(\infty)$ *and* $\hat{a} \neq \min \hat{s}^R(M(\infty))$, $\hat{s}^R$ *takes at least* q *different actions on* $M(\infty) \cap [0, \hat{a} - \Delta]$ *if* $\left[0, g^{-1}(\infty; \hat{a}) - \Delta\right]$ *has a forward solution of size-*q. *Let* $\tilde{a} \equiv \max \hat{s}^R([0, \hat{a} - \Delta] \cap M(\infty))$. *Then* $\tilde{a}$ *is greater than or equal to the largest action on the size-*q *forward solution on* $\left[0, g^{-1}(\infty; \hat{a}) - \Delta\right]$ *and* $g^{-1}(\infty; a_q) \geq t_{q-1}^q\left(\left[0, g^{-1}(\infty; \hat{a}) - \Delta\right]\right)$ *where* a_q *is the smallest message equivalent to* $\tilde{a}$.

Proof. See the Appendix. □

Remark 2.3. *We prove it by showing that it is necessary for the limiting set. The arguments do not depend on the finiteness assumption. As a corollary, this is also a necessary condition for the limiting set under* $NIAL$ *even if* $T = A = M = [0, 1]$. *In fact, we do not need the assumptions we impose in the discrete case.*

2.5.2 Relating $NIAL$ to Equilibria in the Game without Language

Denote by $EQ(G)$ the set of equilibria in G, where G represents the game without language. Recall that $NIAL$ is iterative admissibility in G_L, the game WITH language. In general, there is no containment between $NIAL$ and $EQ(G)$. As a non-equilibrium concept, $NIAL$ naturally gives rise to non-equilibrium outcome being contained in $NIAL$. Proposition 2.3 implies that $NIAL$ may eliminate some of the less informative equilibria. However, in this section, we present an example demonstrating that $NIAL$ can be disjoint from $EQ(G)$.

types of S sending message m **in** $S^S(1)$	$\emptyset$	0	$\frac{1}{2}, 1$		
s^R **in language** $\setminus$ **message** m	"0"	"$\frac{1}{2}$"	"1"	in $S^R(1)$	in $S^R(2)$
	0	0	0		
	0	$\frac{1}{2}$	$\frac{1}{2}$	v	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	v	
	1	1	1		
	0	1	1	v	
	$\frac{1}{2}$	$\frac{1}{2}$	1	v	
	0	$\frac{1}{2}$	1	v	
s^R_{nice}	0	0	1	v	v

Table 2.4: Receiver Strategy Set in Language

In our example, the unique equilibrium in G is the babbling equilibrium, while the unique prediction given by $NIAL$ is partially informative.

Example 2.4. *There are three types: type* $0, \frac{1}{2}$ *and* 1*. The common prior is such that* $\pi(0) = \frac{1}{3}$*;* $\pi\left(\frac{1}{2}\right) = \frac{4}{9}$ *and* $\pi(1) = \frac{2}{9}$*. Both the Sender and the Receiver have quadratic loss function:* $u^R(t,a) = -(t-a)^2$ *and* $u^S(t,a) = -\left(t + \frac{1}{2} - a\right)^2$*.*

The unique equilibrium in this game without language is babbling. Because both type $\frac{1}{2}$ and type 1 Senders prefer a higher action to a lower one, it is impossible to separate these two types in any equilibrium. To show that there is no informative equilibria, let's suppose to the contrary that there is an equilibrium in which type 0 separates from type $\frac{1}{2}$ and 1. The best action against pooling of type $\frac{1}{2}$ and type 1 is action $\frac{1}{2}$, while the best action against type 0 is action 0. However, this cannot be an equilibrium because type 0 prefers action $\frac{1}{2}$ to action 0, and therefore would have an incentive to imitate type $\frac{1}{2}$ and type 1. Thus, in an equilibrium, type 0 cannot separate from type $\frac{1}{2}$ and type 1. Since the best action with respect to the prior is $\frac{1}{2}$, the unique equilibrium in the game without language is babbling, where all types pool and the Receiver takes action $\frac{1}{2}$ after receiving

every message.

Now let's derive the solution to this game under $NIAL$. Write s^R as a 3-tuple of actions taken after messages $0, \frac{1}{2}$ and 1 respectively, i.e., $s^R = \left(s^R(0), s^R\left(\frac{1}{2}\right), s^R(1)\right)$. The bottom part of table 2.4 shows all the Receiver strategies in S_L^R. In the first round of deletion, $(0,0,0)$ and $(1,1,1)$ are eliminated since the unique best response to pooling of all types is $\frac{1}{2}$. For every other strategy $\hat{s}^R$, a totally mixed Sender strategy $\hat{\sigma}^S$ can be constructed such that $\hat{s}^R$ is a best response to $\hat{\sigma}^S$. Thus no further Receiver strategies can be eliminated in the first round.

Now we must determine the set of Sender strategies that survive the first round. Both type 1 and type $\frac{1}{2}$ Sender prefer a higher action to a lower one. For these two types, both message 0 and message $\frac{1}{2}$ are weakly dominated by message 1 because Receiver strategies in S_L^R are weakly increasing, and hence message 1 induces the weakly highest action. A type 0 Sender prefers action $\frac{1}{2}$ the most, and is indifferent between action 0 and action 1. Recall that by the absolute meaning property of language, s^R satisfies the following inequalities if $s^R(0) \neq s^R\left(\frac{1}{2}\right)$:

$$\begin{cases} s^R\left(\frac{1}{2}\right) > s^R(0) \\ s^R\left(\frac{1}{2}\right) > 0 \\ s^R(0) < \frac{1}{2} \end{cases}.$$

This implies that $s^R(0) = 0$ and $s^R\left(\frac{1}{2}\right) \geq \frac{1}{2}$ if $s^R(0) \neq s^R\left(\frac{1}{2}\right)$. That is, whenever message 0 and message $\frac{1}{2}$ induce different actions, message 0 induces action 0 while message $\frac{1}{2}$ induce either action $\frac{1}{2}$ or action 1. As the type 0 Sender weakly prefer both action $\frac{1}{2}$ and action 1 over action 0, and strictly prefers action $\frac{1}{2}$ over action 0, message 0 is weakly dominated by message $\frac{1}{2}$ for the type 0 Sender. Similarly, message 1 is weakly dominated by message $\frac{1}{2}$

	Equilibrium	**IA**	ID
No Language	s_{babble}	everything	everything
Language	s_{babble}, s_{nice}	s_{nice}	everything

Table 2.5: Comparison of Predictions

for a type 0 Sender among $S^S(0)$. In conclusion, after the first round of deletion, a type 0 Sender will send only message $\frac{1}{2}$, and both type $\frac{1}{2}$ and type 1 Sender will send only message 1. This Sender strategy, called s^S_{nice}, is shown in the first row of table 2.4.

Now we show that $S^R(2) = \left\{s^R_{nice}\right\}$ where $s^R_{nice} = (0,0,1)$. In the second round, the only conjecture the Receiver can hold about the Sender's strategy is s^S_{nice}. Under s^S_{nice}, no type of Sender ever sends message 0. Therefore the Receiver's predetermined response to message 0 is irrelevant. Hence, the relevant difference among Receiver strategies $\left(\frac{1}{2},\frac{1}{2},1\right), \left(0,\frac{1}{2},1\right), (0,1,1)$ and $(0,0,1)$ lies only in their responses at message $\frac{1}{2}$. When receiving message $\frac{1}{2}$, action 0 is the best because only type 0 sends this message. Therefore, Receiver strategy $(0,0,1)$ yields a higher utility than either strategy $\left(\frac{1}{2},\frac{1}{2},1\right), \left(0,\frac{1}{2},1\right)$ or $(0,1,1)$. Then we need only compare s^R_{nice} with the strategy $\left(0,\frac{1}{2},\frac{1}{2}\right)$. Simple calculation of ex ante utility shows that $U^R\left(s^S_{nice},(0,0,1)\right) > U^R\left(s^S_{nice},\left(0,\frac{1}{2},\frac{1}{2}\right)\right)$. So $S^R(2) = \left\{s^R_{nice}\right\}$. The process then stops and $S(\infty) = \left(\left\{s^S_{nice}\right\},\left\{s^R_{nice}\right\}\right)$. Call this strategy profile s_{nice}. $NIAL$ predicts that type 0 Sender receives action 0 and both type $\frac{1}{2}$ and type 1 Sender receive action 1.

Table 2.5 summarizes this game's predictions under different combinations of language restriction and solution concepts. s_{nice} emerges as an equilibrium in the game with language, though babbling is the unique equilibrium in the game without language. In arriving at $S(\infty)$, we previously showed that s^R_{nice} is optimal with respect to s^S_{nice} among S^R_L. s^S_{nice}

types of S sending message m **in** s^S_{nice}	$\emptyset$	0	$\frac{1}{2}, 1$	
s^R \ **message** m	"0"	"$\frac{1}{2}$"	"1"	in language
s^R_{nice}	0	0	1	yes
s^R_{cheat}	0	0	$\frac{1}{2}$	no
s^R_{ignore}	0	$\frac{1}{2}$	$\frac{1}{2}$	yes

Table 2.6:

is optimal among S^S with respect to s^R_{nice} because every other Sender strategy is weakly dominated by s^S_{nice} with respect to S^R_L. It follows that s_{nice} is an equilibrium in G_L.

To understand why s_{nice} is not an equilibrium in G, note that according to s^S_{nice}, message 1 is transmitted by either type $\frac{1}{2}$ or type 1. The best response against pooling of these two types is action $\frac{1}{2}$, not action 1. Therefore, the strategy $\left(0, 0, \frac{1}{2}\right)$ yields a higher utility than $(0, 0, 1)$ with respect to s^S_{nice}. It then follows that s_{nice} is not an equilibrium in G. Table 2.6 illustrates all the relevant strategies. To see why s_{nice} is an equilibrium in G_L but not an equilibrium in G, note that $\left(0, 0, \frac{1}{2}\right)$ does not satisfy the language assumptions, and therefore does not belong to S^R_L. Recall that by the literal meaning assumption of language (see definition 2.3), if action $\frac{1}{2}$ belongs to the range of a strategy, action $\frac{1}{2}$ must be taken in response to message $\frac{1}{2}$. Thus, if the Receiver wants to take action $\frac{1}{2}$ after receiving message 1, he must choose strategy $\left(0, \frac{1}{2}, \frac{1}{2}\right)$. Though strategy $\left(0, \frac{1}{2}, \frac{1}{2}\right)$ yields a higher interim utility than strategy $(0, 0, 1)$ when message 1 is received, it yields a lower interim utility when message $\frac{1}{2}$ is received, because only type 0 sends message $\frac{1}{2}$, and action 0 is the best against type 0. When deriving the solution to $NIAL$, we have shown that strategy $(0, 0, 1)$ gives a higher ex ante payoff than strategy $\left(0, \frac{1}{2}, \frac{1}{2}\right)$ against the Sender strategy s^S_{nice}. Therefore, $(0, 0, 1)$ is optimal among S^R_L with respect to s^S_{nice}, though it is not optimal among the unrestricted strategy set.

This example points out that in a game with language, ex ante utility maximization does not necessarily imply interim utility maximization on every information set on the equilibrium path. We showed above that s^R_{nice} is ex ante optimal against s^S_{nice}. However, s^R_{nice} takes a suboptimal action against the posterior generated by s^S_{nice} in response to message 1, which is reached with positive probability by the profile s_{nice}. Thus, in the game with language, s^R_{nice} is not interim optimal with respect to s^S_{nice} even on the equilibrium path.

The break down of the link between ex ante optimality and interim optimality on the equilibrium path results from the non-separability of the second-stage-action space created by the language restriction. In the first stage of the two-stage sender-receiver game, the Sender decides to send a message based on her private information. In the second stage, the Receiver takes an action in response to the message from the Sender. In a standard game, the action space available to a player at a given information set (a particular message in a sender-receiver game) does not depend on the action the player plans to take at any other information set. We can conceive of the second stage action space as "separable". The language restriction breaks the separability: the set of actions available to the Receiver upon receiving a message depends on which actions he plans to take in response to other messages. Although for any single message taken in isolation, language does not restrict the Receiver to a strict subset of his action space, when holding fixed the Receiver's responses to other messages, language assumption does often impose restrictions on the available action space. In Example 2.4, if the Receiver wants to respond optimally to message 1 with respect to s^S_{nice}, he should take action $\frac{1}{2}$. However, if he takes action $\frac{1}{2}$ after receiving message

1, he must then take action $\frac{1}{2}$ at message $\frac{1}{2}$ by the literal meaning condition of language. Note that our assumption prohibits the Receiver from any strategy which violates language. Therefore, the set of actions available to the Receiver at message $\frac{1}{2}$ is $\left\{\frac{1}{2}\right\}$ given that he takes action $\frac{1}{2}$ after receiving message 1. Thus it is often the case that when the Receiver decides to take an optimal action in response to a message based on a conjecture he holds about the Sender's behavior, he will be forced to take a suboptimal action in response to another message reached with positive probability. The Receiver gauges the gains and losses ex ante and chooses one that maximizes his ex ante utility.

Though the language restriction does not limit strategic contents, in that $EQ(G) \subset EQ(G_L)$ for every sender-receiver game G, it does provide an artificial commitment device that may make $EQ(G)$ strictly contained in $EQ(G_L)$. As is often the case, commitment makes the Receiver weakly better off. For example, s_{nice} gives the Receiver a higher ex ante payoff than the babbling outcome. However, the Receiver does not really have a commitment device. Incorporating language with a normal form approach fails to take into account sequential rationality, since in the game with language, interim optimality on the equilibrium path is no longer implied by ex ante optimality. This prompts us to develop an extensive form version incorporating language, iterative admissibility and sequential rationality.

2.6 Conclusion

This chapter is an exercise to demonstrate the power of incorporating the asymmetry implied by language, when language is regarded as one way to transmit a given amount of

information. Taking a literal approach, we model common knowledge of language by directly restricting players' strategy sets without a priori ruling out any outcome. We then characterize the solution to this new game under iterative admissibility. Applying the general framework to sender-receiver games a la Crawford and Sobel (1982), we assume that strategies satisfy "language" if and only if they satisfy the literal meaning condition and the convexity condition. Using normal form iterative admissibility, under a regularity condition, we show that all outcomes are at least as informative (in terms of number of distinct actions possibly received by the sender) as the most informative equilibrium.

However, we illustrate through an example that this procedure may eliminate even the most informative equilibrium, and we point out the tension among language, iterative admissibility and sequential rationality. These conflicts arise because modeling language through physically restricting a player's strategy set gives language the highest priority. Therefore, language always overrides interim optimality, although normal form iterative admissibility takes care of ex ante optimality. We will make an attempt to address this issue in the following chapter.

2.7 Appendix

2.7.1 Proofs for Section 2.3

Proof for Lemma 2.12.2. We first establish the equivalence between strong dominance and best response. Lemma 2.1 then follows using the same method in the proof of lemma 4 in the Appendix of Pearce (1984). For the completeness of the argument, we restate the proof below.

Suppose that $\hat{s}^R$ is not a best response to any $\sigma^S \in \Pi_t\left(\Delta^+ X^S(t)\right)$. Define

$$A = \left\{\sigma^R \in \Delta X^R : U^R\left(\sigma^S, \sigma^R\right) = U^R\left(\sigma^S, \hat{s}^R\right) \forall \sigma^S \in \Pi_t\left(\Delta X^S(t)\right)\right\}.$$

Let k_t be the number of pure strategies in $X^S(t)$ and let $k = \Pi_{t\in T} k_t$, and κ be the open interval $\left(0, \frac{1}{k}\right)$. Define

$$\begin{aligned}
\delta_\varepsilon^t &= \left\{\sigma^S \in \Delta X^S(t) : \sigma_i^S \geq \varepsilon \forall i = 1, 2, ..., k_t\right\}, \\
\delta_\varepsilon &= \Pi_t \kappa_\varepsilon^t \\
B_\varepsilon &= \left\{\sigma^R \in \Delta X^R : U^R\left(\sigma^R, \sigma^S\right) > U^R\left(\hat{s}^R, \sigma^S\right) \forall \sigma^S \in \kappa_\varepsilon\right\}, \\
W_\varepsilon &= \left\{\sigma^R \in \Delta X^R : U^R\left(\sigma^R, \sigma^S\right) \geq U^R\left(\hat{s}^R, \sigma^S\right) \forall \sigma^S \in \kappa_\varepsilon\right\}.
\end{aligned}$$

$\hat{s}^R$ is not a best response to any $\sigma^S \in \Pi_t\left(\Delta^+ X^S(t)\right)$, so for each $\varepsilon \in \kappa$, $\hat{s}^R$ is not a best response to any $\sigma^S \in \delta_\varepsilon$. If we view δ_ε^t as set of strategies for type t Sender, then the equivalence between strong dominance and never best response establishes that B_ε is nonempty. Since W_ε is closed and nonempty, for each $\varepsilon \in \kappa$ we can choose $s_\varepsilon^R \in \Delta X^R$ that is a best response in W_ε to $\sigma_\varepsilon^S \in \delta_\varepsilon$, where $\sigma_\varepsilon^S(t)$ puts probability $\frac{1}{k_t}$ on every pure strategy in $X^S(t)$. Notice that s_ε^R yields player R strictly higher utility against σ_ε^S than $\hat{s}^R$, since $B_\varepsilon \sqsubseteq W_\varepsilon$. Choose a sequence of $\varepsilon_i' s$ in T converging to 0, such that $\left\{\sigma_{\varepsilon_i}^R\right\}$ converges. Let σ_*^R be the limit of the sequence $\left\{\sigma_{\varepsilon_i}^R\right\}$. We will show that σ_*^R weakly dominates $\hat{s}^R$.

Continuity of U^R guarantees that σ_*^R is at least as good for player R as $\hat{s}^R$ against all $\sigma^S \in \Pi\left(\Delta X^S(t)\right)$. It remains only to show that $\sigma_*^R \notin A$. If $\exists \sigma'^R \in A$ with $\sigma'^R = \sigma_*^R$, then for all sufficiently small ε_i, $\sigma_{\varepsilon_i}^R$ gives positive weight to very pure strategy given positive

weight by σ'^R. Then $\lambda > 0$ can be chosen sufficiently small so that all components of

$$\bar{\sigma}^R_{\varepsilon_i} = \left(\sigma_{\varepsilon_i} - \lambda\sigma'^R\right)\frac{1}{1-\lambda}$$

are nonnegative. For any $\sigma^R \in \delta_{\varepsilon_i}$,

$$U^R\left(\sigma^S, \bar{\sigma}^R_{\varepsilon_i}\right) - U^R\left(\sigma^S, \sigma^R_{\varepsilon_i}\right) = \frac{\lambda}{1-\lambda}\left[U^R\left(\sigma^R_{\varepsilon_i}, \sigma^S\right) - U^R\left(\hat{s}^R, \sigma^S\right)\right] \geq 0$$

because $\sigma^R_{\varepsilon_i} \in W_{\varepsilon_i}$. Moreover, the inequality is strict when σ^S is such that, for every t, $\sigma^S(t)$ puts probability $\frac{1}{k_t}$ on every pure strategy in $X^S(t)$ (denote it by $\tilde{\sigma}^S$). Thus $\bar{\sigma}^R_{\varepsilon_i}$ is in W_{ε_i} and yields player R higher utility than $\sigma^R_{\varepsilon_i}$ against $\tilde{\sigma}^S$, a contradiction.

Claim 2.5. *s^R is strongly dominated w.r.t. $\left(\Pi_t X^S(t)\right) \times X^R$ if and only if there does not exist a $\sigma^S(t) \in \Delta X^S(t)$ for every t such that $s^R \in \arg\max_{s' \in X^R} U^R\left(\left(\sigma^S(t)\right)_t, s'\right)$.*

Proof. The "only-if" part is trivial. To show the "if" part, suppose to the contrary that $\hat{s}^R$ is not a best response to any $\sigma^S \in \Pi_t\left(\Delta X^S(t)\right)$. Then there exists a function $b : \Pi_t\left(\Delta X^S(t)\right) \to X^R$ with $U^R\left(\sigma^S, b\left(\sigma^S\right)\right) > U^R\left(\sigma^S, \hat{s}^R\right)\ \forall \sigma^S$. Consider the zero-sum game

$$\bar{G} = \left(T, \Pi_t X^S(t), X^R, \bar{U}^S(;t), \bar{U}^R\right)$$

where $\bar{U}^R\left(s^S, s^R\right) = U^R\left(s^S, s^R\right) - U^R\left(s^S, \hat{s}^R\right)$ and $\bar{U}^S\left(s^S, s^R; t\right) = -U^R\left(s^S, s^R\right)\ \forall t$. Let $\left(\Pi_t \sigma^S_*(t), \sigma^R_*\right)$ be a Bayesian Nash equilibrium of $\bar{G}$. Since the interim interpretation results in the same equilibria as the ex ante interpretation,

$$\bar{U}^S\left(\Pi_t \sigma^S_*(t), \sigma^R_*\right) \geq \bar{U}^S\left(\Pi_t \sigma^S(t), \sigma^R_*\right)$$

for any $\Pi_t\sigma^S(t) \neq \Pi_t\sigma_*^S(t)$. For any $\sigma^S \in \Delta X^S$,

$$\begin{aligned}
\bar{U}^R\left(\Pi_t\sigma^S(t), \sigma_*^R\right) &\geq \bar{U}^R\left(\Pi_t\sigma_*^S(t), \sigma_*^R\right) \\
&\geq \bar{U}^R\left(\Pi_t\sigma_*^S(t), b\left(\Pi_t\sigma_*^S(t)\right)\right) \\
&= U^R\left(\Pi_t\sigma_*^S(t), b\left(\Pi_t\sigma_*^S(t)\right)\right) - U^R\left(\Pi_t\sigma_*^S(t), \hat{s}^R\right) \\
&> 0
\end{aligned}$$

So

$$U^R\left(\Pi_t\sigma^S(t), \sigma_*^R\right) > U^R\left(\Pi_t\sigma^S(t), \hat{s}^R\right) \ \forall \Pi_t\sigma^S(t) \in \Pi_t\Delta X^S(t)$$

So $\hat{s}^R$ is strongly dominated by σ_*^R. □

□

2.7.2 NIAL Results under the Interim Interpretation

To show lemma 2.5 strictly under the interim interpretation, we need the following lemmas.

Lemma 2.10. [3] *If $y^S(t) \geq t + b \forall t$, then $S(k)$ satisfies the following properties for all k:*

1. *Given any messages m_0, m_1 where $m_0 < m_1 \leq y^S(t)$ and $m_0 \in S^S(k;t)$, then m_1 belongs to $S^S(k;t)$.*

2. *If $l(k;t) < y^S(t)$, then $[l(k;t), y_S(t)] \in S^S(k;t)$;*

3. *$\forall m_0 < m_1$ such that $m_1 \in M(k)$ and $\exists s^R \in S^R(k)$ such that $s^R(m_0) \neq s^R(m_1)$, there exists $\hat{s}^R \in S^R(k)$ such that $\hat{s}^R(m_0) < \hat{s}^R(m_1) \leq m_1$.*

[3] I have a proof for property 3 for ex ante representation. The proof for property 4 relies on interim representation. Do not know whether it holds under ex ante representation.

4. For all message $\hat{m} \in M(k)$ such that there exists a strategy $s_1^R \in S^R(k)$ where $s_1^R(\hat{m}-\Delta) < \hat{m} < s_1^R(\hat{m})$, there exists another strategy $s_2^R \in S^R(k)$ where

$$\begin{aligned} s_2^R(\hat{m}+\Delta) &= s_1^R(\hat{m}+\Delta) \\ &> s_2^R(\hat{m}) \geq s_1^R(\hat{m}-\Delta). \end{aligned}$$

Proof. Prove by induction. It is obvious that properties 1 through 4 hold for $k = 0$. Suppose they hold for $j = 0, ..., k$. Property 2 is a re-phrasing of property 1. Property 1 follows from property 4. To show property 4 holds for $j = k+1$, suppose $\hat{s}^R \in S^R(k+1)$ is such that $\hat{s}^R(\hat{m}-\Delta) \neq \hat{s}^R(\hat{m})$. If $l(k;t) > \hat{m}$ for all $t > \hat{m}$, then message $\hat{m}$ can only come from types smaller or equal to $\hat{m}$. $\hat{m} \in M(k)$, so it can be shown that s^R such that $s^R(\hat{m}) > \hat{m}$ cannot be a best response to any $\sigma^S \in \Pi_{t\in T}\left(\Delta^+ S^S(k;t)\right)$. So if $\hat{s}^R(\hat{m}) \neq \hat{s}^R(\hat{m}-\Delta)$ and $\hat{s}^R \in S^R(k+1)$, then $\hat{s}^R(\hat{m}) = \hat{m}$. Property 3 is thus shown to hold and property 4 holds automatically since there does not exist $s^R \in S^R(k)$ such that $s^R(\hat{m}) > \hat{m}$. Now discuss the case where $l(k;t) \leq \hat{m}$ for some $t > \hat{m}$. According to the procedure, there exists $\hat{\sigma}^S \in \Pi_{t\in T}\left(\Delta^+ S^S(k;t)\right)$ to which $\hat{s}^R$ is a best response. Recall that $\sigma^S(.;t) \in \Delta M$. For type $t > \hat{m}$, construct $\hat{\sigma}_2^S(.;t)$ to be such that the weight on message $\hat{m}$ is moved to message $\hat{m}+\Delta$, i.e.

$$\begin{aligned} \hat{\sigma}_2^S(\hat{m}+\Delta;t) &= \hat{\sigma}^S(\hat{m};t) + \hat{\sigma}^S(\hat{m}+\Delta;t) \\ \hat{\sigma}_2^S(\hat{m};t) &= 0 \\ \hat{\sigma}_2^S(m;t) &= \hat{\sigma}^S(m;t) \ \forall m < \hat{m} \end{aligned}$$

Define $\hat{\sigma}_2^S(.;t) = \hat{\sigma}^S(.;t)$ for every type $t \leq \hat{m}$. Message $\hat{m} + \Delta$ belongs to $S^S(k;t)$ $\forall t > \hat{m}$ if $\hat{\sigma}^S(\hat{m};t) > 0$ because $y^S(t) > t + b \geq \hat{m} + b \geq \hat{m} + \Delta$ and property 2 implies that $\left[\hat{m}, y^S(t)\right] \subset S^S(k;t)$ if $\hat{m} \in S^S(k;t)$. Therefore $\hat{\sigma}_2^S(.;t) \in \Delta S^S(k;t)$ $\forall t$. Define $\hat{\sigma}_\alpha^S(.;t) \equiv (\alpha)\hat{\sigma}_2^S(.;t) + (1-\alpha)\hat{\sigma}^S(.;t)$. Because $\hat{s}^R$ is a best response to $\hat{\sigma}^S$ and $\hat{s}^R(\hat{m}) > \hat{m}$, it must be the case that $\arg\max_a U^R|_{\{\hat{m}\}}\left(\hat{\sigma}^S, a\right) > \hat{m}$. Since types that send message $\hat{m}$ under $\hat{\sigma}_2^S$ must be smaller or equal to $\hat{m}$, $\arg\max_a U^R|_{\{\hat{m}\}}\left(\hat{\sigma}_2^S, a\right) \leq \hat{m}$. So there exists $\hat{\alpha} \in (0,1)$ such that $\arg\max_a U^R|_{\{\hat{m}\}}\left(\hat{\sigma}_{\hat{\alpha}}^S, a\right) = \hat{m}$. This comes from the condition that $\frac{\partial^2 u}{\partial a^2} < 0$ and can be shown by mean value theorem.

Let $s_{\hat{\alpha}}^R$ be a best response to $\hat{\sigma}_{\hat{\alpha}}^S$. Since $\hat{\sigma}_{\hat{\alpha}}^S(t) \in \Delta^+ S^S(k;t)$ $\forall t$, it follows that $s_{\hat{\alpha}}^R \in S^R(k+1)$. It remains to show that $s_{\hat{\alpha}}^R(\hat{m}) = \hat{m}$ and $s_{\hat{\alpha}}^R(\hat{m} + \Delta) \geq \hat{s}^R(\hat{m} + \Delta)$ for property 4 to hold for $j = k+1$. If $s_{\hat{\alpha}}^R(\hat{m} - \Delta) \leq \hat{m} - \Delta$, then property 3 is shown to hold for $j = k+1$. Otherwise, define $\hat{\sigma}_3^S$ to be such that all types smaller than $\hat{m}$ send messages smaller than $\hat{m}$ and all types geater than or equal to $\hat{m}$ send messages greater than or equal to $\hat{m}$. Then if $\tilde{s}^R$ is a best response to σ^S close to $\hat{\sigma}_3^S$, it must be the case that $\tilde{s}^R(\hat{m} - \Delta) \leq \hat{m} - \Delta$ and $\tilde{s}^R(\hat{m}) \geq \hat{m}$. If $\tilde{s}^R(\hat{m}) = \hat{m}$ then property 3 is shown to hold. Otherwise, $\tilde{s}^R(\hat{m}) > \hat{m} > \tilde{s}^R(\hat{m} - \Delta)$ and we can apply the technique for $\hat{s}^R$ again to show that there exists $\tilde{s}_{\tilde{\alpha}}^R \in S^R(k+1)$ such that $\tilde{s}_{\tilde{\alpha}}^R(\hat{m}) = \hat{m}$ and $\tilde{s}_{\tilde{\alpha}}^R(\hat{m} - \Delta) \leq \hat{m} - \Delta$.

Now show that $s_{\hat{\alpha}}^R(\hat{m}) = \hat{m}$ and $s_{\hat{\alpha}}^R(\hat{m} + \Delta) \geq \hat{s}^R(\hat{m} + \Delta)$. If s^R is such that $s^R(\hat{m}) = s^R(\hat{m} + \Delta)$, then $U^R\left(\hat{\sigma}_2^S, s^R\right) = U^R\left(\hat{\sigma}^S, s^R\right)$ and therefore $U^R\left(\hat{\sigma}_{\hat{\alpha}}^S, s^R\right) = U^R\left(\hat{\sigma}^S, s^R\right)$. Construct a strategy $\phi_-\left(\hat{s}^R, \hat{m}\right)$ which is equal to $\hat{s}^R$ except on $\hat{m}$ and $\phi_-\left(\hat{s}^R, \hat{m}\right)(\hat{m}) = \hat{m}$. The strategy $\phi_-\left(\hat{s}^R, \hat{m}\right) \in S_L^R$ because $\hat{s}^R(\hat{m}) > \hat{m} > \hat{s}^R(\hat{m} - \Delta)$ by construction. It's

easy to show that

$$U^R\left(\hat{\sigma}^S_{\hat{\alpha}}, \phi_-\left(\hat{s}^R, \hat{m}\right)\right) > U^R\left(\hat{\sigma}^S_{\hat{\alpha}}, \hat{s}^R\right)$$

because $\arg\max_a U^R|_{\{\hat{m}\}}\left(\hat{\sigma}^S_{\hat{\alpha}}, a\right) = \hat{m}$. From the construction that $\hat{s}^R$ is a best response to $\hat{\sigma}^S_{\hat{\alpha}}$,

$$\begin{aligned} U^R\left(\hat{\sigma}^S_{\hat{\alpha}}, \phi_-\left(\hat{s}^R, \hat{m}\right)\right) &> U^R\left(\hat{\sigma}^S_{\hat{\alpha}}, \hat{s}^R\right) \\ &= U^R\left(\hat{\sigma}^S, \hat{s}^R\right) \\ &\geq U^R\left(\hat{\sigma}^S, s^R\right) \end{aligned}$$

for all s^R such that $s^R(\hat{m}) = s^R(\hat{m}+\Delta)$. Therefore, being a best response to $\hat{\sigma}^S_{\hat{\alpha}}$ by construction, $s^R_{\hat{\alpha}}(\hat{m}) \neq s^R_{\hat{\alpha}}(\hat{m}+\Delta)$ and hence $s^R_{\hat{\alpha}}(\hat{m}) \leq \hat{m}$. A similar argument can be used to show that $s^R_{\hat{\alpha}}(\hat{m}) = \hat{m}$ because $\arg\max_a U^R|_{\{\hat{m}\}}\left(\hat{\sigma}^S_{\hat{\alpha}}, a\right) = \hat{m}$. Now construct a strategy $\phi\left(s^R_{\hat{\alpha}}, \hat{m}, \hat{s}^R\right)$ which is equal to $s^R_{\hat{\alpha}}$ for $m \leq \hat{m}$ and is equal to $\hat{s}^R$ for $m > \hat{m}$. This new strategy $\phi\left(s^R_{\hat{\alpha}}, \hat{m}, \hat{s}^R\right)$ belongs to the language S^R_L because $s^R_{\hat{\alpha}}(\hat{m}) \leq \hat{m} < \hat{s}^R(\hat{m}+\Delta)$. For $s^R_{\hat{\alpha}}$ to be a best response to $\hat{\sigma}^S_{\hat{\alpha}}$, it has to be the case that $U^R\left(\hat{\sigma}^S_{\hat{\alpha}}, s^R_{\hat{\alpha}}\right) \geq U^R\left(\hat{\sigma}^S_{\hat{\alpha}}, \phi\left(s^R_{\hat{\alpha}}, \hat{m}, \hat{s}^R\right)\right)$. Let $\phi_+\left(s^R_{\hat{\alpha}}, \hat{m}\right)$ be a strategy which is equal to $s^R_{\hat{\alpha}}$ except on message $\hat{m}$ and $\phi_+\left(s^R_{\hat{\alpha}}, \hat{m}\right)(\hat{m}) = s^R_{\hat{\alpha}}(\hat{m}+\Delta)$. Since $s^R_{\hat{\alpha}}(\hat{m}+\Delta) > \hat{m}$, the new strategy $\phi_+\left(s^R_{\hat{\alpha}}, \hat{m}\right)$

belongs to S_L^R. Therefore,

$$\begin{aligned}
0 &\leq U^R\left(\hat{\sigma}_{\hat{\alpha}}^S, s_{\hat{\alpha}}^R\right) - U^R\left(\hat{\sigma}_{\hat{\alpha}}^S, \phi\left(s_{\hat{\alpha}}^R, \hat{m}, \hat{s}^R\right)\right) \\
&= U^R|_{[\hat{m}+\Delta,1]}\left(\hat{\sigma}_{\hat{\alpha}}^S, s_{\hat{\alpha}}^R\right) - U^R|_{[\hat{m}+\Delta,1]}\left(\hat{\sigma}_{\hat{\alpha}}^S, \hat{s}^R\right) \\
&= \sum_{t\geq\hat{m}+\Delta}\left(\hat{\sigma}^S\left(\hat{m}+\Delta;t\right)+\alpha\times\hat{\sigma}^S\left(\hat{m};t\right)\right)\left[u^R\left(t, s_{\hat{\alpha}}^R\left(\hat{m}+\Delta\right)\right) - u^R\left(t, \hat{s}^R\left(\hat{m}+\Delta\right)\right)\right] \\
&\quad + \sum_{t\leq\hat{m}}\left(\hat{\sigma}^S\left(\hat{m}+\Delta;t\right)\right)\left[u^R\left(t, s_{\hat{\alpha}}^R\left(\hat{m}+\Delta\right)\right) - u^R\left(t, \hat{s}^R\left(\hat{m}+\Delta\right)\right)\right] \\
&\quad + \sum_{m\geq\hat{m}+2\Delta}\sum_t \hat{\sigma}^S\left(m;t\right)\left[u^R\left(t, s_{\hat{\alpha}}^R\left(m\right)\right) - u^R\left(t, \hat{s}^R\left(m\right)\right)\right] \\
&= \sum_t\left(\hat{\sigma}^S\left(\hat{m}+\Delta;t\right)+\hat{\sigma}^S\left(\hat{m};t\right)\right)\left[u^R\left(t, s_{\hat{\alpha}}^R\left(\hat{m}+\Delta\right)\right) - u^R\left(t, \hat{s}^R\left(\hat{m}+\Delta\right)\right)\right] \\
&\quad + \sum_{m\geq\hat{m}+2\Delta}\sum_t \hat{\sigma}^S\left(m;t\right)\left[u^R\left(t, s_{\hat{\alpha}}^R\left(m\right)\right) - u^R\left(t, \hat{s}^R\left(m\right)\right)\right] \\
&\quad - \left\{\begin{array}{c} \sum_{t\leq\hat{m}}\hat{\sigma}^S\left(\hat{m};t\right)\left[u^R\left(t, s_{\hat{\alpha}}^R\left(\hat{m}+\Delta\right)\right) - u^R\left(t, \hat{s}^R\left(\hat{m}+\Delta\right)\right)\right] \\ +\left(1-\alpha\right)\sum_{t\geq\hat{m}+\Delta}\hat{\sigma}^S\left(\hat{m};t\right)\left[u^R\left(t, s_{\hat{\alpha}}^R\left(\hat{m}+\Delta\right)\right) - u^R\left(t, \hat{s}^R\left(\hat{m}+\Delta\right)\right)\right] \end{array}\right\} \\
&= U^R\left(\hat{\sigma}^S, \phi\left(\hat{s}^R, \hat{m}-\Delta, \phi_+\left(s_{\hat{\alpha}}^R, \hat{m}\right)\right)\right) - U^R\left(\hat{\sigma}^S, \hat{s}^R\right) \\
&\quad - \left(U^R|_{\{\hat{m}\}}\left(\hat{\sigma}_{\hat{\alpha}}^S, s_{\hat{\alpha}}^R\left(\hat{m}+\Delta\right)\right) - U^R|_{\{\hat{m}\}}\left(\hat{\sigma}_{\hat{\alpha}}^S, \hat{s}^R\left(\hat{m}+\Delta\right)\right)\right)
\end{aligned}$$

Thus,

$$\begin{aligned}
& U^R|_{\{\hat{m}\}}\left(\hat{\sigma}_{\hat{\alpha}}^S, s_{\hat{\alpha}}^R\left(\hat{m}+\Delta\right)\right) - U^R|_{\{\hat{m}\}}\left(\hat{\sigma}_{\hat{\alpha}}^S, \hat{s}^R\left(\hat{m}+\Delta\right)\right) \\
\leq\ & U^R\left(\hat{\sigma}^S, \phi\left(\hat{s}^R, \hat{m}-\Delta, \phi_+\left(s_{\hat{\alpha}}^R, \hat{m}\right)\right)\right) - U^R\left(\hat{\sigma}^S, \hat{s}^R\right) \\
\leq\ & 0
\end{aligned}$$

because $\hat{s}^R$ is a best response to $\hat{\sigma}^S$. Since

$$\arg\max_{a\in A} U^R|_{\{\hat{m}\}}\left(\hat{\sigma}_{\hat{\alpha}}^S, a\right) = \hat{m} < \hat{s}^R\left(\hat{m}+\Delta\right)$$

and $U^R|_{\{\hat{m}\}}\left(\hat{\sigma}^S_{\hat{\alpha}}, a\right)$ as a function of a inherits the concavity from u^R, we get

$$s^R_{\hat{\alpha}}(\hat{m}+\Delta) \geq \hat{s}^R(\hat{m}+\Delta)$$

It has now been shown that $s^R_{\hat{\alpha}}(\hat{m}+\Delta) \geq \hat{s}^R(\hat{m}+\Delta)$ and $s^R_{\hat{\alpha}}(\hat{m}) = \hat{m} \geq \hat{s}^R(\hat{m}-\Delta)$.

□

Define η_{k+1} iteratively to be the largest type $\hat{t} < \eta_k$ such that $l(k;t) \leq t$. That is, define

$$\eta_{k+1} \equiv \max\{t < \eta_k | l(k;t) \leq t\}$$

. Define

$$l^{-1}(k;m) = \max\{t | l(k;t) \leq m\}$$

. Then by definition, $l^{-1}(k;\eta_k) = \eta_k$ and $l^{-1}(k;m) < m$ for all $m > \eta_k$.

Lemma 2.11. *There exists $s^R \in S^R(k+1)$ such that $s^R(\eta_k) \neq s^R(\eta_k - \Delta)$, for any k.*

Proof. This can be done by showing that there exists $\hat{s}^S \in S^S(k)$ such that $\hat{s}^S(\eta_k) = \eta_k$ and $\hat{s}^S(t) \leq \eta_k - \Delta$ for all $t \leq \eta_k - \Delta$. Then show that $s^R(m) \leq m$ for all $m \geq \eta_k$ given any $s^R \in C^R_*(k+1)$. Then $\hat{s}^R$ where $\hat{s}^R(\eta_k) = \eta_k$ and $\hat{s}^R(\eta_k - \Delta) \leq \eta_k - \Delta$ does strictly better w.r.t. $\hat{\sigma}^S$ close to $\hat{s}^S$ than any other s^R. So there exists $\hat{s}^R \in C^R_*(k+1)$ where $\hat{s}^R(\eta_k) = \eta_k$ and $\hat{s}^R(\eta_k - \Delta) \leq \eta_k - \Delta \neq \hat{s}^R(\eta_k)$. □

Now lemma 2.5 follows.

Lemma 2.12. *$l(\infty;t) \geq t$ for all $t \in T$. Moreover, either $l(\infty;t) \geq y^S(t)$ or $s^R(l(\infty;t)) = s^R\left(y^S(t)\right)$ for all $s^R \in S^R(\infty)$.*

Proof. Suppose given k, there exists a type t such that $l(k;t) \leq t$. Then η_k is well defined. So for all $s^R \in S^R(k+1)$, $s^R(\eta_k) \leq \eta_k$ and there exists $\hat{s}^R \in S^R(k+1)$ where $\eta_k = \hat{s}^R(\eta_k) \neq \hat{s}^R(\eta_k - \Delta)$. So for every type t where $y_S(t) \geq \eta_k$, message $\eta_k - \Delta$ is weakly dominated by message η_k, because they prefer action η_k to any smaller action. It then follows that every message $m \leq \eta_k - \Delta$ is also weakly dominated by message η_k. So $l(k+2;t) \geq \eta_k$ for all t where $y_S(t) \geq \eta_k$. So $l(k+2;t) \geq \eta_k$ for all $t \geq \eta_k - b$. It follows that $\eta_{k+2} \leq \eta_k - b$ and thus the process does not stop at round k. So when the process stops, it has to be the case that $l(\infty;t) \geq t$ for all type t. Furthermore, either $l(\infty;t) \geq y_S(t)$ or $s^R(l(\infty;t)) = s^R\left(y^S(t)\right)$ for all $s^R \in S^R(\infty)$ because otherwise , message $l(\infty;t)$ is weakly dominated by message $y^S(t)$ since $s^R\left(y^S(t)\right) \leq y^S(t)$ for all $s^R \in S^R(\infty)$ and thus type t always prefers the action induced by message $y^S(t)$ to that induced by message $l(\infty;t)$, which contradicts the definition of $l(\infty;t)$. □

2.7.3 NIAL Results under Ex Ante Interpretation

Ex ante interpretation is equivalent to assuming that different types of Sender hold the same belief about the behavior of the Receiver. Therefore, $S^S(k)$ is no longer a product space of $S^S(k;t)$, and the proof under the interim interpretation does not apply. We'll make an assumption which is satisfied when neither type space nor action space is discretized. We make use of the assumption in the proof when type space and action space are discretized.

Assumption For any $a_1 < a_2 < a_3$ where $a_j \in A$ for $j = 1,2,3$ and $a_1 \geq y^S(0)$, there exists $\hat{t} \in T$ such that $a_2 = \arg\max_{j=1,2,3} u^S\left(\hat{t}, a_j\right)$.

Proof for Lemma 2.5

Suppose to the contrary that there exists type $\hat{t}$ such that $l\left(\infty;\hat{t}\right)<\hat{t}$ and $l\left(\infty;t\right)\geq t$ for all $t>\hat{t}$. From the definition of b, we know that every type greater or equal to $1-b$ prefers a higher action to a lower one. Thus, $l\left(\infty;t\right)=1$ for all $t\geq 1-b$ because any message smaller than 1 induces a weakly smaller action. So $\hat{t}<1-b$. $l\left(\infty;t\right)\geq t$ for all $t>\hat{t}$ implies that $l^{-1}\left(\infty;m\right)\leq m$ for all $m\geq\hat{t}$. From observation 2.5.1, we know that $s^{R}\left(m\right)\leq m$ for all $s^{R}\in S^{R}\left(\infty\right)$ and $m\in M\left(\infty\right)\cap\left[\hat{t},1\right]$. In particular, $s^{R}\left(\hat{t}\right)\leq\hat{t}$ for all $s^{R}\in S^{R}\left(\infty\right)$. If there exists $\tilde{s}^{R}\in S^{R}\left(\infty\right)$ such that $\tilde{s}^{R}\left(\hat{t}\right)\neq\tilde{s}^{R}\left(\hat{t}-\Delta\right)$, then $\tilde{s}^{R}\left(\hat{t}\right)=\hat{t}>\tilde{s}^{R}\left(\hat{t}-\Delta\right)$. Since $b\geq\Delta$, $y^{S}\left(\hat{t}\right)\geq\hat{t}+\Delta$. Therefore, any message smaller than $\hat{t}$ is weakly dominated for type $\hat{t}$ by the message $\hat{t}$. This contradicts the assumption that $l\left(\infty;\hat{t}\right)<\hat{t}$. Likewise, if there exists $m\in\left[\hat{t},y^{S}\left(\hat{t}\right)\right]$ such that there exists $s^{R}\in S^{R}\left(\infty\right)$ where $s^{R}\left(m\right)\neq s^{R}\left(l\left(\infty;\hat{t}\right)\right)$, then the message $l\left(\infty;\hat{t}\right)$ is weakly dominated by the message m. Since $l\left(\infty;\hat{t}\right)<\hat{t}$, there exists $\hat{m}\geq y^{S}\left(\hat{t}\right)$ such that $s^{R}\left(m\right)=s^{R}\left(l\left(\infty;\hat{t}\right)\right)$ for all $m\in\left[l\left(\infty;\hat{t}\right),\hat{m}\right]$.

Moreover, $s^{R}\left(l\left(\infty;\hat{t}-\Delta\right)\right)=s^{R}\left(l\left(\infty;\hat{t}\right)\right)$ for all $s^{R}\in S^{R}\left(\infty\right)$ because type $\hat{t}-\Delta$ prefers action $\hat{t}$ to any smaller action, due to the assumption that $b\geq\Delta$. So $s^{R}\left(m\right)=s^{R}\left(l\left(\infty;\hat{t}-\Delta\right)\right)$ for all $m\in\left[l\left(\infty;\hat{t}-\Delta\right),\hat{m}\right]$.

Now we need to construct $s_{*}^{S}\in S^{S}\left(\infty\right)$ such that the Receiver's ex ante best response to s_{*}^{S} must be a non-constant on the interval $\left[l\left(\infty;\hat{t}-\Delta\right),\hat{m}\right]$, which would contradict the construction of $\hat{m}$ and we would be done. We can find such sender strategy in $S^{S}\left(1\right)$. We'll suppose we can find one in $S\left(k\right)$, and then show that we can find one in $S\left(k+1\right)$. Then by induction, we an find one in $S\left(\infty\right)$.

We first show that a separating Sender strategy exists, that is, if there exists a Sender

strategy surviving round k where type $\hat{t}$ sends a message greater than or equal to $\tilde{m}$ and a Sender strategy surviving round k where type $\hat{t}-\Delta$ sends a message lower than $\tilde{m}$, then there exists a Sender strategy surviving round k where type $\hat{t}$ sends a message greater than or equal to $\tilde{m}$ and type $\hat{t}-\Delta$ sends a message lower than $\tilde{m}$. We will prove it for the case where $\tilde{m}$ is smaller than the Receiver action most preferred by type $\hat{t}-\Delta$. This is what we'll need for the proof of the lemma. But the other case can be shown with the same technique.

Lemma 2.13. *Given $\tilde{m}$ smaller than the Receiver action most preferred by type $\hat{t}-\Delta$, that is, $\tilde{m} < y^S\left(\hat{t}-\Delta\right)$, if $l\left(k;\hat{t}-\Delta\right) < \tilde{m}$ and $g\left(k;\hat{t}\right) \geq \tilde{m}$, and there exists a Receiver strategy $\hat{s}^R \in S^R\left(k-1\right)$ and a message $m > \tilde{m}$ such that*

$$u^S\left(\hat{t}, \hat{s}^R\left(m\right)\right) > u^S\left(\hat{t}, \hat{s}^R\left(\tilde{m}-\Delta\right)\right),$$

then there exists a Sender strategy $s^S_{*k} \in S^S\left(k\right)$ such that $s^S_{*k}\left(\hat{t}\right) \geq \tilde{m}$ and $s^S_{*k}\left(\hat{t}-\Delta\right) \leq \tilde{m}-\Delta$.

Proof. Define

$$m^0 := \max\left\{S^S\left(k;\hat{t}-\Delta\right) \cap \left[0, \tilde{m}-\Delta\right]\right\}$$

and

$$m^1 := \min\left\{\begin{array}{c} m \in S^S\left(k;\hat{t}\right) \mid \\ \text{there exists } s^R \in S^R\left(k\right) \text{ where} \\ u^S\left(\hat{t}, s^R\left(m\right)\right) > u^S\left(\hat{t}, s^R\left(\hat{t}-\Delta\right)\right) \end{array}\right\}.$$

We can assume w.l.o.g. that message m^1 is not equivalent to message $m^1-\Delta$ nor to message $m^1+\Delta$ for type $\hat{t}$. By definition, there exists $s^{R,h}, s^{R,l} \in S^R\left(k-1\right)$ such that type $\hat{t}$ prefers

$s^{R,h}\left(m^1\right)$ to $s^{R,h}\left(m^1+\Delta\right)$, and type $\hat{t}$ prefers $s^{R,l}\left(m^1\right)$ to $s^{R,l}\left(m^1-\Delta\right)$. If $m^1 \leq y^S\left(\hat{t}\right)$, then

$$m^1 \in \arg\max_m u^S\left(\hat{t}, s^{R,h}\left(m\right)\right)$$

and for ε sufficiently small,

$$m^1 = \arg\max_m u^S\left(\hat{t}, \left(\left(1-\varepsilon\right)s^{R,h}+\varepsilon s^{R,l}\right)\left(m\right)\right)$$

and $u^S\left(\hat{t}, \left(\left(1-\varepsilon\right)s^{R,h}+\varepsilon s^{R,l}\right)\left(m\right)\right)$ is decreasing on $\left[m^1, 1\right]$. Otherwise, $m^1 > y^S\left(\hat{t}\right)$, then

$$m^1 \in \arg\max_m u^S\left(\hat{t}, s^{R,l}\left(m\right)\right)$$

and for ε sufficiently small,

$$m^1 = \arg\max_m u^S\left(\hat{t}, \left(\left(1-\varepsilon\right)s^{R,l}+\varepsilon s^{R,h}\right)\left(m\right)\right)$$

and $u^S\left(\hat{t}, \left(\left(1-\varepsilon\right)s^{R,l}+\varepsilon s^{R,h}\right)\left(m\right)\right)$ is decreasing on $\left[m^1, 1\right]$. We conclude that there exists $\sigma^{R,1} \in \Delta S^R\left(k-1\right)$ such that

$$m^1 = \arg\max_m u^S\left(\hat{t}, \sigma^{R,1}\left(m\right)\right)$$

and $u^S\left(\hat{t}, \sigma^{R,1}\left(m\right)\right)$ is decreasing on $\left[m^1, 1\right]$. If

$$m' \leq \arg\max_m u^S\left(\hat{t}-\Delta, \sigma^{R,1}\left(m\right)\right),$$

then $u^S\left(\hat{t}-\Delta, \sigma^{R,1}\left(m\right)\right)$ is increasing on $[0, m']$. Likewise, there exists $\sigma^{R,0} \in \Delta S^R\left(k-1\right)$

such that

$$m^0 = \arg\max_m u^S\left(\hat{t}-\Delta, \sigma^{R,0}(m)\right)$$

and $u^S\left(\hat{t}-\Delta, \sigma^{R,0}(m)\right)$ is increasing on $\left[0, m^0\right]$ and $u^S\left(\hat{t}, \sigma^{R,0}(m)\right)$ is decreasing on $[m'', 1]$ for any $m'' \geq \arg\max_m u^S\left(\hat{t}, \sigma^{R,0}(m)\right)$. If either

$$\arg\max_m u^S\left(\hat{t}-\Delta, \sigma^{R,1}(m)\right) \leq \tilde{m}-\Delta$$

or

$$\arg\max_m u^S\left(\hat{t}, \sigma^{R,0}(m)\right) \geq \tilde{m},$$

then we are done. Otherwise, define

$$m^+ := \min\left\{ m \geq \tilde{m} \middle| \begin{array}{c} u^S\left(\hat{t}-\Delta, \sigma^{R,1}(m)\right) \geq u^S\left(\hat{t}-\Delta, \sigma^{R,1}\left(\hat{t}-\Delta\right)\right); \\ \text{either } m \in S^S\left(k; \hat{t}-\Delta\right), \\ \text{or there exists } s^R \in S^R(k-1) \\ \text{such that } u^S\left(\hat{t}-\Delta, s^R(m)\right) > u^S\left(\hat{t}-\Delta, s^R(\tilde{m}-\Delta)\right) \end{array} \right\}$$

and

$$m^- := \max\left\{ m \leq \tilde{m}-\Delta \middle| \begin{array}{c} u^S\left(\hat{t}, \sigma^{R,0}(\tilde{m})\right) \leq u^S\left(\hat{t}, m\right); \\ \text{there exists } s^R \in S^R(k-1) \\ \text{such that } u^S\left(\hat{t}, s^R(m)\right) > u^S\left(\hat{t}, s^R(\tilde{m})\right) \end{array} \right\}.$$

Then

$$\begin{aligned} u^S\left(\hat{t}, \sigma^{R,0}\left(m^-\right)\right) &\geq u^S\left(\hat{t}, \sigma^{R,0}\left(m^+\right)\right) \\ u^S\left(\hat{t}-\Delta, \sigma^{R,0}\left(m^-\right)\right) &> u^S\left(\hat{t}-\Delta, \sigma^{R,0}\left(m^+\right)\right) \\ u^S\left(\hat{t}, \sigma^{R,1}\left(m^-\right)\right) &< u^S\left(\hat{t}, \sigma^{R,1}\left(m^+\right)\right) \\ u^S\left(\hat{t}-\Delta, \sigma^{R,1}\left(m^-\right)\right) &\leq u^S\left(\hat{t}, \sigma^{R,1}\left(m^+\right)\right); \end{aligned}$$

both $u^S\left(\hat{t}, \sigma^{R,1}(m)\right)$ and $u^S\left(\hat{t}, \sigma^{R,0}(m)\right)$ are decreasing on $\left[m^1, 1\right]$ and increasing on $\left[0, m^0\right]$, while both $u^S\left(\hat{t}-\Delta, \sigma^{R,1}(m)\right)$ and $u^S\left(\hat{t}-\Delta, \sigma^{R,0}(m)\right)$ are both increasing on $\left[0, m^0\right]$ and decreasing on $\left[m^1, 1\right]$. And

$$u^S\left(\hat{t}-\Delta, \sigma^{R,1}(m)\right) \leq u^S\left(\hat{t}-\Delta, \sigma^{R,1}(\tilde{m}-\Delta)\right)$$

for every $m \in [\tilde{m}, m^+ - \Delta]$ and

$$u^S\left(\hat{t}, \sigma^{R,0}(m)\right) \leq u^S\left(\hat{t}, \sigma^{R,0}(\tilde{m})\right)$$

for every $m \in [m^- + \Delta, \tilde{m} - \Delta]$. By the single crossing condition,

$$\begin{aligned} & u^S\left(\hat{t}, \sigma^{R,1}\left(m^+\right)\right) - u^S\left(\hat{t}, \sigma^{R,1}\left(m^-\right)\right) \\ > \; & u^S\left(\hat{t}-\Delta, \sigma^{R,1}\left(m^+\right)\right) - u^S\left(\hat{t}-\Delta, \sigma^{R,1}\left(m^-\right)\right) \\ \geq \; & 0 \end{aligned}$$

and

$$\begin{aligned} & 0 \\ \leq \; & u^S\left(\hat{t}, \sigma^{R,0}\left(m^+\right)\right) - u^S\left(\hat{t}, \sigma^{R,0}\left(m^-\right)\right) \\ < \; & u^S\left(\hat{t}-\Delta, \sigma^{R,0}\left(m^+\right)\right) - u^S\left(\hat{t}-\Delta, \sigma^{R,0}\left(m^-\right)\right) \end{aligned}$$

and

$$\begin{aligned} & u^S\left(\hat{t}, \left(\alpha\sigma^{R,1} + (1-\alpha)\,\sigma^{R,0}\right)\left(m^+\right)\right) \\ & -u^S\left(\hat{t}, \left(\alpha\sigma^{R,1} + (1-\alpha)\,\sigma^{R,0}\right)\left(m^-\right)\right) \\ > \; & u^S\left(\hat{t}-\Delta, \left(\alpha\sigma^{R,1} + (1-\alpha)\,\sigma^{R,0}\right)\left(m^+\right)\right) \\ & -u^S\left(\hat{t}-\Delta, \left(\alpha\sigma^{R,1} + (1-\alpha)\,\sigma^{R,0}\right)\left(m^-\right)\right) \end{aligned}$$

for every α in $[0,1]$. Then by intermediate value theorem, there exists $\hat{\alpha}$ such that

$$\begin{aligned} & u^S\left(\hat{t}-\Delta, \left(\hat{\alpha}\sigma^{R,1} + (1-\hat{\alpha})\,\sigma^{R,0}\right)\left(m^+\right)\right) \\ & -u^S\left(\hat{t}-\Delta, \left(\hat{\alpha}\sigma^{R,1} + (1-\hat{\alpha})\,\sigma^{R,0}\right)\left(m^-\right)\right) = 0. \end{aligned}$$

Then there exists α^* slightly below $\hat{\alpha}$ such that

$$\begin{aligned} & u^S\left(\hat{t}, \left(\alpha^* \sigma^{R,1} + (1-\alpha^*)\,\sigma^{R,0}\right)\left(m^+\right)\right) \\ & -u^S\left(\hat{t}, \left(\alpha^* \sigma^{R,1} + (1-\alpha^*)\,\sigma^{R,0}\right)\left(m^-\right)\right) \\ > \; & 0 \\ > \; & u^S\left(\hat{t}-\Delta, \left(\alpha^* \sigma^{R,1} + (1-\alpha^*)\,\sigma^{R,0}\right)\left(m^+\right)\right) \\ & -u^S\left(\hat{t}-\Delta, \left(\alpha^* \sigma^{R,1} + (1-\alpha^*)\,\sigma^{R,0}\right)\left(m^-\right)\right). \end{aligned}$$

Since $u^S\left(\hat{t}, \left(\alpha^* \sigma^{R,1} + (1-\alpha^*)\,\sigma^{R,0}\right)(m)\right)$ and $u^S\left(\hat{t}-\Delta, \left(\alpha^* \sigma^{R,1} + (1-\alpha^*)\,\sigma^{R,0}\right)(m)\right)$ are both decreasing on $\left[m^1, 1\right]$ and increasing on $\left[0, m^0\right]$, any maximizer for both functions in m must be in $\left[m^0, m^1\right]$. We worry if both maximizers are on the same side of $\tilde{m}$. If they are both to the right ot $\tilde{m}$, then define

$$\sigma_1^{R,1} := \alpha^* \sigma^{R,1} + (1-\alpha^*)\,\sigma^{R,0}$$

and

$$m_1^+ := \min\left\{ m \geq \tilde{m} \middle| \begin{array}{c} u^S\left(\hat{t}-\Delta, \sigma_1^{R,1}(m)\right) \geq u^S\left(\hat{t}-\Delta, \sigma_1^{R,1}\left(\hat{t}-\Delta\right)\right); \\ \text{either } m \in S^S\left(k; \hat{t}-\Delta\right), \\ \text{or there exists } s^R \in S^R(k-1) \\ \text{such that } u^S\left(\hat{t}-\Delta, s^R(m)\right) > u^S\left(\hat{t}-\Delta, s^R(\tilde{m}-\Delta)\right) \end{array} \right\}.$$

Since type $\hat{t}-\Delta$ prefers $\sigma^{R,1}(\tilde{m}-\Delta)$ to $\sigma^{R,1}(m)$ and $\sigma^{R,0}(\tilde{m}-\Delta)$ to $\sigma^{R,0}(m)$ for any $m \in [\tilde{m}, m^+ - \Delta]$, it has to be the case that $m_1^+ > m^+$. Define m_1^- in the analygous way if both maximizers are to the left of $\tilde{m}-\Delta$. Likewise, $m_1^- < m^-$. Do this iteratively. It

has to stop when either $m_j^- = m^0$ or $m_j^+ = m^1$. We are done. □

Definition 2.6. *Say that property * holds for j if there exists $m \in S^S\left(j;\hat{t}\right) \cap \left[\hat{t}, \hat{m} - \Delta\right]$ and $\sigma^R \in \Delta S^R\left(j-1\right)$ such that $u^S\left(\hat{t}, \sigma^R\left(m\right)\right) > u^S\left(\hat{t}, \sigma^R\left(\hat{t} - \Delta\right)\right)$.*

Suppose k is such that property * holds for $j = 1, 2, ..., k$. Let

$$m_1^k \equiv \min \left\{ \begin{array}{c} m \in S^S\left(k;\hat{t}\right) \cap \left[\hat{t}, \hat{m} - \Delta\right] \mid \\ \exists \sigma^R \in \Delta S^R\left(k-1\right) \text{ s.t.} \\ u^S\left(\hat{t}, \sigma^R\left(m\right)\right) > u^S\left(\hat{t}, \sigma^R\left(\hat{t} - \Delta\right)\right) \end{array} \right\}.$$

Let $s_0^S \in S^S\left(\infty\right)$ be such that $s_0^S\left(\hat{t} - \Delta\right) \leq \hat{t} - \Delta$. Then there exists $\sigma_0^S \in \Delta^+ S^R\left(\infty\right)$ to which s_0^S is a best response. Define

$$\bar{m}_k = \max\left\{ S^S\left(k; \hat{t} - \Delta\right) \cap \left[0, \hat{t} - \Delta\right] \right\}.$$

Step 1 Show that it cannot be the case that $s_{*k-1}^R\left(\hat{m} + \Delta\right) = s_{*k-1}^R\left(s_{*k-1}^S\left(\hat{t}\right)\right)$ and type $\hat{t}$ prefers $s_{*k-1}^R\left(\tilde{m}\right)$ to $s_{*k-1}^R\left(s_{*k-1}^S\left(\hat{t}\right)\right)$ for some $\tilde{m} \leq \bar{m}_{k-1}$.

Proof. Suppose the opposite is true. Since $s_{*k-1}^R\left(\bar{m}_{k-1}\right) \leq \hat{t} - \Delta$, type $\hat{t}$ must also prefer $s_{*k-1}^R\left(\bar{m}_{k-1}\right)$ to $s_{*k-1}^R\left(s_{*k-1}^S\left(\hat{t}\right)\right)$. Then it has to be the case that $s_{*k-1}^R\left(\hat{m} + \Delta\right) \leq s_{*k-1}^S\left(\hat{t}\right)$, because otherwise, it is strictly better off to take action $\hat{t}$ after receiving message $s_{*k-1}^S\left(\hat{t}\right)$ against s_{*k-1}^S, which contradicts the construction that s_{*k-1}^R is a best response to s_{*k-1}^S. It also has to be the case that $s_{*k-1}^R\left(s_{*k-1}^S\left(\hat{t}\right)\right) > y^S\left(\hat{t}\right)$. Let $\tilde{k}$ be the largest $k' < k-1$ such that $m_1^{k'} < s_{*k}^R\left(s_{*k-1}^S\left(\hat{t}\right)\right)$. Then $m_1^{\tilde{k}} < s_{*k}^R\left(s_{*k-1}^S\left(\hat{t}\right)\right)$ and $m_1^{\tilde{k}+1} \geq s_{*k}^R\left(s_{*k-1}^S\left(\hat{t}\right)\right)$. Then we have $s_{*\tilde{k}}^S\left(\hat{t}\right) \leq m_1^{\tilde{k}-1}$ and $s_{*\tilde{k}}^S\left(\hat{t} - \Delta\right) = \bar{m}_{\tilde{k}}$. Then there

exists

$$\hat{s}^R_{*\tilde{k}} \in BR^R\left((1-\varepsilon)\, s^S_{*k-1} + \varepsilon s^S_{*\tilde{k}}\right) \cap S^R\left(\tilde{k}+1\right)$$

where ε is sufficiently small such that

$$\begin{aligned}
\hat{s}^R_{*\tilde{k}}\left(s^S_{*k-1}\left(\hat{t}\right)\right) &= s^R_{*k-1}\left(s^S_{*k-1}\left(\hat{t}\right)\right) > y^S\left(\hat{t}\right) \\
\hat{s}^R_{*\tilde{k}}\left(s^S_{*\tilde{k}}\left(\hat{t}\right)\right) &= \hat{t} \\
\hat{s}^R_{*\tilde{k}}\left(\bar{m}_{k-1}\right) &= s^R_{*k-1}\left(\bar{m}_{k-1}\right) \leq \hat{t} - \Delta \\
\hat{s}^R_{*\tilde{k}}\left(\bar{m}_{\tilde{k}}\right) &\leq \hat{t} - \Delta
\end{aligned}$$

because

$$s^S_{*k-1}(T) \cap \left[\bar{m}_{k-1} + \Delta, s^S_{*k-1}\left(\hat{t}\right) - \Delta\right] = \emptyset.$$

Therefore, type $\hat{t}$ must prefer action $\hat{t}$ to action $s^R_{*k-1}\left(\bar{m}_{k-1}\right)$. By assumption, type $\hat{t}$ prefers action $s^R_{*k-1}\left(\bar{m}_{k-1}\right)$ to action $s^R_{*k-1}\left(s^S_{*k-1}\left(\hat{t}\right)\right)$. It follows that type $\hat{t}$ prefers action $\hat{t}$ to action $\hat{s}^R_{*\tilde{k}}\left(s^S_{*k-1}\left(\hat{t}\right)\right)$. So

$$\arg\max_m u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}(m)\right) \subset \left[\bar{m}_{\tilde{k}} + \Delta, s^S_{*k-1}\left(\hat{t}\right) - \Delta\right]$$

If

$$\max_m u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}(m)\right) \neq u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}\left(s^S_{*\tilde{k}}\left(\hat{t}\right)\right)\right),$$

then

$$\max_m u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}(m)\right) > u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}\left(s^S_{*\tilde{k}}\left(\hat{t}\right)\right)\right)$$

and thus

$$\arg\max_m u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}(m)\right) \subset \left[s^S_{*\tilde{k}}\left(\hat{t}\right) + \Delta, s^S_{*k-1}\left(\hat{t}\right) - \Delta\right].$$

It follows that there exists

$$m^* \in S^S\left(\tilde{k}+1; \hat{t}\right) \cap \left[s^S_{*\tilde{k}}\left(\hat{t}\right) + \Delta, s^S_{*k-1}\left(\hat{t}\right) - \Delta\right]$$

where

$$\begin{aligned} u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}(m^*)\right) &> u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}\left(s^S_{*\tilde{k}}\left(\hat{t}\right)\right)\right) \\ &\geq u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}\left(\hat{t} - \Delta\right)\right). \end{aligned}$$

Thus $m_1^{\tilde{k}+1} \leq s^S_{*k-1}\left(\hat{t}\right) - \Delta$, contradiction! Therefore, it has to be the case that

$$\begin{aligned} \max_m u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}(m)\right) &= u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}\left(s^S_{*\tilde{k}}\left(\hat{t}\right)\right)\right) \\ &= u^S\left(\hat{t}, \hat{t}\right). \end{aligned}$$

Then

$$\begin{aligned} \arg\max_m u^S\left(\hat{t} - \Delta, \hat{s}^R_{*\tilde{k}}(m)\right) &= \arg\max_m u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}(m)\right) \\ &\subset \left[\bar{m}_{\tilde{k}} + \Delta, s^S_{*k-1}\left(\hat{t}\right) - \Delta\right]. \end{aligned}$$

If

$$\left[\bar{m}_{\tilde{k}} + \Delta, \hat{t} - \Delta\right] \cap \arg\max_m u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}(m)\right) \cap S^S\left(\tilde{k}+1, \hat{t}\right) \neq \emptyset,$$

then define $\tilde{m}$ to be the largest message of that set. Let $\tilde{m}_j$ be the largest message greater

than $\tilde{m}_j$ such that there exists $\tilde{s}_j^R \in S^R(j)$ such that $s^R(\tilde{m}) \neq s^R(\tilde{m}_j)$. It follows that

$$\begin{aligned}\arg\max_m u^S\left(\hat{t}, (1-\varepsilon)\,\hat{s}^R_{*\tilde{k}} + \varepsilon \tilde{s}_j^R\right) &\subset \left[\bar{m}_{\tilde{k}} + \Delta, \tilde{m}_j - \Delta\right] \\ &\subset \left[\bar{m}_{\tilde{k}} + \Delta, s^S_{*\tilde{k}}\left(\hat{t}\right) - \Delta\right]\end{aligned}$$

for every $j \leq \tilde{k}$, and message m is equivalent to message $\tilde{m}_j$ w.r.t. $S^R(j)$ for every $m \in [\tilde{m} + \Delta, \tilde{m}_j - \Delta]$. Therefore,

$$\arg\max_m u^S\left(\hat{t} - \Delta, (1-\varepsilon)\,\hat{s}^R_{*\tilde{k}} + \varepsilon \tilde{s}_j^R\right) \cap S^S\left(j+1, \hat{t} - \Delta\right) \cap \left[\bar{m}_{\tilde{k}} + \Delta, \tilde{m}\right] \neq \emptyset$$

for every $j \leq \tilde{k}$. But then

$$\begin{aligned}\emptyset &\neq S^S\left(\tilde{k}+1, \hat{t} - \Delta\right) \cap \left[\bar{m}_{\tilde{k}} + \Delta, \tilde{m}\right] \\ &\subset S^S\left(\tilde{k}+1, \hat{t} - \Delta\right) \cap \left[\bar{m}_{\tilde{k}} + \Delta, \hat{t} - \Delta\right].\end{aligned}$$

Contradiction! Therefore,

$$\left[\bar{m}_{\tilde{k}} + \Delta, \hat{t} - \Delta\right] \cap \arg\max_m u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}(m)\right) \cap S^S\left(\tilde{k}+1, \hat{t}\right) = \emptyset.$$

And thus

$$\arg\max_m u^S\left(\hat{t}, \hat{s}^R_{*\tilde{k}}(m)\right) \subset \left[\hat{t}, s^S_{*k-1}\left(\hat{t}\right) - \Delta\right].$$

It follows that $m_1^{\tilde{k}+1} < s^S_{*k-1}\left(\hat{t}\right)$, which contradicts the construction of $\tilde{k}$. Done! □

Step 2 Show that there exists

$$m^* \in \left(\arg\max_m u^S\left(\hat{t}, s^R_{*k}(m)\right)\right) \cap S^S\left(k+1; \hat{t}\right) \cap \left[\hat{t}, 1\right]$$

where there exists $\sigma^{*R} \in \Delta S^R(k)$ such that

$$u^S\left(\hat{t}, \sigma^{*R}(m^*)\right) > u^S\left(\hat{t}, \sigma^{*R}\left(\hat{t}-\Delta\right)\right).$$

Proof. Suppose to the contrary that, for every s^R_{*k} in

$$BR^R\left(s^S_{*k}\right) \cap S^R(k),$$

and every $\tilde{m}$ in

$$\arg\max_m u^S\left(\hat{t}, s^R_{*k}(m)\right) \cap S^S\left(k+1;\hat{t}\right),$$

either

$$\tilde{m} \leq \hat{t} - \Delta$$

or

$$u^S\left(\hat{t}, s^R(\tilde{m})\right) = u^S\left(\hat{t}, s^R\left(\hat{t}-\Delta\right)\right)$$

for every $s^R \in S^R(k)$.

We will first show that $\bar{m}_{k-1} = \max_{t\in[0,\hat{t}-\Delta]} \max\left(S^S(k-1;t) \cap \left[0, \hat{t}-\Delta\right]\right)$.

If $S^S\left(k+1;\hat{t}\right) \cap \left[\bar{m}_{k-1}+\Delta, \hat{t}-\Delta\right] = \emptyset$, then either $s^R_{*k-1}\left(s^S_{*k-1}\left(\hat{t}\right)\right) \neq s^R_{*k-1}(\hat{m}+\Delta)$ which implies that $s^R_{*k-1}\left(s^S_{*k-1}\left(\hat{t}\right)\right) = \hat{t}$ and

$$\arg\max_m u^S\left(\hat{t}, s^R_{*k-1}(m)\right) \subset \left[\hat{t}, s^S_{*k-1}\left(\hat{t}\right)\right]$$

and we are done, or $s^R_{*k-1}\left(s^S_{*k-1}\left(\hat{t}\right)\right) = s^R_{*k-1}(\hat{m}+\Delta)$, then Step 2 implies that

$$\arg\max_m u^S\left(\hat{t}, s^R_{*k-1}(m)\right) \subset \left[\hat{t}, s^S_{*k-1}\left(\hat{t}\right)\right].$$

We are done!

Now we discuss the case that

$$S^S\left(k+1;\hat{t}\right)\cap\left[\bar{m}_{k-1}+\Delta,\hat{t}-\Delta\right]\neq\emptyset.$$

Then there exists $\tilde{s}^S\in S^S\left(k+1\right)$ such that $\tilde{s}^S\left(\hat{t}\right)\in\left[\bar{m}_{k-1}+\Delta,\hat{t}-\Delta\right]$. Let

$$m_{up}=\min\left(S^S\left(k+1;\hat{t}+\Delta\right)\cap\left[\hat{t}+\Delta,1\right]\right).$$

Using the technique in Step 1, we can show that there exists $\tilde{s}'^S\in S^S\left(k+1\right)$ such that $\tilde{s}'^S\left(\hat{t}\right)=\tilde{s}^S\left(\hat{t}\right)\in\left[\bar{m}_{k-1}+\Delta,\hat{t}-\Delta\right]$ and $\tilde{s}'^S\left(\hat{t}+\Delta\right)\geq\hat{t}$. Then for every $s^R_{*k-1}\in BR^R\left(\left(1-\varepsilon\right)s^S_{*k-1}+\varepsilon\tilde{s}'^S\right)$ where ε is sufficiently small,

$$\begin{aligned}s^R_{*k-1}\left(\tilde{s}\left(\hat{t}\right)\right)&=\hat{t}\\ s^R_{*k-1}\left(\bar{m}_{k-1}\right)&\leq\hat{t}-\Delta.\end{aligned}$$

From the assumption,

$$\arg\max_m u^S\left(\hat{t},s^S_{*k-1}\left(m\right)\right)\cap\left[\bar{m}_{k-1}+\Delta,\hat{t}-\Delta\right]\neq\emptyset.$$

Then Step 1 implies that

$$\begin{aligned}\max_m u^S\left(\hat{t},s^R_{*k-1}\left(m\right)\right)&=u^S\left(\hat{t},s^R_{*k-1}\left(\tilde{s}^S\left(\hat{t}\right)\right)\right)\\&=u^S\left(\hat{t},s^R_{*k-1}\left(s^S_{*k-1}\left(\hat{t}\right)\right)\right)\\&=u^S\left(\hat{t},\hat{t}\right).\end{aligned}$$

Therefore,

$$\arg\max_m u^S\left(\hat{t}-\Delta, s^R_{*k-1}(m)\right) \subset \left[\bar{m}_{k-1}+\Delta, \hat{m}\right].$$

Since $\tilde{s}^S \in S^S(k+1)$, there exists $\tilde{s}^R \in S^R(k)$ such that

$$u^S\left(\hat{t}, \tilde{s}^R\left(\tilde{s}^S\left(\hat{t}\right)\right)\right) > u^S\left(\hat{t}, \tilde{s}^R\left(\tilde{s}^S\left(\hat{t}\right)+\Delta\right)\right).$$

Then

$$\begin{aligned}\arg\max_m u^S\left(\hat{t}-\Delta, \left(\left(1-\varepsilon\right)s^R_{*k-1}+\varepsilon\tilde{s}^R\right)(m)\right) &\subset \left[\bar{m}_{k-1}+\Delta, \tilde{s}^S\left(\hat{t}\right)\right] \\ &\subset \left[\bar{m}_{k-1}+\Delta, \hat{t}-\Delta\right].\end{aligned}$$

So

$$\left[\bar{m}_{k-1}+\Delta, \hat{t}-\Delta\right] \cap S^S\left(k+1; \hat{t}-\Delta\right) \neq \emptyset.$$

Contradiction! □

Therefore,

$$\arg\max_m u^S\left(\hat{t}, \left(\left(1-\varepsilon\right)\sigma^R_0+\varepsilon s^R_{*k}\right)(m)\right) \subset \left[\hat{t}, \hat{m}\right]$$

and

$$\max_m u^S\left(\hat{t}, \left(\left(1-\varepsilon\right)\sigma^R_0+\varepsilon s^R_{*k}\right)(m)\right) > u^S\left(\hat{t}, \left(\left(1-\varepsilon\right)\sigma^R_0+\varepsilon s^R_{*k}\right)\left(\hat{t}-\Delta\right)\right).$$

We have thus shown that property $*$ holds for $k+1$. By induction, it holds for $S^S(\infty)$ and we are done!

Chapter 3

Sender-Receiver Game — Extensive Form

To resolve this tension presented in example 2.4 of Chapter One , we propose a notion of weak sequential rationality with language and an extensive form iterative procedure. The key observation motivating our definition of weak sequential rationality is that the Sender does not distinguish between messages that induce the same action, and hence the Receiver does not either. We view messages as a coordination device to achieve a mapping from types to actions, which is called an outcome. We decompose a strategy profile into the usage of messages and the induced outcome. In order to capture the idea that language takes care of the usage of messages, while rationality concerns determine the set of possible outcomes, we define a concept of sequential rationality in terms of the outcome induced by a strategy profile, instead of the strategy profile itself. This extensive form iterative procedure always yields a nonempty limiting set. When the original game has multiple equilibria, it eliminates some of the less informative outcomes like the normal form procedure. When babbling is

the unique equilibrium in the original game, it also yields babbling as the unique prediction.

The Rest of the chapter is organized as follows. Section 3.1 motivates and defines weak sequential rationality and the procedure for $EIAL$. Section 3.2 characterizes the solution to $EIAL$.

3.1 Weak Sequential Rationality and the Extensive Form Procedure

Sequential rationality is not a novel issue, and a natural first step is to add the requirement into the iterative procedure. Recall the standard definition of sequential rationality. A Receiver strategy σ^R is sequentially rational with respect to a belief σ^S if and only if $\sigma^R(m)$ is optimal at every message m according to the Bayesian update of σ^S. However, with a simple opposing-interest example, we show that this definition may clash with language combined with iterative admissibility. We argue that a weaker notion of sequential rationality, in terms of the induced outcome instead of the strategy profile, can better capture the idea of language, because messages serve only as coordination device. We then develop the extensive form procedure $EIAL$. Example 2.4 is revisited to show the predictions of IA combined with different sequential rationality notion. It is shown that the limiting set of $EIAL$ is nonempty.

3.1.1 The Opposing-interest Game

Let's look at the game in figure 3.1 where the Sender and the Receiver have opposing interest. When the true state is $West$, the Receiver wants to take action W while the Sender wants the Receiver to take action E and vice versa when the true state is $East$. The

		a	
		West	East
t	West	0,1	2,0
	East	2,0	0,1

Table 3.1: Opposing Interest Game

probability that the true state is $West$ is $\frac{2}{3}$ and the probability that the true state is $East$ is $\frac{1}{3}$. If the players cannot communicate before the Receiver takes an action, it's optimal for the Receiver to take action W. This game has a unique babbling equilibrium. $NIAL$ gives a unique solution where the Receiver takes action W to both messages, which is the same as in the babbling equilibrium.

We derive the solution to $NIAL$ in this game as follows. The bottom part of table 3.2 shows all the Receiver strategies in G_L, the game with language. In the first round of deletion, the strategy *Stubborn E* is eliminated because it is strongly dominated by *Stubborn W* since taking action W is optimal without communication. Nothing else can further be eliminated for the Receiver. For type $West$ Sender, sending message "*west*" is weakly dominated by sending message "*east*," because type $West$ prefers action E to action W, and either both messages lead to the same action, or message "*west*" leads to action W and message "*east*" leads action E. Similarly, for type $East$ Sender, sending message "*east*" is weakly dominated by sending message "*west*." In summary, the only strategy that survives the first round of deletion for each type of Sender is to utter the desired action. Call it s^S_{prefer}. That is, $s^S_{prefer}(West) =$ "*east*" and $s^S_{prefer}(East) =$ "*west*". In the second round of deletion, the only conjecture the Receiver can have about the Sender's behavior is s^S_{prefer}. *Stubborn W* strategy strictly dominates *Literal* strategy with respect to s^S_{prefer}. The two strategies differ only on the actions taken after receiving message "*east*." At round

types of Sender sending m in s^S_{prefer}	$East$	$West$
$s^R \backslash$**message**	**"West"**	**"East"**
Stubborn W	W	W
Stubborn E	E	E
Literal	W	E

Table 3.2: Language in Opposing Interest Game

2, message **"***east***"** can only come from a type $West$ Sender, and *Stubborn W* strategy takes action W there, which is better against type $West$ than action E, the action taken by *Literal* strategy. Therefore, we end up with a unique prediction $S(\infty) = \left\{s^S_{prefer}, Stubborn\ W\right\}$, which gives the babbling outcome.

However, *Stubborn W* is not interim optimal with respect to s^S_{prefer} when the Receiver receives message "*west*", even though $NIAL$ prediction is equal to the unique equilibrium outcome in the original game. The top row in table 3.2 illustrates the correspondence between messages and types under the sender strategy s^S_{prefer}. As is shown, message "*west*" can only come from type $East$. But when the true state is $East$, the optimal action is action E , not action W which is taken by the *Stubborn W* strategy. The unique strategy which is sequentially rational with respect to s^S_{prefer} is the *Opposite* strategy (E, W). But (E, W) does not belong to language, and therefore is physically unavailable. Since s^S_{prefer} is the only conjecture the Receiver can have in the second round, none of the strategies in language satisfies standard sequential rationality in the second round. Therefore, imposing standard sequential rationality in the iterative procedure would yield an empty set.

3.1.2 Weak Sequential Rationality

To see what drives this result and how to tackle it, it might be worthwhile to look at sequential rationality by its components. In this two-stage game, sequential rationality

can be broken down into ex ante rationality and interim rationality. Ex ante rationality means utility maximization at the hypothetical initial node, before the receiver receives the message. Interim rationality means utility maximization at every information set in the second stage, after receiving the message. Interim rationality implies ex ante rationality. In the game without language, ex ante rationality implies interim rationality. We showed that the latter does not hold in the game with language. While ex ante rationality is taken care of by normal form analysis, the problem lies in interim rationality. The above discussion shows that no Receiver strategy in language is sequentially rational with respect to s^S_{prefer}. This is because no Receiver strategy in language is interim rational with respect to s^S_{prefer}. Since every outcome can be achieved in the game with language by some strategy profile, and thus any information can be successfully transmitted by some message usage specified by language, we wonder what s^S_{prefer} represents in the game with language: is it meant to convey information?

This project focuses on the set of outcomes: language specifies how messages are used to achieve a given set of outcomes. But standard interim rationality is defined in terms of strategy profiles, not outcomes. In addition, it is sensitive to the number of messages employed to convey the given information. Consider the opposing-interest game without language. We want to find the smallest set of strategy profiles which satisfies the following two properties: 1) it contains all babbling strategies; 2) it contains all pure strategies getting positive weight in the set; 2) it is closed under standard interim rationality. Suppose the message space is trivial and contains only one message. Then the smallest such set contains only babbling outcome. But if there are two messages in the message space, then one

babbling strategy for each type of the sender is to randomize over the two messages. Then to contain all supporting pure strategies, this set needs to contain the two sender strategies where type $West$ sends one message and type $East$ sends the other message. One such strategy is s^S_{prefer}, where type $West$ utters "*east*" and type $East$ says "*west*." Let's call the other strategy s^S_{honest}. Then to contain Receiver strategies that are interim rational with respect to these two sender strategies, this set needs to contain both *Literal* and *Opposite*, where the Receiver takes different actions after receiving different messages. This set then has to contain two separating outcomes. However, if we look at interim rationality in terms of outcomes, ignoring altogether how messages are used, we'll avoid this dependence. Since we use language to take care of how messages are used, it might be natural to look for a notion of sequential rationality that deals only with the outcomes.

Given a Receiver strategy s^R and a belief σ^S. Typically we say that the profile $\left(s^R, \sigma^S\right)$ gives rise to an outcome which is a mapping from the type space to distributions over the action space. From the Receiver's point of view, however, the profile $\left(\sigma^S, s^R\right)$ gives rise to an association between actions in the range of s^R and distributions over the type space. Let $\beta_{(\sigma^S, s^R)}$ denote the association induced by the profile $\left(\sigma^S, s^R\right)$ and $s^R(M)$ denote the range of s^R. Then, $\beta_{(\sigma^S, s^R)} : s^R(M) \to \Delta T$ is defined by

$$\beta_{(\sigma^S, s^R)}(a)(t) = \frac{\sum_{m \in (s^R)^{-1}(a)} \sigma^S(m; t)}{\sum_{t'} \sum_{m \in (s^R)^{-1}(a)} \sigma^S(m; t')},$$

which is simply a Bayesian update. For each action a that the Receiver takes in response to some message $m \in M$, $\beta_{(\sigma^S, s^R)}$ associates with it a probability distribution on the type space T, which represents the distribution of the types of the Sender that might receive

this action a under the profile $\left(\sigma^S, s^R\right)$. Standard interim rationality looks at $\left(\sigma^S, s^R\right)$. We propose checking interim rationality from the point of view of $\beta_{(\sigma^S, s^R)}$. It is formally stated as follows:

Definition 3.1 (Outcome Interim Rationality). *Let B denote a subset of A. Say that $\beta : B \to \Delta T$ is outcome interim rational if and only if*

$$a \in \arg\max_{a' \in A} \sum_t \beta(a)(t)\, u^R\left(t, a'\right)$$

for all $a \in B$.

Definition 3.2 (Weak Interim Rationality). *Say that s^R is weakly interim rational with respect to σ^S if and only if $\beta_{(\sigma^S, s^R)}$ is outcome interim rational.*

It is easy to see that *Stubborn W* is weakly interim rational with respect to s^S_{prefer}. Actually, *Stubborn W* is weakly interim rational with respect to every $\sigma^S \in \Delta S^S$. In general, for every $\sigma^S \in \Delta S^S$, there exists a s^R in language that is weakly interim rational with respect to σ^S. Using outcome interim rationality, we avoid the problem that there might exist some conjectures σ^S with respect to which no Receiver strategy in language is interim rational. However, unlike standard interim rationality, weak interim rationality does not necessarily imply ex ante rationality. Given a belief σ^S, there are typically many Receiver strategies that are weakly interim rational with respect to σ^S and can be Pareto ranked. The idea of sequential rationality is that strategies that are not "rational" at the interim stage are not credible. This motivates a weaker notion of ex ante rationality: compare ex ante payoff among only "credible" Receiver strategies. More precisely, given σ^S, only strategies that are weakly interim rational with respect to σ^S are credible. Ex

ante, the Receiver picks among these "credible" strategies one that gives him the highest ex ante payoff. We combine the weaker notion of ex ante rationality and weak interim rationality analogously to define weak sequential rationality. Breaking standard sequential rationality into the two parts and putting them back this way does not alter the implication, since every Receiver strategy that is interim rational with respect to a conjecture σ^S is ex ante payoff equivalent to each other.

Definition 3.3 (Weak Sequential Rationality). *Let $X^R \subset S^R$. Say that s^R is weakly sequentially rational among X^R with respect to σ^S if and only if s^R is ex ante optimal with respect to σ^S among Receiver strategies that are weakly interim rational with respect to σ^S. That is,*

$$s^R \in \arg \max_{\substack{s^{R\prime} \in X^R \\ s^{R\prime} \text{ is weakly interim optimal} \\ \text{w.r.t. } \sigma^S}} U^R\left(\sigma^S, s^{R\prime}\right).$$

Call $\left(\sigma^S, \sigma^R\right)$ a weak sequential equilibrium if and only if σ^S is sequentially rational with respect to σ^R and s^R is weakly sequentially rational with respect to σ^S for every s^R in the support of σ^R. Let $WSEQ\left(G\right)$ denote the set of weak sequential equilibrium in the game without language and $WSEQ\left(G_L\right)$ denote the set of weak sequential equilibrium in the game with language. Then $WSEQ\left(G\right) = WSEQ\left(G_L\right)$. Recall that in Example 2.4, where $NIAL$ selects a unique informative outcome while the unique equilibrium in the original game is babbling, the set of equilibrium outcomes in the game with language strictly contain the set of equilibrium outcomes in the game without language. That is, $EQ\left(G\right) \subsetneqq EQ\left(G_L\right)$. It is then not that surprising that $NIAL$, being iterative admissibility on G_L, does not select any equilibrium in $EQ\left(G\right)$. Imposing weak sequential rationality restores the equilibrium outcomes in G_L to the equilibrium outcomes in G. This gives us

hope that this definition might work.

The motivation for outcome interim rationality is that, instead of truly conveying information, s^S_{prefer} might simply be a supporting pure strategy of the mixed babbling sender strategy. But in the second round, s^S_{prefer} is the only conjecture the Receiver can hold. If s^S_{prefer} represents only a supporting pure strategy of the mixed babbling strategy, the other supporting pure strategy should also be contained as a possible conjecture held by the Receiver.

We now explain how the combination of language and weak dominance selects $\left\{s^S_{prefer}\right\}$ as the unique conjecture the Receiver can hold in the second round and why it is more properly viewed as a pure strategy supporting the mixed babbling sender strategy. In the first round, we eliminated all sender strategies except s^S_{prefer}. The elimination takes place because the sender takes into account the possibility of the strategy *Literal* being used. Suppose instead that the sender believes that *Literal* is not going to be used in the game with language, and therefore the Receiver always ignores messages. Then the two messages have exactly the same implication to the Sender, and therefore she might as well randomize. No message is weakly dominated for either type, and all sender strategies are possible. This points to the well-known force of weak dominance: the reason for eliminating one strategy might later be eliminated. To show the role language plays, consider the game without language. If the sender takes into account all four strategies, she is not sure which one induces her preferred action more often. The two messages again look the same to her, though she might prefer one message under some conjecture, while another under another conjecture. We'll end up with everything in the prediction, which is not clear whether it

represents no information transmission, or simply no predicting power.

Language gives a bite by specifying the asymmetry. Though it does not rule out any outcome, weak dominance forces the sender to take into account all communication outcomes. Babbling is present in every cheap talk game because it is self-fulfilling: if the receiver always takes the same action, and the sender wholeheartedly believes that, then the sender sees the two messages as the same and might very well randomize between the two. This in turn makes it optimal for the Receiver to treat the two messages equally and therefore always take the same action. Language and weak dominance breaks out of this by making the sender take into account all outcomes. But the danger lies in going to the other extreme and taking into account outcomes that cannot happen in the game in question. This gives rise to selecting $\left\{s^S_{prefer}\right\}$ as the unique sender strategy profile in a babbling outcome, because the sender takes into account even outcomes that are not possible in the situation.

3.1.3 The Procedure for the Extensive Iterative Admissibility with Language (EIAL)

The idea is to let language take care of how messages are used and use weak sequential rationality to take care of rationality. That $WSEQ(G_L) = WSEQ(G)$ is encouraging. We then define the iterative procedure in an analogous way. We call this procedure extensive form iterative admissibility with language ($EIAL$).

Let $ES^S(0;t) = M$, $\forall t$ and $ES^R(0) = S^R_L$.

Procedure 1. $s^R \in ES^R(k+1)$ iff

a. $s^R \in ES^R(k)$

b. there exists a totally mixed conjecture $\sigma^S \in \Pi_{t\in T}\left(\Delta^+ ES^S(k;t)\right)$ such that s^R is weakly sequentially rational with respect to σ^S.

2. $s^S(t) \in ES^S(k+1;t)$ iff

a. $s^S(t) \in ES^S(k;t)$

b. there exists a totally mixed conjecture $\sigma^R \in \Delta^+ ES^R(k)$ such that $s^S(t)$ is a best response among $ES^S(k+1;t)$ with respect to σ^R.

Definition 3.4. *Write* $\cap_{k=0}^{\infty} ES^i(k) = ES^i(\infty)$ *and* $ES(k) = \left(ES^S(k), ES^R(k)\right)$.

It is easy to see that $EIAL$ gives the same prediction as $NIAL$ in the opposing interest game. Now let's look at the prediction of $EIAL$ on the game in example 2.4 in section 2.5.2.

Example 1 Revisited There are three types: type $0, \frac{1}{2}$ and 1. The common prior is such that $\pi(0) = \frac{1}{3}$; $\pi\left(\frac{1}{2}\right) = \frac{4}{9}$ and $\pi(1) = \frac{2}{9}$. Both the Sender and the Receiver have quadratic loss function: $u^R(t,a) = -(t-a)^2$ and $u^S(t,a) = -\left(t+\frac{1}{2}-a\right)^2$.

Recall that the unique equilibrium in the game without language is babbling, while $NIAL$ selects a unique informative strategy $\left(s^S_{nice}, s^R_{nice}\right)$. Table 2.4 shows all the Receiver strategies in language. The first round of deletion is the same as in the normal form procedure: for the Receiver, every strategy in language except $(0,0,0)$ and $(1,1,1)$ are retained; for each type of the Sender, only $s^S_{nice}(t)$ is retained. The second round of the extensive form procedure is different from that of $NIAL$. Suppose s^R survives the second round of deletion. Then it is necessary for s^R to be weakly interim rational with respect to s^S_{nice}, which is the only conjecture the Receiver can hold at the second round. It is then

	Language	No Language
EQ	s^R_{babble}, s^R_{nice}	s^R_{babble}
WSEQ	s^R_{babble}	s^R_{babble}
IA	s^R_{nice}	everything
IA+Standard Sequential Rationality	empty	everything
Weak IA+Weak Sequential Rationality	s^R_{babble}	everything

necessary that s^R takes the same action at both message $\frac{1}{2}$ and message 1. Suppose to the contrary that $s^R\left(\frac{1}{2}\right) \neq s^R(1)$, then by the assumption of language, $s^R(1) = 1$. From the first two rows in table 2.4, we can see that both type $\frac{1}{2}$ and type 1 senders send only message 1. Since $s^R(1) \neq s^R\left(\frac{1}{2}\right)$, action 1 is associated with a posterior belief that puts probability $\frac{2}{3}$ on type $\frac{1}{2}$ and probability $\frac{1}{3}$ on type 1. The best action given this distribution is action $\frac{1}{2}$, not action 1. So s^R is not weakly interim rational with respect to s^S_{nice}. We've then shown that to be weakly interim rational with respect to s^S_{nice}, it is necessary to take the same action at both message $\frac{1}{2}$ and message 1. Then every type receives the same action since all types of the Sender send either message $\frac{1}{2}$ or message 1. Thus $\left(0, \frac{1}{2}, \frac{1}{2}\right)$ and $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ are both weakly sequentially rational with respect to s^S_{nice}. The unique outcome predicted by $EIAL$ is that all types of the Sender receive action $\frac{1}{2}$, which is the same as the babbling outcome.

Table 3.1.3 summarizes the predictions of the game in example 3.1 under different procedures.

We now establish nonemptiness of the limit. We need one notation here. Define $\varepsilon\left(\sigma^S, \sigma^{S\prime}\right) : T \to \Delta M$ by $\varepsilon\left(\sigma^S, \sigma^{S\prime}\right)(t) \equiv (1-\varepsilon)\,\sigma^S(t) + \varepsilon\sigma^{S\prime}(t)$.

Lemma 3.1. *$ES(\infty)$ is nonempty.*

Proof. Since $S_L = \left(S^S, S^R_L\right)$ is finite, the elimination process must stop after finite steps.

It suffices to show that $ES(k+1)$ is nonempty if $ES(k)$ is nonempty. It is obvious from the iterative procedure that $ES^S(k+1)$ is nonempty, since S_L is finite and for every $\sigma^R \in \Delta^+ES^R(k)$, there exists $s^S \in ES^S(k)$ which attains the maximum payoff among Sender strategies in $ES^S(k)$. To show that $ES^R(k+1)$ is nonempty, it suffices to show that there exists a totally mixed conjecture on $ES^S(k)$, i.e. $\sigma^S \in \Delta^+ES^S(k)$, such that there exists s^R in $ES^R(k)$ which is weakly interim rational with respect to σ^S. It is obvious that for every $\sigma^S \in \Delta^+ES^S(0) = \Delta^+S^S$, there exists $s^R \in ES^R(0) = S^R$ which is weakly interim rational with respect to σ^S, since a constant s^R which plays the best action against the prior is weakly interim rational with respect to σ^S. Let us pick any $\sigma^S_{k-1} \in \Delta^+ES^S(k-1)$. Since $ES^S(k) \subset ES^S(k-1)$, $\varepsilon\left(\sigma^S_k, \sigma^S_{k-1}\right) \in \Delta^+ES^S(k-1)$ for any σ^S_k in $ES^S(k)$. To shorten the notation, write $\varepsilon\left(\sigma^S_k, \sigma^S_{k-1}\right)$ simply as $\sigma^S_{k,\varepsilon}$. Lemma 3.2 implies that for every ε small enough,

$$\begin{aligned} & \left\{s^R \in ES^R(k-1) : s^R \text{ is weakly interim rational with respect to } \sigma^S_{k,\varepsilon}\right\} \\ \subset \quad & \left\{s^R \in ES^R(k) : s^R \text{ is weakly interim rational with respect to } \sigma^S_k\right\}. \end{aligned}$$

By hypothesis, the set

$$\left\{s^R \in ES^R(k-1) : s^R \text{ is weakly interim rational with respect to } \sigma^S_{k,\varepsilon}\right\}$$

is nonempty. Since $ES^R(k-1)$ is finite, there exists a Receiver strategy $s^R \in ES^R(k-1)$ which attains the maximum expected utility among all strategies that are weakly interim

rational with respect to $\sigma^S_{k,\varepsilon}$. Therefore, the set

$$\left\{s^R \in ES^R(k) : s^R \text{ is weakly sequentially rational with respect to } \sigma^S_k\right\}$$

is nonempty. The proof is then completed by induction. □

Lemma 3.2. *Given any $\sigma^S \in \Delta S^S$, there exists $\bar{\varepsilon} > 0$ such that for all $\sigma^{S\prime} \in \Delta S^S$ and $\varepsilon < \bar{\varepsilon}$,*

$$\begin{aligned} & \left\{s_R \in S_R | s_R \textit{ is weakly sequentially rational with respect to } \varepsilon\left(\sigma^S, \sigma^{S\prime}\right)\right\} \\ \subset \quad & \left\{s_R \in S_R | s_R \textit{ is weakly sequentially rational with respect to } \sigma^S\right\}. \end{aligned}$$

The proof is left to the Appendix.

3.2 Characterization

The extensive form procedure ($EIAL$) is motivated by the example illustrating that $NIAL$ might not select any equilibrium outcome of the original game. We showed that $EIAL$ restores babbling as the unique prediction in that example. In this section, we show that this result is general, i.e., $EIAL$ selects babbling as the unique outcome when babbling is the unique equilibrium in the original game. However, we are able to show that $EIAL$ contains at least one equilibrium outcome in the original game only under monotonicity condition (M) (defined in section 2.4) and with the interim interpretation. On the other

hand, with $EIAL$, we are able to show the lower bound on the amount of information transmission only under ex ante interpretation. We do not have a tight characterization when the monotonicity condition (M) is satisfied. Showing inclusion of strategies under interim representation is easier, while showing exclusion of strategies under ex ante representation is easier. Therefore, our current results under $EIAL$ depend on whether interim representation or ex ante representation is employed.

Proposition 3.1. *If babbling is the unique equilibrium, then babbling is the only outcome under $EIAL$.*

Proof. Say that $X \subset S$ contains an informative outcome if there exists $\left(s^S, s^R\right) \in X$ such that there are two different types $t_1 \neq t_2$ where $s^R\left(s^S\left(t_1\right)\right) \neq s^R\left(s^S\left(t_2\right)\right)$. We show that if $ES\left(k\right)$ contains an informative outcome, then the iterative process does not stop, i.e., $ES\left(k+1\right) \subsetneqq ES\left(k\right)$. Since we have shown that the limiting set is nonempty, it is necessary that $ES\left(\infty\right)$ contains no informative outcomes. Thus $EIAL$ predicts that every type receives the same action. Since the strictly best constant strategy is to play the best action against the prior, we get the babbling outcome.

To see why the iterative process does not stop when $ES\left(k\right)$ contains an informative outcome, note that for babbling to be the unique equilibrium in the original game, it has to be the case that every type t prefers to be pooled with all higher types than with all lower types. Whenever it is not the case that all messages induce the same action, some type t will want to discard a message which always receives the lowest action.

The details of the proof is left to the Appendix. □

Proposition 3.2. *If condition (M) holds, then the most informative equilibrium outcome is contained in EIAL under the interim interpretation.*

We prove it by showing that if every type exaggerates the most they want and sends the highest message they might use in $ES^S(\infty)$, the best the Receiver can do without violating either language or weak interim rationality is to play the most informative equilibrium strategy. The details are left in the Appendix.

When ex ante interpretation is employed, however, we need to make sure that there exists one single Receiver strategy with respect to which every type of the Sender wants to exaggerate the most. Therefore, the proof of proposition 3.2 does not carry through directly. Under the interim interpretation, different types are allowed to hold different beliefs about the behavior of the Receiver. Therefore, it is easier to construct a sender strategy profile in the limiting set, and therefore easier to show that a Receiver strategy belongs to the limiting set.

Now we state the result of the lower bound on the amount of information transmission. Proposition 3.3 says that every Receiver strategy in the limit partitions the set of messages used in the limit into at least L intervals.

Proposition 3.3. *With ex ante interpretation, under EIAL, there exists a nontrivial lower bound on the number of different actions taken on $ES^R(\infty)$. Specifically, if the game admits a non-babbling equilibrium, then the number of different actions taken in $ES^R(\infty)$ is at least 2.*

Proof. A Receiver strategy s^R partitions the message space. If s^R is weakly interim rational with respect to σ^S, then a partition determines s^R. A finer partition is unambiguously

better. However, a finer partition might violate the language restriction. We show that a Receiver strategy cannot have a step that is too wide, because otherwise there exists a finer partition that satisfies language and is weakly interim rational. The same logic is used to attain a lower bound. □

We need the following two observations to proceed with the proof. The first claim relates the minimum action the Receiver might take at message $m-\Delta$ to the value m. It holds only under ex ante interpretation. The second claim gives a relation between the lowest type that might send messages in $[m,1]$ to the value m.

$s^R(m-\Delta) \geq E([0,\rho(m)])$ for all $s^R \in ES^R(1)$ such that $s^R(m) \neq s^R(m-\Delta)$.

Claim Type $g^{-1}(\infty;m)$ prefers action $a=m$ to action $E\left(\left[0,g^{-1}(\infty;m)-\Delta\right]\right)$.

In particular, $g^{-1}\left(\infty;E\left(\left[t_1^2,1\right]\right)\right) \geq t_1^2$. Given any $\sigma^S \in \Delta S^S$, define

$$s^R_{sep2}(m) \equiv \begin{cases} \arg\max_a U^R|_{\left[0,E\left(\left[t_1^2,1\right]\right)-\Delta\right]}\left(\sigma^S,a\right) & m \in \left[0,E\left(\left[t_1^2,1\right]\right)-\Delta\right] \\ \arg\max_a U^R|_{\left[E\left(\left[t_1^2,1\right]\right),1\right]}\left(\sigma^S,a\right) & m \in \left[E\left(\left[t_1^2,1\right]\right),1\right] \end{cases}$$

It is obvious that s^R_{sep2} is weakly interim rational with respect to σ^S. To show that s^R_{sep2} belongs to language, we need to show that

$$\arg\max_a U^R|_{\left[0,E\left(\left[t_1^2,1\right]\right)-\Delta\right]}\left(\sigma^S,a\right) \leq E\left(\left[t_1^2,1\right]\right)-\Delta \tag{3.1}$$

$$\arg\max_a U^R|_{\left[E\left(\left[t_1^2,1\right]\right),1\right]}\left(\sigma^S,a\right) \geq E\left(\left[t_1^2,1\right]\right). \tag{3.2}$$

Ex ante interpretation implies that every pure Sender strategy $s^S \in ES^S(1)$ is weakly increasing in t. Therefore, $\arg\max_a U^R|_{[0,m]}\left(s^S,a\right) \leq E([0,1])$ for every m and every

$s^S \in ES^S(1)$. It follows that $\arg\max_a U^R|_{[0,m]}\left(\sigma^S, a\right) \leq E([0,1])$ for every m and every $\sigma^S \in \Delta ES^S(1)$. Thus,

$$\begin{aligned} \arg\max_a U^R|_{[0,E([t_1^2,1])-\Delta]}\left(\sigma^S, a\right) &\leq E([0,1]) \\ &< E\left(\left[t_1^2, 1\right]\right). \end{aligned}$$

This gives us inequality 3.1. We showed that the smallest type that can send any message higher than or equal to message $E\left(\left[t_1^2,1\right]\right)$ is greater than $t_1^2 - \Delta$. Therefore, $E\left(\left[g^{-1}\left(\infty; E\left(\left[t_1^2,1\right]\right)\right),1\right]\right) \geq E\left(\left[t_1^2,1\right]\right)$. Inequality 3.2 then follows. So a constant Receiver strategy cannot be weakly sequentially rational with respect to σ^S, because s^R_{sep2} is weakly interim rational with respect to σ^S and gives a higher ex ante payoff than a constant Receiver strategy.

Therefore, s^R must partition $M(\infty)$ into at least two intervals. Let $\{a_1, ..., a_q\}$ be the set of actions taken by s^R on $M(\infty)$, where $a_j < a_{j+1}$. Let m_j be the smallest message on which s^R takes the value a_j. Let

$$\hat{m}_q \equiv \max\left\{m \in M | E\left(\left[g^{-1}(\infty; m), 1\right]\right) \leq m\right\}.$$

It follows that $\hat{m}_q \geq E\left(\left[t_1^2, 1\right]\right)$.

Claim $a_q \geq \hat{m}_q$ or a_q is such that

$$g(\infty; \rho(a_k)) \geq E\left(\left[\rho(a_k), g^{-1}(\infty; \hat{m}_q) - \Delta\right]\right). \tag{3.3}$$

To show this, define

$$s^R_{sep2}(m) \equiv \begin{cases} \arg\max_a U^R|_{[m_q, E([t_1^2,1])-\Delta]}(\sigma^S, a) & m \in [m_q, \hat{m}_q - \Delta] \\ \arg\max_a U^R|_{[E([t_1^2,1]),1]}(\sigma^S, a) & m \in [\hat{m}_q, 1] \\ s^R(m) & \text{otherwise} \end{cases}.$$

For s^R to be weakly interim rational with respect to σ^S, it has to be the case that $g\left(\infty; \rho_{[0,1]}(a_q)\right) \geq m_q$. If $g(\infty; \rho(a_k)) < E\left(\left[\rho(a_k), g^{-1}(\infty; \hat{m}_q) - \Delta\right]\right)$, then

$$\begin{aligned} m_q &\leq g(\infty; \rho(a_k)) \\ &< E\left(\left[\rho(a_k), g^{-1}(\infty; \hat{m}_q) - \Delta\right]\right) \\ &\leq \arg\max_a U^R|_{[m_q, E([t_1^2,1])-\Delta]}(\sigma^S, a). \end{aligned}$$

Therefore s^R_{sep2} satisfies language. Since s^R_{sep2} is weakly interim rational w.r.t. σ^S by construction, we have thus reached a contradiction.

The above claim gives a lower bound on a_q. This in turn gives a lower bound on $\arg\max_a U^R|_{[0,m_q-\Delta]}(\sigma^S, a)$. Look at σ^S restricted on the interval $[0, m_q - \Delta]$. We can then apply the same argument and get a lower bound on a_{q-1}.

Define $\psi(\tau_1)$ to be the longest forward solution with an initial condition τ_1. That is, $\psi(\tau_1) \equiv \{0, \tau_1, \psi_2, \psi_3, ..., \psi_n\}$ is a forward solution on $[0, \psi_n]$ where $\psi_n \leq 1$, and there does not exist a forward solution $\left\{0, \tau_1, \psi'_2, ..., \psi'_{n'}\right\}$ where $\psi'_{n'} \leq 1$ and $n' > n$. Define $\lambda(\tau_1) \equiv n$ where n is the size of the forward solution $\psi(\tau_1)$. A necessary condition is that either $\lambda(\rho(a_k)) = 1$ or $\psi_2(\rho(a_q)) \geq t_1^2([0,1])$. So when there is a size-2 forward solution on $\left[0, t_1^2([0,1])\right]$, $q - 1 \geq 2$. A lower bound of the lower bound

on a_{q-1} can be interpreted this way by restricting types to the subset $[0, \rho(a_k) - \Delta]$. So if there is a size-2 forward solution on $\left[0, t_1^2\left(\left[0, t_1^2([0,1])\right)\right)\right]$, then $q-2 \geq 2$. Define $f_1 \equiv t_1^2([0,1])$, and $f_{j+1} \equiv t_1^2([0, f_j])$ whenever $[0, f_j]$ has a size-2 forward solution. The process ends when we reach $[0, f_l]$ where there is no size-2 forward solutions on it.

3.3 Appendix

3.3.1 Proof for Lemma 3.2

Proof. The idea is that, if s_2^R is not $\sigma^S - compatible$, then there must exist an action a_2 taken by s_2^R exactly on some interval I_2 such that a_2 does not maximize expected utility conditional on I_2. If ε is small enough, $U^R|_{I_2}\left(\varepsilon\left(\sigma^S, \sigma^{S\prime}\right), a\right)$ is very close to $U^R|_{I_2}\left(\sigma^S, a\right)$ as a function of a, then a_2 cannot maximize expected utility conditional on I_2, hence s_2^R is not $\varepsilon\left(\sigma^S, \sigma^{S\prime}\right) - compatible$ either.

u^R is bounded, so

$$\bar{D}^R \equiv \max_{(t,a),(t',a') \in T \times A} \left| u^R(t,a) - u^R\left(t', a'\right) \right|$$

is well-defined. Then for any $\left(\sigma_1^S, \sigma_1^R\right), \left(\sigma_2^S, \sigma_2^R\right) \in \Delta S_L$,

$$\begin{aligned} & \left| U^R\left(\sigma_1^S, \sigma_1^R\right) - U^R\left(\sigma_2^S, \sigma_2^R\right) \right| \\ \leq \; & \bar{D}^R \end{aligned}$$

Given an interval $I \subset M$, let $U^R|_I$ be the expected Receiver utility conditional on receiving a message in I. Let a denote both action $a \in A$ and the constant strategy which reacts to

every message with action a. Then

$$\begin{aligned}
& \left|U^R|_I\left(\sigma^S, a\right) - U^R|_I\left(\sigma_\varepsilon^S\left(\sigma^{S\prime}\right), a\right)\right| \\
= & \left|U^R|_I\left(\sigma^S, a\right) - (1-\varepsilon)\, U^R|_I\left(\sigma^S, a\right) - \varepsilon U^R|_I\left(\sigma^{S\prime}, a\right)\right| \\
= & \ \varepsilon\left|U^R|_I\left(\sigma^S, a\right) - U^R|_I\left(\sigma^{S\prime}, a\right)\right| \\
\leq & \ \varepsilon \bar{D}^R
\end{aligned}$$

The bound is does not depend on $\sigma^S, \sigma^{S\prime}, a$ or I. A^R is finite, so a best response $a \in A$ to any conjecture σ^S gives a strictly higher expected utility than any non-best response a'. Let d_{I,σ^S} denote the difference in expected utility conditional on I against conjecture σ^S between the best action and the second best action. Formally, define

$$d_{I,\sigma^S} \equiv \min_{a_2 \notin \arg\max_{a'} U^R|_I(\sigma^S, a')} \left(\left(\max_{a''} U^R|_I\left(\sigma^S, a''\right)\right) - U^R|_I\left(\sigma^S, a_2\right)\right)$$

Then $d_{I,\sigma^S} > 0$.

For all $\varepsilon < \frac{1}{2}\frac{d_{I,\sigma^S}}{\bar{D}^R}$, $a \notin \arg\max_{a'} U^R|_I\left(\sigma^S, a'\right)$ and $a^* \in \arg\max_{a'} U^R|_I\left(\sigma^S, a'\right)$,

$$\begin{aligned}
& U^R|_I\left(\varepsilon\left(\sigma^S, \sigma^{S\prime}\right), a\right) - U^R|_I\left(\varepsilon\left(\sigma^S, \sigma^{S\prime}\right), a^*\right) \\
= & \ U^R|_I\left(\varepsilon\left(\sigma^S, \sigma^{S\prime}\right), a\right) - U^R|_I\left(\sigma^S, a\right) + U^R|_I\left(\sigma^S, a\right) - U^R|_I\left(\sigma^S, a^*\right) \\
& + U^R|_I\left(\sigma^S, a^*\right) - U^R|_I\left(\varepsilon\left(\sigma^S, \sigma^{S\prime}\right), a^*\right) \\
\leq & \ \varepsilon \bar{D}^R - d_{I,\sigma^S} + \varepsilon \bar{D}^R \\
< & \ \frac{d_{I,\sigma^S}}{2} - d_{I,\sigma^S} + \frac{d_{I,\sigma^S}}{2} = 0
\end{aligned}$$

So

$$\arg\max_a U^R|_I\left(\varepsilon\left(\sigma^S,\sigma^{S\prime}\right),a\right)\subset\arg\max_a U^R|_I\left(\sigma^S,a\right) \tag{3.4}$$

Define

$$\bar{\varepsilon}_{\sigma^S}\equiv\min_{I\subset M} d_{I,\sigma^S}$$

Since M is finite, $\bar{\varepsilon}_{\sigma^S}$ is well defined. So the containment relation 3.4 holds for any $\varepsilon<\bar{\varepsilon}_{\sigma^S}$, and for any $\sigma^{S\prime}\in\Delta S^S$. If s_1^R is $\varepsilon\left(\sigma^S,\sigma^{S\prime}\right)$ for some $\varepsilon<\bar{\varepsilon}_{\sigma^S}$, then for any $\hat{m}$ which is sent by some type with strictly positive probability given the conjecture $\varepsilon\left(\sigma^S,\sigma^{S\prime}\right)$, and for the interval $I_{\hat{m}}$ on which s_1^R takes the same value as $s_1^R(\hat{m})$,

$$\begin{aligned} s_1^R(\hat{m}) &\in \arg\max_a U^R|_{I_{\hat{m}}}\left(\varepsilon\left(\sigma^S,\sigma^{S\prime}\right),a\right) \\ &\subset \arg\max_a U^R|_{I_{\hat{m}}}\left(\sigma^S,a\right) \end{aligned}$$

Since $\varepsilon\left(\sigma^S,\sigma^{S\prime}\right)(t)=(1-\varepsilon)\,\sigma^S(t)+\varepsilon\sigma^{S\prime}(t)$ for all t, any message that receives positive probability given the conjecture $\sigma^S(t)$ also receives positive probability under $\varepsilon\left(\sigma^S,\sigma^{S\prime}\right)(t)$, it is just shown that s_1^R is also σ^S-*compatible*. □

3.3.2 Proof for Proposition 3.1

Proof. Assume to the contrary there exist $m_1<m_2\in M(k-1)$ which receive different reactions under some $s^R\in S^R(k)$, i.e. $s^R(m_1)\neq s^R(m_2)$. Consider $\hat{m}_k$ being such that $\hat{m}_k$ always attains the minimum on $M(k-1)$ for any $s^R\in S^R(k)$ and that there exists $s^R\in S^R(k)$ such that $s^R(\hat{m}_k)\neq s^R(\hat{m}_k)$. Suppose $\hat{t}$ is the highest type that sends messages smaller or equal to $\hat{m}_k$. Then since $\hat{m}_k$ always takes on the minimum of s^R for any s^R in $C_R^*(k)$, the highest values $\hat{m}_k$ and $\hat{m}_k+\Delta$ can take on when $s^R(\hat{m}_k)\neq s^R(\hat{m}_k+\Delta)$ would

be $E\left(\left[0,\hat{t}_k\right]\right)$ and $E\left(\left[\hat{t}_k+\Delta,1\right]\right)$ respectively. But since there is only babbling equilibria, for every Sender type t, she prefers being thought of as pooling with all higher types than pooling with all lower types. So $\hat{t}_k$ would prefer $E\left(\left[\hat{t}_k+\Delta,1\right]\right)$ to $E\left(\left[0,\hat{t}_k\right]\right)$, where $E\left(\left[0,\hat{t}_k\right]\right)$ is the best $\hat{t}_k$ can hope for from sending message $\hat{m}_k$ (because $E\left(\left[0,\hat{t}_k\right]\right) \leq \hat{t}_k$ is on the increasing part of $\hat{t}_k$'s utility curve) and $E\left(\left[\hat{t}_k+\Delta,1\right]\right)$ is the worst $\hat{t}_k$ would anticipate from sending message $\hat{m}_k+\Delta$ when $\hat{m}_k$ induces a different action from $\hat{m}_k+\Delta$. So sending message $\hat{m}_k$ is weakly dominated by sending message $\hat{m}_k+\Delta$ for type $\hat{t}_k$. Hence in $\Pi_{t\in T}S^S\left(k+1;t\right)$, the highest type that sends messages smaller or equal to $\hat{m}_k$ would be strictly smaller than $\hat{t}_k$ and thus $\Pi_{t\in T}S^S\left(k+1;t\right) \subsetneqq \Pi_{t\in T}S^S\left(k-1;t\right)$ (in particular, $S^S\left(k+1;\hat{t}\right) \subsetneqq S^S\left(k-1;\hat{t}\right)$) and the process does not stop at round k.

Formally, define

$$\hat{m}_k := \min\left\{\begin{array}{c} m \in M\left(k-1\right)| \\ \exists s^R \in S^R\left(k\right) \text{ such that} \\ m \in \arg\min_{m'\in M(k)} s^R\left(m'\right) \\ \text{and } s^R\left(m\right) \neq s^R\left(m+\Delta\right) \end{array}\right\}$$

From the definition, $s^R\left(\hat{m}_k\right) = \min_{m'\in M(k-1)} s^R\left(m'\right)$ for all $s^R \in S^R\left(k\right)$ and there exists $\hat{s}^R \in S^R\left(k\right)$ such that $\hat{s}^R\left(\hat{m}_k\right) \neq \hat{s}^R\left(\hat{m}_k+\Delta\right)$. From weak monotonicity of $\hat{s}^R$ and the construction,

$$\begin{aligned} \hat{s}^R\left(\min M\left(k-1\right)\right) &= \min_{m'\in M(k-1)} \hat{s}^R\left(m'\right) \\ &= \hat{s}^R\left(\hat{m}_k\right) \\ &\neq \hat{s}^R\left(\hat{m}_k\right) \end{aligned}$$

That is, the interval that $\hat{s}^R$ takes on the same value as on $\hat{m}_k$ is $[\min M^*(k-1), \hat{m}_k]$. Let the interval that $\hat{s}^R$ takes on the sme value as on $\hat{m}_k + \Delta$ be $[\hat{m}_k + \Delta, \bar{m}_k]$. By the procedure, there exists $\hat{\sigma}^S \in \Pi_{t \in T}\left(\Delta^+ S^S(k-1;t)\right)$ to which $\hat{s}^R$ is $\hat{\sigma}^S - compatible$. It then follows that

$$\hat{s}(\hat{m}_k) \in \arg\max_{a \in A} U^R|_{[\min M(k-1), \hat{m}_k]}\left(\hat{\sigma}^S, a\right)$$

$$\begin{aligned} & U^R|_{[\min M(k-1), \hat{m}_k]}\left(\hat{\sigma}^S, a\right) \\ = & \sum_{s^S \in \Pi_{t \in T} S^S(k-1;t)} \hat{\sigma}^S\left(s^S\right) \sum_{\substack{t \in T: \\ s^S(t) \in [\min M(k-1), \hat{m}_k]}} \pi(t)\, u^R(t,a) \\ = & \sum_{s^S \in \Pi_{t \in T} S^S(k-1;t)} \sum_{\substack{t \leq \\ s^S(t) \in [\min M^*(k-1), \hat{m}_k]}} \hat{\sigma}^S\left(s^S\right) \end{aligned}$$

□

3.3.3 Proof for Proposition 3.2

Proposition 3.2 follows immediately from the following claim.

Claim For all k, there exists $s^R \in S^R$ such that.

1. $s^R \in ES^R(k)$, and $s^R(M(k)) = \left\{\alpha_1, ..., a_{N(b)}\right\}$ where $\alpha_i = E([t_{i-1}, t_i - \Delta])$;
2. $\forall m \in [\alpha_i, \alpha_{i+1} - \Delta]$, either there exists $m' < m$ such that $u^S\left(t_i - \Delta, s^R(m)\right) \leq u^S\left(t_i - \Delta, s^R(m')\right)$ for all $s^R \in ES^R(k)$, or $s^R(m) = \alpha_{i-1}$.

Proof Show by induction. Suppose they hold for k. Then there exists $\hat{s}^R \in ES^R(k)$ satisfying condition 1 and 2. From the definition that $\left\{t_0, ..., t_{N(b)}\right\}$ is a forward

solution and that $\alpha_i = E([t_{i-1}, t_i - \Delta])\ \forall i = 1, ..., N(b)$, every type $t \in [t_{i-1}, t_i - \Delta]$ strictly prefers action α_i the most in the range of $\hat{s}^R$. Therefore, there exists one message m such that $\hat{s}^R(m) = \alpha_i$ and $m \in ES^S(k+1; t)$. Since $\hat{s}^R(\alpha_{i+1}) = \alpha_{i+1} > \alpha_i$, such message must be smaller than $\alpha_{i+1} - \Delta$. Thus $l(k+1; t) \leq \alpha_{i+1} - \Delta$ for all $t \in [t_{i-1}, t_i - \Delta]$. Therefore, we can define

$$s^S_{big}(t) \equiv \max \left\{ \begin{array}{c} m \in ES^S(k+1; t), m \leq \alpha_{i+1} - \Delta \\ \text{where } i \text{ is such that } t \in [t_{i-1}, t_i - \Delta] \end{array} \right\} \forall t.$$

By definition, $s^S_{big} \in ES^S(k+1)$, and thus $s^S_{big} \in ES^S(k)$.

Claim s^S_{big} is increasing in t.

Proof Given $\hat{t}$. Let i be such that $\hat{t} + \Delta \in [t_{i-1}, t_i - \Delta]$. To show that $s^S_{big}(\hat{t} + \Delta) \geq s^S_{big}(\hat{t})$, it suffices to show that

$$\left[s^S_{big}(\hat{t}), \alpha_{i+1} - \Delta\right] \cap ES^S(\hat{t} + \Delta; k) \neq \emptyset.$$

We break the discussion into two cases.

Case 1 $\hat{s}^R\left(s^S_{big}(\hat{t})\right) \leq \alpha_i - \Delta$. Then $\left(\hat{s}^R\right)^{-1}(\alpha_i) \subset \left[s^S_{big}(\hat{t}) + \Delta, \alpha_{i+1} - \Delta\right]$ by the construction of $\hat{s}^R$ and the assumption that $\hat{s}^R\left(s^S_{big}(\hat{t})\right) \leq \alpha_i - \Delta$. Since

$$\left(\hat{s}^R\right)^{-1}(\alpha_i) \cap ES^S(k; \hat{t} + \Delta) \neq \emptyset,$$

we know that

$$\left[s^S_{big}(\hat{t}) + \Delta, \alpha_{i+1} - \Delta\right] \cap ES^S(k; \hat{t} + \Delta) \neq \emptyset.$$

Case 2 $\hat{s}^R\left(s^S_{big}\left(\hat{t}\right)\right) \geq \alpha_i$. Then $\hat{s}^R\left(s^S_{big}\left(\hat{t}\right)\right) = \alpha_i$, because $\hat{t} \leq t_i - \Delta$ and by construction of s^S_{big}, $s^S_{big}\left(\hat{t}\right) \leq \alpha_{i+1} - \Delta$, and the assumption that $\hat{s}^R$ satisfies condition 2. Let $\tilde{\sigma}^R \in \Delta^+ ES^R\left(k-1\right)$ such that

$$s^S_{big}\left(\hat{t}\right) \in \arg\max_m U^S\left(\hat{t}, \tilde{\sigma}^R\left(m\right)\right).$$

(Existence is guaranteed by the definition of $s^S_{big}\left(\hat{t}\right)$) Then by super modularity of U^S and weak monotonicity of s^R in the support of $\tilde{\sigma}^R$,

$$U^S\left(\hat{t}+\Delta, \tilde{\sigma}^R\left(s^S_{big}\left(\hat{t}\right)\right)\right) > U^S\left(\hat{t}+\Delta, \tilde{\sigma}^R\left(m\right)\right)$$

for all $m < s^S_{big}\left(\hat{t}\right)$. Since $\hat{s}^R\left(s^S_{big}\left(\hat{t}\right)\right) = \alpha_i$, we know that $s^S_{big}\left(\hat{t}\right) \in \arg\max_m U^S\left(\hat{t}+\Delta, \hat{s}^R\left(m\right)\right)$. Therefore, for ε very small,

$$\begin{aligned} & \arg\max_m U^S\left(\hat{t}+\Delta, \left(\left(1-\varepsilon\right)\hat{s}^R + \varepsilon\tilde{\sigma}^R\right)\left(m\right)\right) \\ \subset \quad & \left[s^S_{big}\left(\hat{t}\right), \alpha_{i+1} - \Delta\right]. \end{aligned}$$

Since $\left(1-\varepsilon\right)\hat{s}^R + \varepsilon\tilde{\sigma}^R$ belongs to $\Delta^+ ES^R\left(k-1\right)$,

$$\begin{aligned} & \arg\max_m U^S\left(\hat{t}+\Delta, \left(\left(1-\varepsilon\right)\hat{s}^R + \varepsilon\tilde{\sigma}^R\right)\left(m\right)\right) \cap ES^S\left(k; \hat{t}+\Delta\right) \\ \neq \quad & \emptyset \end{aligned}$$

and thus

$$\left[s^S_{big}\left(\hat{t}\right), \alpha_{i+1} - \Delta\right] \cap ES^S\left(k; \hat{t}+\Delta\right) \neq \emptyset.$$

Lemma 3.2 implies that $ES^R\left(k+1\right)$ must contain one Receiver strategy that is

weakly sequentially rational with respect to s^S_{big}. Suppose

$$s^R_{big} \in \arg \max_{\substack{s^R \in S^R \\ s^R \text{ is interim rational} \\ \text{w.r.t. } s^S_{big}}} U^R\left(s^S_{big}, s^R\right),$$

then $s^R_{big}\left(s^S_{big}(t)\right)$ is increasing in t because s^R_{big} is increasing and $s^S_{big}(t)$ is increasing in t. Therefore, s^R_{big} partitions the type space into $\{\tau_0, ..., \tau_n\}$ where $\tau_0 = 0$ and $\tau_n = 1$. By definition,

$$s^R_{big}\left(s^S_{big}(t)\right) = s^R_{big}\left(s^S_{big}\left(t'\right)\right)$$

if and only if t and t' both belong to the same step $[\tau_{i-1}, \tau_i - \Delta]$ for some i and

$$s^R_{big}\left(s^S_{big}(\tau_{i-1})\right) = E\left([\tau_{i-1}, \tau_i - \Delta]\right)$$

for $i = 1, ..., n$.

Claim $[0, \tau_{i+1} - \Delta]$ has a forward solution of size $i+1$ and $\tau_i \leq t_i^{i+1}\left([0, \tau_{i+1} - \Delta]\right)$ for $i = 1, ..., n$.

Proof Show by induction. Suppose $[0, \tau_{j+1} - \Delta]$ has a forward solution of size $j+1$ and $\tau_j \leq t_j^{j+1}\left([0, \tau_{j+1} - \Delta]\right)$ for all $j = 1, ..., i-1$. Condition (M) implies that $\tau_{j+1} > t_j$ for all $j = 1, ..., i-1$ because $\left\{t_0, ..., t_{N(b)}\right\}$ is the largest forward solution on $[0, 1]$. First we want to show that type $\tau_i - \Delta$ must weakly prefer action $E\left([\tau_{i-1}, \tau_i - \Delta]\right)$ to action $E\left([\tau_i, \tau_{i+1} - \Delta]\right)$.

Case 1 $\tau_i \neq t_q$ for any q.

Therefore, there exists q such that $\tau_i - \Delta, \tau_i \in [t_{q-1}, t_q - \Delta]$. By construc-

tion, $s^S_{big}(\tau_i) < \alpha_{q+1} - \Delta$. By the construction of $\hat{s}^R$, $\hat{s}^R\left(s^S_{big}(\tau_i)\right) = \alpha_q$. Suppose to the contrary that type $\tau_i - \Delta$ prefers action $E([\tau_i, \tau_{i+1} - \Delta])$ to action $E([\tau_{i-1}, \tau_i - \Delta])$. By the definition of $\{\tau_0, \tau_1, ...\tau_n\}$ and the construction that s^R_{big} is sequentially rational w.r.t. s^S_{big}, we know that

$$s^R_{big}\left(s^S_{big}(\tau_i)\right) = E([\tau_i, \tau_{i+1} - \Delta])$$

and

$$s^R_{big}\left(s^S_{big}(\tau_i - \Delta)\right) = E([\tau_{i-1}, \tau_i - \Delta]).$$

Therefore, given the Receiver strategy s^R_{big}, type $\tau_i - \Delta$ prefers message $s^S_{big}(\tau_i)$ to message s^S_{big}, and

$$\begin{aligned} & \arg\max_m U^S\left(\tau_i - \Delta, \left((1-\varepsilon)\hat{s}^R + \varepsilon s^R_{big}\right)(m)\right) \\ \subset \quad & \left[s^S_{big}(\tau_i - \Delta) + \Delta, \alpha_{i+1} - \Delta\right]. \end{aligned}$$

Since $s^R_{big} \in ES^R(k+1) \subset ES^R(k)$,

$$\begin{aligned} & \left[s^S_{big}(\tau_i - \Delta) + \Delta, \alpha_{i+1} - \Delta\right] \cap ES^S(k+1; \tau_i - \Delta) \\ \neq \quad & \emptyset. \end{aligned}$$

But this contradicts the construction of $s^S_{big}(\tau_i - \Delta)$.

Case 2 $\tau_i = t_q$ for some q.

We've shown that $\tau_i > t_{i-1}$. So $q \geq i$. Suppose $q > i$. But then s^R_{big} can be improved upon by partitioning $[0, \tau_i]$ as $\{0, t_1, ..., t_q\}$ by the monotonic-

ity condition (M), and there exists a Receiver strategy that is interim rational w.r.t. s^S_{big} which does that partition. So $q = i$. But then by the same argument, $\tau_j = t_j$ for all $j < i$. In particular, $\tau_{i-1} = t_{i-1}$. Suppose to the contrary, type $t_i - \Delta$ prefers action $E([t_i, \tau_{i+1} - \Delta])$ to action $E([t_{i-1}, t_i - \Delta])$. Then it has to be the case that $E([t_i, \tau_{i+1} - \Delta]) < E([t_i, t_{i+1} - \Delta]) = \alpha_{i+1}$. By the literal condition of the language assumption, $s^R_{big}(E([t_i, \tau_{i+1} - \Delta])) = E([t_i, \tau_{i+1} - \Delta])$. Therefore, given s^R_{big}, type $t_i - \Delta$ prefers message $E([t_i, \tau_{i+1} - \Delta])$ to message $s^S_{big}(t_i - \Delta)$. So there exists some message $m \geq E([t_i, \tau_{i+1} - \Delta])$ such that $m \in ES^S(k; t_i - \Delta)$. Since $E([t_i, \tau_{i+1} - \Delta]) \leq \alpha_{i+1} - \Delta$, from the assumption that condition 2 holds for k, $\hat{s}^R(E([t_i, \tau_{i+1} - \Delta])) = \alpha_i$. So

$$\begin{aligned} & \arg\max_m U^S\left(t_i - \Delta, \left(\left(1-\varepsilon\right)\hat{s}^R + \varepsilon s^R_{big}\right)(m)\right) \\ \subset \quad & \left[s^S_{big}(t_i - \Delta) + \Delta, \alpha_{i+1} - \Delta\right]. \end{aligned}$$

And it follows that

$$\begin{aligned} & \left[s^S_{big}(t_i - \Delta) + \Delta, \alpha_{i+1} - \Delta\right] \cap ES^S(k+1; \tau_i - \Delta) \\ \neq \quad & \emptyset. \end{aligned}$$

A contradiction.

By assumption, $\tau_{i-1} \leq t^i_{i-1}([0, \tau_i - \Delta])$. So $E([\tau_{i-1}, \tau_i - \Delta]) \leq \alpha^i_i([0, \tau_i - \Delta])$. We have just shown that type $\tau_i - \Delta$ prefers $E([\tau_{i-1}, \tau_i - \Delta])$ to $E([\tau_i, \tau_{i+1} - \Delta])$. Thus, type $\tau_i - \Delta$ must prefer action $\alpha^i_i([0, \tau_i - \Delta])$ to action $E([\tau_i, \tau_{i+1} - \Delta])$

because

$$E\left(\left[\tau_{i-1},\tau_i-\Delta\right]\right)\leq\alpha_i^i\left(\left[0,\tau_i-\Delta\right]\right)<\tau_i\leq E\left(\left[\tau_i,\tau_{i+1}-\Delta\right]\right).$$

So there exists $\bar{t}\in[\tau_i,\tau_{i+1}-\Delta]$ such that type $\tau_i-\Delta$ prefers action $\alpha_i^i\left(\left[0,\tau_i-\Delta\right]\right)$ to action $E\left(\left[\tau_i,\bar{t}\right]\right)$ and type τ_i prefers action $E\left(\left[\tau_i,\bar{t}\right]\right)$ to action $\alpha_i^i\left(\left[0,\tau_i-\Delta\right]\right)$. By definition of $\bar{t}$, $\tau_i=t_i^{i+1}\left(\left[0,\bar{t}\right]\right)$. By the monotonicity condition (M),

$$t_i^{i+1}\left(\left[0,\bar{t}\right]\right)\leq t_i^{i+1}\left(\left[0,\tau_{i+1}-\Delta\right]\right)$$

because $\bar{t}\leq\tau_{i+1}-\Delta$. It follows that $\tau_i\leq t_i^{i+1}\left(\left[0,\tau_{i+1}-\Delta\right]\right)$. Moreover, $[0,\bar{t}]$ has a forward solution of size $i+1$, so $[0,\tau_{i+1}]$ has a forward solution of size $i+1$.

Claim 3.3.3 implies that

$$\tau_{n-1}\leq t_{n-1}^n\left(\left[0,\tau_n-\Delta\right]\right)=t_{n-1}^n\left(\left[0,1\right]\right)$$

and that $[0,1]$ has a forward solution of size n. Since $N(b)$ is the maximum of the size of a forward solution on $[0,1]$, $n\leq N(b)$. So $\tau_i\leq t_i$ for all $i=1,...,n$. Condition (M) implies that

$$U^R\left(s_{big}^S,s_{big}^R\right)\leq U^R\left(s_{big}^S,\hat{s}^R\right)$$

because $\hat{s}^R\left(s_{big}^S\right)$ partitions the type space into $\left\{0,t_1,...,t_{N(b)-1},1\right\}$ and this is a better partition. But by assumption, $\hat{s}^R\in ES^R(k)$ where $\hat{s}^R\left(M\left(k\right)\right)=$

$\left\{\alpha_1, ..., a_{N(b)}\right\}$ and $\alpha_i = E\left([t_{i-1}, t_i - \Delta]\right)$ for $i = 1, ..., N(b)$. $\hat{s}^R$ is weakly interim rational with respect to s^S_{right}, so

$$\max_{\substack{s^R \in ES^R(k); \\ s^R \text{ is weakly interim rational} \\ \text{w.r.t. } s^S_{big}}} U^R\left(s^S_{big}, s^R\right) \geq U^R\left(s^S_{big}, \hat{s}^R\right).$$

Therefore, equality holds and for any $\tilde{s}^R \in ES^R(k+1)$ such that $\tilde{s}^R$ is weakly sequentially rational w.r.t. s^S_{big}, $\tilde{s}^R$ partitions the type space into $\left\{0, t_1, ..., t_{N(b)-1}, 1\right\}$.

Suppose $m \in [\alpha_i, \alpha_{i+1} - \Delta]$ is such that there exists $\tilde{s}^R$ belongs to $ES^R(k+1)$ where

$$u^S\left(t_i - \Delta, \tilde{s}^R(m)\right) > u^S\left(t_i - \Delta, \tilde{s}^R(m - \Delta)\right).$$

Since statement 2 holds for k, $\hat{s}^R(m) = \alpha_i$. So

$$\arg\max_m U^S\left(t_i - \Delta, \left((1-\varepsilon)\hat{s}^R + \tilde{s}^R\right)(m)\right) \subset [m, \alpha_{i+1} - \Delta].$$

Therefore,

$$ES^S(k+1; t_i - \Delta) \cap [m, \alpha_{i+1} - \Delta] \neq \emptyset.$$

It follows that $s^S_{big}(t_i - \Delta) \geq m$ and therefore

$$s^R_{big}(m) \leq s^R_{big}\left(s^S_{big}(t_i - \Delta)\right) = \alpha_i.$$

We have thus shown that statement 2 holds for $k+1$.

Chapter 4

Coordination Games

4.1 Introduction

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

The difficulty in formalizing the credibility criterion lies in how to incorporate the strategy of the hypothetical speaker who intends to *not* carry out her statement into the analysis . Baliga and Morris (2002) tackle this problem by expanding the original game into one

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

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

In this expanded game, an instruction is actually a recommendation for the Receiver to

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

We focus on stage games where the best response correspondences are functions. So, the stage game is self-signaling when the Sender always prefers the Receiver to take his best response, and the stage game is self-committing if the Receiver should take an action whenever one is recommended, given that the Sender believes that the recommendation will be followed. With these definitions in mind, we find that if the stage game is self-committing and strongly self-signalling, there is a unique iterative admissible outcome of the language

game which gives the Sender her Stackelberg payoff. On the other hand, if the stage game is self-committing, but the Sender's preference over the Receiver's actions does not depend her own action, every rationalizable action profile is the outcome of an iterative admissible strategy profile in the language game.

The rest of this chapter is structured as follows. Section 4.2 provides three simple examples to illustrate the role of the self-signalling condition and to motivate the language assumption. Section 4.3 describes the model and the language assumptions. Section 4.4 presents our main results described above. Section 4.5 briefly reviews the main results in Baliga and Morris (2002) and compares theirs with ours. Section 4.6 concludes.

4.2 Motivating Examples

The main idea of this chapter is best understood through examples. The battle-of-the-sex game example in section 4.2.1 illustrates that self-signalling is sufficient to guarantee Stackleberg payoff for the speaker. The investment game example in section 4.2.2 shows that a severe violation of the self-signalling criterion makes communication ineffective. The partial-common-interest game in section 4.2.3 motivates the hierarchical messages and language assumptions formally described in section 4.3.1.

4.2.1 Coordination without positive spillovers

In the Battle-of-the-Sexes game in table 4.2.1, there are two Nash equilibria: both go to the Opera and both go to the Club. The Sender prefers the first equilibrium and the Receiver prefers the second. The promise "I will go to the opera" is self-committing because if the Sender believes that the Receiver will believe this statement and play his best response

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

Table 4.1: Battle of Sex Game

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

Table 4.2: Receiver's Strategies in Battle-of-the-Sex Game

Opera, the Sender would prefer to go to the *Opera* and carry out her promise. The promise is self-signalling as well because had the Sender not intended to go to the *Opera*, i.e. had she intended to go to the *Club*, she would prefer the Receiver to go to the *Club* instead of the *Opera* and hence she would not want the Receiver to believe the promise "I will go to the *Opera*".

Suppose $M = \{\text{“}opera\text{”}, \text{“}club\text{”}\}$. It can be interpreted as a promise to carry out a certain action, or a recommended action for the Receiver. The Sender sends a message $m \in M$, and then plays an action $a^S \in A^S$ in the stage game. The set of strategies for the Sender is thus

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

while the set of strategies for the Receiver is listed in table 4.2.1. As in the motivating example of chapter 1, both the *Always Opera* and *Always Club* strategies ignore the messages completely. *Literal* strategy and *Opposite* strategy both respond to a message by going to the *Opera* and the other message by going to the *Club*, and hence are essentially the same up to their renaming.

In addition, We can see from table 4.2.1 that message "*opera*" and message "*club*" are complete symmetric in the sense that if we swap the names of these two messages, we end up with exactly the same strategy set S^R as in table 4.2.1. This should not be surprising because in traditional economic models of communication, messages have no inherent meanings — the meaning is determined by the equilibrium.

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

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

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

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

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

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

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

4.2.2 Coordination with positive spillovers

To understand the role of the self-signalling criterion, let's look at the Investment game in figure 4.2.2. As in the Battle-of-the-Sexes game, there are two Nash equilibria in this game: $(Invest, Invest)$ and (Not, Not). The promise "I'm going to invest" is self-committing

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

Table 4.3: Investment Game

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

Table 4.4: Receiver's Strategies in Investment Game

because if the Sender believes that the Receiver is going to believe the statement and play his best response, it is optimal for the Sender to carry out the promise and play the strategy *Invest*. In Farrell's point of view, this message is thus credible and should be believed. Aumann argues that this promise is not self-signalling and hence is not credible. Even if the Sender intends to play *Not*, possibly due to lack of confidence that Receiver is really going to *Invest*, she still prefers the Receiver to use the strategy *Invest*. Therefore, she would like the Receiver to believe her promise regardless of her intended action. If she is pessimistic about the effect of communication and believes that, with high probability, the Receiver is going to *Not Invest* regardless of what she says, then she would prefer to *Not Invest*. However, if the probability that the Receiver uses the strategy *Invest* is higher after hearing the promise "I'm going to invest", the Sender would like to make that promise even though she does not intend to carry it out.

Let's look at the cheap talk extension game in detail. Suppose $M = \{\text{“}invest\text{”}, \text{“}not\text{”}\}$. Then the set of strategies for the Sender is

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

while the set of strategies for the Receiver is listed in table 4.2.2. Suppose it is common knowledge that the Receiver follows the language convention and never uses the strategy *Opposite*. In the transformed game G_L, the set of strategies for the Receiver is thus

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

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

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

the role of the self-signalling criterion.

4.2.3 Partial Common Interest

The Fighting-Couple game in table 4.5 shows that communication can help players avoid bad equilibria, even though their preferences are not fully aligned. The game has one pure strategy equilibrium: $(Home, Home)$. In one of the two mixed strategy equilibria, both go to the *Opera* with probability $\frac{1}{2}$ and go to the *Club* with probability $\frac{1}{2}$. In the other mixed strategy equilibrium, both go to the *Opera* with probability $\frac{1}{8}$, go to the *Club* with probability $\frac{1}{8}$ and stay *Home* with probability $\frac{3}{4}$. The mixed-strategy equilibrium where both staying *Home* with probability 0 is the efficient one. Both going to the *Opera* and going to the *Club* is consistent with going out, as opposed to staying home. One prefers to stay *Home* if and only if the other stays *Home*. Moreover, avoiding staying *Home* and restricting themselves to the submatrix $\{Opera, Club\} \times \{Opera, Club\}$ is mutually beneficial for both players.

Now suppose the Sender has an opportunity to leave a voice message before they play the one-shot game in table 4.5. She cannot possibly persuade the Receiver to go to the *Opera*, nor can she persuade the Receiver to go to the *Club*, because they have conflict of interest regarding the two actions. However, it is self-committing for her to say "you should go out," in the sense that if the Receiver is persuaded and goes out, the Sender will go out, i.e., she will choose to either go to the *Opera* or go to the *Club*, in which case, the Receiver prefers to go out. In addition, the suggestion "you should go out" is also self-signalling in the sense that the Sender prefers the Receiver to go out only if she plans to go out herself.

Consider the suggestion "Definitely go out tonight, dear. Regarding where to go, you

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

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

Then the preceding discussion suggests the type of language assumption that restricts the Receiver's strategies to those in table 4.6.

We will now show that, in the language game G_L, no player uses the action $Home$ after three rounds of deletion of weakly dominated strategies. In the first round of deletion, the strategy ("$\{Opera, Club\}\{Opera\}$", $Home$) is weakly dominated by ("$\{Home\}$", $Home$) for the Sender. And both ("$\{Home\}$", $Opera$) and ("$\{Opera, Club\}\,\{Opera\}$", $Opera$) are weakly dominated by ("$\{Opera, Club\}\,\{Club\}$", $Opera$) for the Sender. Therefore, after the first round of deletion of weakly dominated strategies, if the Sender suggests any non-violent action (either $Opera$ or $Club$), she definitely plans to not fight; if the Sender

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

Table 4.5: Fighting-Couple Game

suggests to fight, she definitely plans to fight. Thus, it is weakly dominated for the Receiver to fight after a non-violent suggestion; it is also weakly dominated for the Receiver to play a non-violent action after the suggestion to fight. However, it is also weakly dominated in the second round of deletion for the Receiver to play the *Completely Literal* strategy, because they have opposing interests when restricting the game to $\{Opera, Club\} \times \{Opera, Club\}$. It can be easily checked that the set of strategies that survive the second round of deletion of weakly dominated strategies is thus$\{Opera\ Home, Club\ Home\}$. Therefore, the Sender can be guaranteed a non-violent response if she says either "$\{Opera, Club\}\ \{Opera\}$" or "$\{Opera, club\}\ \{Club\}$". Since she prefer any outcome in the submatrix $\{Opera, Club\} \times \{Opera, Club\}$ to anyone outside of that matrix, the strategy ("$\{Fight\}$", $Fight$) is strictly dominated in the third round.

The strategy set for the Sender that survives iterative admissibility is

$$\{(\text{“}\{Opera, Club\}\ \{Opera\}\text{”}, Club)\,, (\text{“}\{Opera, Club\}\ \{Club\}\text{”}, Opera)\}\,,$$

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

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

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

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

Table 4.6: Receiver's Strategies in the Fighting Couple Game

4.3 The Model

In this chapter, we apply the general framework described in chapter one to one-sided communication in finite two-player games with complete information. The Sender (S) and the Receiver (R) simultaneously choose an action a^S, a^R from a finite set A^S and A^R respectively. Their payoffs are given by $g^S : A^S \times A^R \to R$ and $g^R : A^S \times A^R \to R$ respectively. Write $g = \left(g^S, g^R\right)$. We will abuse the notation and denote the stage game also by g. In the one-sided cheap talk extension game G, the Sender gets to send a message from a finite set M before they play the stage game g. A strategy for the Sender in the reduced-form cheap talk extension game G, denoted by s^S, is a message $m \in M$ and an action $a^S \in A^S$. A strategy for the Receiver in G, denoted by s^R, is a mapping from M to A^R. To apply the general framework, we first transform the cheap talk game G into the language game G_L by directly restricting the set of strategies for the Receiver. We need to modify the language assumptions in chapter one because in the class of games we deal with here, there is no natural order on the set of actions. After the language assumptions are laid out, we apply normal form iterative admissibility to the language game G_L.

In this chapter, we will focus on games where changing exactly one player's action in the action profile changes the payoff. That is, games such that $g^i\left(a^S, a^R\right) \neq g^i\left(a'^S, a^R\right)$ and

$g^i\left(a^S, a^R\right) \neq g^i\left(a^S, a'^R\right)$ for every $a^S \neq a'^S$, $a^R \neq a'^R$, and $i = S, R$. This implies that, in particular, the best response correspondences for both players are well-defined functions. This condition is weaker than genericity, which is a common assumption, and does not exclude any of the motivating games in section 4.2.

4.3.1 Incorporating Language

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

Some may argue that the Receiver would want to take the *Opposite* strategy in a matching penny game i.e.: the strategy that takes the action opposite to its meaning according to L. However, if the Sender knows the payoff structure of a matching penny game, and if she knows that the Receiver uses language L' in a matching penny game, she will give

recommendations according to L', thereby destroying the incentive for the Receiver to use language L'. In this case, one may argue that the Receiver randomizes his actions after receiving a message. This could be achieved by randomizing between the *Always B* strategy, and the *Never B* strategy.

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

In the Battle-of-the-Sexes game, if we let B refer to going to the opera, then "not B," i.e. "not go to the opera," is equivalent to "go to the club," since this is the only choice other than going to the opera. In the Fighting-Couple game, the expression, "go out", is saying exactly the same thing as, "go to the opera or go to the club", and the expression, "do not go out", says the same thing as, "go home". If the Receiver responds differently to the two recommendations, "go out", and, "go home", then we see from previous discussion that he responds to, "go out", by going out. However, the Receiver still has to decide whether to go to the opera or the club. Carrying this idea forward, lets suppose that

the subset of Receiver actions B contains a strict subset B_2, and the language L contains an expression for B_2. Suppose further that L contains an expression ξ_{00} that is simply a concatenation of ξ_0 and the expression for B_2. Then with the expression for logical negation, L contains an expression ξ_{01} which is the concatenation of ξ_0 meaning ,"do not do B", and the expression for, "within B_1, do not do B_2". That the receiver may decide that the messages, "go out", and, "go home", convey separate information, but decide to ignore the finer differences between, "go out; furthermore, go to the opera", and "go out; and then go to the club'. Then the Receiver takes the same action after receiving both message ξ_{00} and ξ_{01}. However, if the Receiver decides not to ignore the finer difference between the recommendations ξ_{00} and ξ_{01}, he refers to his knowledge of L and responds to message ξ_{00} with action B_2 and to message ξ_{01} with an action in B_1 but not in B_2.

Let M denote every message that the Sender could possibly utter. We also assume that the language L the players commonly speak contains an expression for every subset of Receiver actions, an expression for logical negation, and an expression for concatenation. Then, the language L contains an expression for every strictly decreasing sequence of subsets of Receiver actions $A_1A_2...A_n$. As a convention, let $A_0 = A^R$ and $A_{n+1} = \emptyset$. Each sequence can be seen as a sequence of instructions with finer and finer details. The set of all such sequences where the last subset has only one element is called the set of hierarchical recommendations, denoted by M^h. Given a hierarchical recommendation $m = A_1...A_n$, we call A_j the j^{th} level of instruction. Define $M(A_1...A_j)$ to be the set of all messages that start with the strictly decreasing sequence $A_1...A_j$. Every message m in $M(A_1...A_j)$ express the same idea of "Do A_1. Further more, take an action in A_2. ....To be even more

precise, do A_j".

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

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

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

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

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

$a^R \in A^R$, define

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

Now we formally define our language assumptions.

Definition 4.1. *$s^R : M \to A$ is a language-based Receiver strategy, denoted by $s^R \in S_L^R$, if and only if s^R is constant on $M^{cstr}\left(m, s^R\left(m\right)\right)$, for every $m \in M$.*

This definition is best illustrated with graphs. Suppose $A^R = \{A, B, C, D\}$. Figure 4.2 shows some hierarchical recommendations in this game. There are many different ways to group A^R. The first level of instruction can be about taking action D or not taking action D, as shown by the two branches $\{D\}$ and $\{A, B, C\}$ that diverge from each other. The first layer of instruction can also tell you whether to take actions in $\{A, C\}$ or not, as shown by the two branches $\{A, C\}$ and $\{B, D\}$. Given a first-layer instruction $\{A, B, C\}$, the second layer of instruction could be about whether to take action A or not, as shown by the two branches $\{A\}$ and $\{B, C\}$ that diverge one node on the branch of $\{A, B, C\}$. In general , expressions that are "opposite to" each other at some level of instruction are drawn to diverge from the same node. We call all the messages that diverge from the same node a message bundle. For example, all the messages in the circle in figure 4.2 constitutes one message bundle. There can be several parallel message bundles on a branch, which represent different ways to subdivide the set of Receiver actions relevant for the branch.

Suppose we choose the branch of $\{A, B, C\}$ and then choose $\{B\}$, we end up with the message "$\{A, B, C\}\{B\}$." The set $M\left(\{A, B, C\}\{A, C\}\right)$ consists of two messages:

"$\{A, B, C\}\{A, C\}\{A\}$" and "$\{A, B, C\}\{A, C\}\{C\}$." Within the broad instruction $\{A, B, C\}$, these two messages are both "opposite to" message "$\{A, B, C\}\{B\}$."

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

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

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

Alternatively, we could start by defining the set of messages that are "in between" two different messages, and define language by literal meaning condition and convexity condition as in the first chapter. Lemma 4.1 shows that these two approaches are equivalent.

Suppose m_A and m_B are two different hierarchical recommendations. They can be represented on a tree as in figure 4.2. We can trace the messages up along the branches they come from and find the first common branch that they both belong to. For example, message "$\{D\}$" and message "$\{A, B, C\}\{A, C\}\{C\}$" belong to the main branch (or rather, the trunk) $\{A, B, C, D\}$. Furthermore, they belong to two subbranches that diverge from the same node on the branch of $\{A, B, C, D\}$. We call the set of all message that belong to either of the two subbranches the set of messages "in between" message m_A and m_B. It is not always the case that two messages belong to two different subbranches that diverge from the same node. If they belong to two different message bundles on a branch, as

message "$\{A, C, D\}\{A, C\}\{C\}$" and "$\{A, C\}\{A\}$" do, or if one of the two messages is not a hierarchical recommendation, we say that the only messages "in between" m_A and m_B are the messages m_A and m_B themselves.

Let $Conv(m_A, m_B)$ denote the set of messages that are "in between" message m_A and message m_B. It is formally defined as follows.

Definition 4.2. *Let m_A and m_B be two messages in M. If m_A and m_B are both hierarchical recommendations, written as $A_1...A_{n_A}$ and $B_1...B_{n_B}$ respectively, and if there exists a positive integer λ such that $A_j = B_j$ for every level of instruction $j = 1, ..., \lambda - 1$ and $B_\lambda = A_{\lambda-1} \backslash A_\lambda$, then define*

$$Conv(m_A, m_B) = M(A_1...A_\lambda) \cup M(B_1...B_\lambda);$$

otherwise, define

$$conv(m_A, m_B) = \{m_A, m_B\}.$$

A pair of messages are more different from each other than the other pair if the convex hull of the previous pair contains the convex hull of the latter pair. For example, in figure 4.2, the messages "$\{A, B, C\}\{B\}$" and "$\{A, B, C\}\{A, C\}\{A\}$" are "in between" the messages "$\{B\}$" and "$\{A, B, C\}\{A, C\}\{C\}$".

Lemma 4.1. *$s^R : M \to A^R$ belongs to S_L^R iff it satisfies the following two conditions:*

1. *(Literal Meaning) If $s^R(\hat{m}) = \hat{a}^R$ for some $\hat{m} \in M$, then $s^R(\tilde{m}) = \hat{a}^R$ for every $\tilde{m} \in M^{cstr}(\hat{m}, \hat{a}^R)$ where $l(\hat{m}) = \hat{a}^R$;*

2. *(Convexity) if $s^R(m_A) = s^R(m_B)$, then $s^R(m) = s^R(m_A)$ for every m "in between"*

m_A and m_B, that is, for every m in $Conv(m_A, m_B)$.

Proof. We first prove the "only if" part. If s^R is a language-based Receiver strategy, then $s^R|_{M^{cstr}(m,s^R(m))} = s^R(m)$ for every $m \in M$. The literal meaning condition is thus easily satisfied. Suppose $s^R(m_A) = s^R(m_B)$. If $conv(m_A, m_B) = \{m_A, m_B\}$, then the convexity condition trivially holds. Otherwise, write $m_A = A_1...A_{n_A}$ and $m_B = B_1...B_{n_B}$, and let λ be such that $A_j = B_j$ for $j = 1, ..., \lambda - 1$ and $B_\lambda = A_{\lambda-1} \backslash B_\lambda$. Then either $s^R(m_A) \in A_\lambda$ or $s^R(m_A) \in B_\lambda = A_{\lambda-1} \backslash A_\lambda$. W.l.o.g. assume $s^R(m_A) \in B_\lambda$. Therefore,

$$\begin{aligned} M^{cstr}\left(m_A, s^R(m_A)\right) &= M(A_1...A_\lambda) \cup M(A_1...A_{\lambda-1}(A_{\lambda-1} \backslash A_\lambda)) \\ &= conv(m_A, m_B), \end{aligned}$$

and $s^R(m') = s^R(m_A)$ for every $m' \in conv(m_A, m_B)$ by the language assumptions.

Now we prove the "if" part. Given a message $m \notin M^h$ or $m \in M^h$ but $l(m) = s^R(m)$, then $M^{cstr}\left(m, s^R(m)\right) = \{m\}$, so $s^R|_{M^{cstr}(m,s^R(m))} = s^R(m)$. If $m \in M^h$ and $l(m) \neq s^R(m)$, then write $m = A_1...A_jA_{j+1}..A_n$ where $s^R(m) \in A_j \backslash A_{j+1}$. Let $\tilde{m} = A_1...A_j(A_j \backslash A_{j+1})\left\{s^R(m)\right\}$. Then $\tilde{m} \in M^{cstr}\left(m, s^R(m)\right)$ and $l(\tilde{m}) = s^R(m)$. By the literal meaning condition, $s^R(\tilde{m}) = s^R(m)$. Since message m and $\tilde{m}$ diverge from the same node on the branch of $A_1...A_j$,

$$\begin{aligned} conv(m, \tilde{m}) &= M(A_1...A_jA_{j+1}) \cup M(A_1...A_j(A_j \backslash A_{j+1})) \\ &= M^{cstr}\left(m, s^R(m)\right). \end{aligned}$$

The convexity condition thus implies that $s^R|_{M^{cstr}(m,s^R(m))} = s^R(m)$. □

4.4 Results

We can generalize the intuition gained from the contrast between the Battle-of-the-Sex game and the Investment game as follows. Section 4.4.1 gives sufficient conditions for one-sided pre-game communication to guarantee coordinated play in a coordination game. Section 4.4.2 shows that, when the Sender's preference over the Receiver's actions is independent of the action she takes, every rationalizable outcome in the stage game is possible.

4.4.1 A Sufficient Condition to Guarantee Stackelberg Payoff for the Sender

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

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

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

Definition 4.3 (Baliga and Morris (2002)). *Claim about intended action a^S is self-*

committing if $b^S\left(b^R\left(a^S\right)\right) = a^S$.

Definition 4.4. *Recommendation* $a^R \in A^R$ *is self-committing if* $b^R\left(b^S\left(a^R\right)\right) = a^R$.

Definition 4.5. *The stage game* g *is self-committing if every recommendation* $a^R \in A^R$ *is self-committing.*

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

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

Baliga and Morris formalizes the definition as follows.

Definition 4.6 (Baliga and Morris (2002)). *The game* g *is self-signalling (for the Sender) if* $g^S\left(a^S, b^R\left(a^S\right)\right) > g^S\left(a^S, a^R\right)$ *for every* $a^S \in A^S$, *and* $a^R \in A^R$ *where* $a^R \neq b^R\left(a^S\right)$.

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

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

Table 4.7: A Stage Game with Three Receiver Actions

Proposition 4.1. *If the stage game g is self-signalling and self-committing, then any strategy profile $\left(\left(m,a^S\right),s^R\right)$ that survives iterative deletion of weakly dominated strategies in the language game gives the Sender her Stackelberg payoff, that is,*

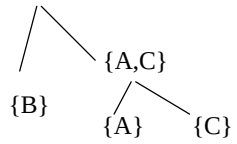
$$\begin{aligned} & g^S\left(a^S,a^R\right) \\ = & \max_{a^S} u^S\left(a^S,b^R\left(a^S\right)\right) \end{aligned}$$

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

An Example

To see the main idea behind the proof, it is easy to start with a simple example. The game is shown in table 4.7. This game has three pure strategy Nash-equilibria: $(a,A),(b,B)$ and (c,C). Payoffs for both players change if only one player changes the action taken. It is obvious that every recommendation is self-committing, and this game is self-signaling.

For ease of exposition, we will assume that the Sender can only give hierarchical recommendations that start with either "$\{B\}$" or "$\{A,C\}$." The top half of figure 4.1 lists every such message. The bottom half of the left panel of figure 4.1 tabulates every Sender strategy that survives the first, third, and fifth round of deletion of weakly dominated strategies.



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

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

Figure 4.1: The Iterative Process for a Game with Three Receiver Actions

Action a is listed in the cell at the intersection of row $S^S(1)$ and column "$\{A, C\}\{A\}$", while action b and c is not listed in that cell. This indicates that taking action a after sending the recommendation "$\{A, C\}\{A\}$" survives the first round of deletion of weakly dominated strategies, while taking action b or action c after sending the recommendation "$\{A, C\}\{A\}$" does not. The right panel in figure "$\{A, C\}\{A\}$" lists Receiver strategies that survive the 0^{th}, the second and the fourth round of deletion of weakly dominated strategies. For example, the Receiver strategy, *First Layer and A*, shown in the fourth row in the right panel of figure 4.1, responds to message "$\{B\}$" with action B, and to both message "$\{A, C\}\{A\}$" and message "$\{A, C\}\{C\}$" with action A. By definition, every language-based Receiver strategy survives the 0^{th} round of deletion of weakly dominated strategies in the language game.

We will first show why the Sender strategy that takes action c after sending the recommendation "$\{A, C\}\{A\}$" does not survive the first round of elimination for the Sender. By the self-signalling condition, the Sender prefers the Receiver action C to any other Receiver action if the Sender is going to take action c. The table in the middle of figure 4.1 shows that, in the language game, message "$\{A, C\}\{A\}$" and message "$\{A, C\}\{C\}$" solicit different actions from the Receiver only if the Receiver plays the *Literal* strategy, where message "$\{A, C\}\{A\}$" induces action A and message "$\{A, C\}\{C\}$" induces action C. It follows that taking action c after sending the recommendation "$\{A, C\}\{A\}$" is weakly dominated by taking the same action c while sending the recommendation "$\{A, C\}\{C\}$".

In a similar way, we show why the Sender strategy that takes action b after sending the recommendation "$\{A, C\}\{A\}$" is weakly dominated by the Sender strategy that takes

action b but sends the recommendation "$\{B\}$." First, the *Completely Literal* strategy responds to message "$\{A, C\}\{A\}$" and message "$\{B\}$" differently. By the self-signalling condition, the Sender prefers the Receiver action B to every other Receiver action if the Sender is going to take action b. If the Receiver plays a language-based strategy that responds to message "$\{B\}$" and message "$\{A, C\}\{A\}$" with different actions, the Receiver must respond to message "$\{B\}$" with action B, while responding to message "$\{A, C\}\{A\}$" with either action A or C. Therefore, the Sender strategy that takes action b after sending message "$\{A, C\}\{A\}$" is weakly dominated by the Sender strategy that takes action b after sending message "$\{B\}$." In other words, giving the recommendation "$\{A, C\}\{A\}$" while taking any action other action a, is weakly dominated for the Sender. Similarly, giving the recommendation "$\{A, C\}\{C\}$" while taking any action other than action c, is weakly dominated for the Sender.

However, it is not the case that every Sender strategy that takes an action which makes the recommendation suboptimal for the Receiver is weakly dominated in the first round of deletion. For example, taking action c after giving the recommendation "$\{B\}$" is not weakly dominated in the first round. We will show why it is not weakly dominated by the Sender strategy that, takes action c but gives the recommendation "$\{A, C\}\{C\}$" instead of the recommendation "$\{B\}$." It suffices to notice that message "$\{B\}$" yields action B while message "$\{A, C\}\{C\}$" yields action A under the receiver strategy *First Layer and A*, and that holding the Sender's action fixed at action c, the Sender prefers the Receiver to play action B over playing action A.

To show that the Sender strategy ("$\{B\}$", c) survives the first round of deletion of

weakly dominated strategies, we have to show that it is not weakly dominated by any other Sender strategy. As the action space grows bigger, the set of hierarchical recommendations grows bigger as well, and this becomes a daunting task. Instead, we invoke lemma 2.1 and construct a totally mixed belief that the Sender can hold about the Receiver's language-based strategies to which ("$\{B\}$", c) is a best response. For example, to show that the Sender strategy ("$\{A, C\}\{C\}$", c) survives the first round of deletion of weakly dominated strategies, we show that it is a best response to the totally mixed Receiver strategy

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

where ε is very small and σ^R is some totally mixed Receiver strategy in S_L^R.

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

We can now show that the Sender strategy ("$\{B\}$", c) is weakly dominated by the Sender strategy ("$\{A, C\}\{C\}$", c) in the third round of deletion. From the self-signaling condition, if the Sender is going to take action c, the recommendation "$\{A, C\}\{C\}$" would be better for her than any other message as long as the Receiver were to follow this recommendation

fully, that is, if the Receiver were to take the ultimately recommended action C. However, if the Receiver follows the recommendation "$\{A, C\}\{C\}$" halfway and takes action A, which is consistent with the first layer of recommendation but not with the last layer, the Sender would prefer message "$\{B\}$" which elicits her second preferred Receiver action when she intends to take action c. However, after the second round of deletion, the Receiver strategy that follows the recommendation "$\{A, C\}\{C\}$" halfway is eliminated. Furthermore, the Receiver strategy *Completely Literal* remains, and thus message "$\{A, C\}\{C\}$" and message "$\{B\}$" are not equivalent w.r.t. $S(2)$. Therefore, the Sender strategy ("$\{B\}$", c) is weakly dominated by ("$\{A, C\}\{C\}$", c) in the third round of deletion. Similarly, the Sender strategy ("$\{B\}$", a) does not survive the third round of deletion.

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

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

where ε is very small and σ^R is any totally mixed Receiver strategy in $S^R(2)$. Therefore, in the fourth round of deletion, the Receiver strategy *Always C* is weakly dominated by the Receiver strategy *Literal*, because given that the Sender sends message "$\{B\}$", the Sender must intend to take action b, to which B is the best action for the Receiver, and given that the Sender sends message "$\{A, C\}\{A\}$", the Sender must intend to take action a, to which A is the best action for the Receiver. Similarly, the Receiver strategy *Always A* is weakly dominated by the Receiver strategy *Completely Literal*.

It follows that, after four rounds of deletion of weakly dominated strategies, the message

"$\{B\}$" will certainly induce action B. Since the strategy profile (b, B) gives the Sender her highest payoff, the Sender strategy that sends message "$\{B\}$" and takes action b strictly dominates any other Sender strategy remaining after four rounds of deletion of weakly dominated strategies. Therefore, the unique outcome surviving iterative deletion of weakly dominated strategies gives the Sender her Stackelberg payoff.

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

The Proof

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

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

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

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

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

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

Figure 4.2 shows a partial set of hierarchical recommendations. Let ϕ be such that $a_{\phi(1)}^R = D$, $a_{\phi(2)}^R = B$, $a_{\phi(3)}^R = C$ and $a_{\phi(4)}^R = D$. Then the messages in the circle of figure 4.2 constitute the set $M_{\phi(1)}$, while the message "$\{D\}$" is the message $m_{\phi(1)}$.

Given a Receiver strategy s^R and a message subset F, let $s^R|_F$ denote the mapping from F to A^R that is equal to s^R conditional on F.

Denote the language-based cheap-talk extension game where $\left|A^R\right| = N$ by $G_L(N)$. There is a one-to-one correspondence between the message subset $M_{\phi(N-k)}$ in the game $G_L(N)$ and the set of hierarchical recommendations in the game $G_L(k)$, where the set of Receiver actions is $\left\{a_{\phi(N-k+1)}^R,...,a_{\phi(N)}^R\right\}$. If we identify Receiver strategies in $G_L(N)$ that are equal on $M_{\phi(N-k)}$ as equivalent, and if we identify Receiver strategies in $G_L(k)$ that are equal on the set of hierarchical recommendations, then there is also a one-to-

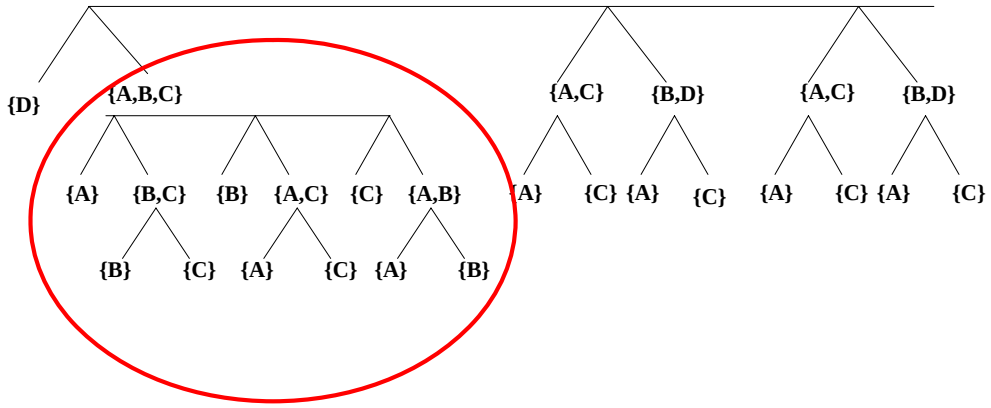


Figure 4.2: Partial Set of Hierarchical Recommendations, $A^R = \{A, B, C, D\}$.

one correspondence between the equivalent classes of Receiver strategies in $G_L(N)$ non-constant on $m_{\phi(N-k)} \cup M_{\phi(N-k)}$, and the equivalent classes of Receiver strategies in $G_L(k)$. The correspondence is follows. Let s^R be a Receiver strategy in $G_L(N)$ non-constant on $m_{\phi(N-k)} \cup M_{\phi(N-k)}$. It can be easily checked that a $s^R|_{M_{\phi(N-k)}}$ is a Receiver strategy in $G_L(k)$ restricted to the set of hierarchical recommendations. Let $\tilde{s}^R$ be a Receiver strategy in $G_L(k)$, then there exists a Receiver strategy s^R in $G_L(N)$ such that $\tilde{s}^R$ restricted to the set of hierarchical recommendations is equal to s^R restricted to $M_{\phi(N-k)}$.

Take a stage game g where the Receiver has N actions $\left\{a_1^R, ..., a_N^R\right\}$. Denote by $g^{\phi(N-k)}$ the truncated game which has the same payoff matrix but the Receiver can only use actions in $\left\{a_{\phi(N-k+1)}^R, ..., a_{\phi(N)}^R\right\}$, while the Sender can only use actions to which the Receiver's best response belong to $\left\{a_{\phi(N-k+1)}^R, ..., a_{\phi(N)}^R\right\}$. Denote the language-based cheap talk extension game of g and $g^{\phi(N-k)}$ by G_L and $G_L^{\phi(N-k)}$ respectively. Identify $M_{\phi(N-k)}$ with the hierarchical recommendation set in $G_L^{\phi(N-k)}$. Since $G_L^{\phi(N-k)}$ is a language-based cheap-

talk extension game with $\left|A^R\right| = N - k$, according to previous discussion, we can identify Receiver strategies in $G_L^{\phi(N-k)}$ with Receiver strategies in G_L non-constant on $M_{\phi(N-k)}$. Lemma 4.2 says that a Sender strategy is eliminated in the first round of deletion if it uses a message in $M_{\phi(N-k)}$ but takes an action whose Receiver best action does not belong to $\left\{a^R_{\phi(N-k+1)}, ..., a^R_{\phi(N)}\right\}$. Therefore, the pure strategy set after the first round of deletion in the game G_L restricted to the message subset $M_{\phi(N-k)}$ is very much related to that in the game $G_L^{\phi(N-k)}$. This is roughly why we could generalize the intuition gained from the Battle-of-the-Sexes game and the Investment game into coordination games with finitely many Receiver actions.

Lemma 4.2. *Given any $q < N - k$, the Sender strategy $\left(m_{\phi(N-k)}, a^S_{\phi(q)}\right)$ is weakly dominated in the first round.*

Proof. According to the self-signaling condition, he Sender prefers $b^R\left(a^S_{\phi(q)}\right) = a^R_{\phi(q)}$ to any other Receiver actions. If a language-based Receiver strategy responds to message $m_{\phi(q)}$ and $m_{\phi(N-k)}$ with different actions, it responds to message $m_{\phi(q)}$ with action $a^R_{\phi(q)}$, and to message $m_{\phi(N-k)}$ with some action a^R_j where $j > q$. In addition, there are language-based Receiver strategies that respond to $m_{\phi(q)}$ and $m_{\phi(N-k)}$ with different actions. It follows that the Sender strategy $\left(m_{\phi(N-k)}, a^S_{\phi(q)}\right)$ is weakly dominated by the Sender strategy $\left(m_{\phi(q)}, a^S_{\phi(q)}\right)$. □

Given a Receiver mixed strategy σ^R in S_L^R and a message bundle B, a Sender action a^S, define

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

Lemma 4.3 and lemma 4.4 give conditions under which there exists a mapping $\psi_d : S^R(j) \to S^R(j)$ such that $\psi_d\left(s^R\right)$ is equal to s^R outside of some message bundle B, and s^R behaves in a certain way on B, for every s^R in the domain which is non-constant on the smallest message bundle containing B. Based on these two lemmas, lemma 4.5 and lemma 4.6 both establish the existence of dominator Sender strategies. Lemma 4.7 combined with claim 4.1 shows that $s^R(m) = b^R\left(a^S\right)$ for every Sender strategy $\left(m, a^S\right) \in S^S(\infty)$ and every Receiver strategy $s^R \in S^R(\infty)$. The proof of the proposition shows that if $\left(m, a^S\right) \in S^S(\infty)$, then a^S is a Stackelberg action for the Sender.

Lemma 4.3. *Suppose Sender strategies $\left(m_1, a_1^S\right)$ and $\left(m_2, a_2^S\right)$ both belong to $S^S(k)$ and there exists a Receiver strategy that is a best response to both Sender strategies, that is,*

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

Let B be a message bundle that contains both $M^{cstr}\left(m_1, b^R\left(a_1^S\right)\right)$ and $M^{cstr}\left(m_2, b^R\left(a_2^S\right)\right)$. Then for every $j \leq k+1$, there exists a mapping $\psi_d : S^R(j) \to S^R(j)$ such that $\psi_d\left(s^R\right) = s^R$ for every Receiver strategy s^R constant on B, while for every s^R non-constant on B, $\psi_d\left(s^R\right)$ is equal to s^R outside of B, but is a best response to both Sender strategy $\left(m_1, a_1^S\right)$ and Sender strategy $\left(m_2, a_2^S\right)$.

Proof. Let $s^R \in S^R(j)$ be non-constant on B. Let σ^S be a totally mixed Sender strategy in $S^S(j-1)$ to which s^R is a best response. Let $\sigma^S|_{\tilde{M}}$ denote the probability distribution over Sender strategies conditional on the message sent being in $\tilde{M}$. Then every Receiver strategy which is a best response to

$$\begin{gathered}\tfrac{1-\varepsilon}{2}\left(m_1, a_1^S\right) + \tfrac{1-\varepsilon}{2}\left(m_2, a_2^S\right) \\ +\varepsilon\sigma^S|_{M\setminus\left(M^{cstr}\left(m_1, b^R\left(a_1^S\right)\right)\cup M^{cstr}\left(m_2, b^R\left(a_2^S\right)\right)\right)}\end{gathered}$$

exhibits the desired properties. Moreover, at least one best response survives the deletion of weakly dominated strategies in that round. We can then define $\psi_d\left(s^R\right)$ to be one such best response in $S^R(j)$. This completes the proof. □

Lemma 4.4. *Let F_1, F_2, and F_3 be three parallel message bundles, and B be the smallest message bundle that strictly contains F_1. If $\left(m_1, a_1^S\right) \in S^S(j-1)$ where*

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

$S^S(j-1)$ contains Sender strategies that use messages in F_2 and F_3 respectively, and $S^R(j)$ contains Receiver strategies s_2^R and s_3^R non-constant on B such that either $s_2^R|_{F_2}$ or $s_3^R|_{F_3}$ is not equivalent for the Receiver to a constant of $b^R\left(a_1^S\right)$ w.r.t. $S^S(j-1)$, then there exists a mapping $\psi_d : S^R(j) \to S^R(j)$ such that $\psi_d\left(s^R\right) = s^R$ for every Receiver strategy s^R constant on B, while for every s^R non-constant on B, $\psi_d\left(s^R\right)$ is equal to s^R outside of $F_1 \cup F_2 \cup F_3$, equal to s_i^R on F_i for $i = 2, 3$, and is a best response to the Sender strategy $\left(m_1, a_1^S\right)$.

Proof. See the Appendix. □

Lemma 4.5. *Let E be a message bundle in $M_{\phi(N-l)}$, where l is an integer between 3 and N. Let F_1 and F_2 be two other parallel message bundles in $M_{\phi(N-l)}$. Suppose a Sender strategy $\left(\hat{m}, \hat{a}^S\right)$ belongs to $S^S(k)$ for some iteration k where $\hat{m} \in E$, and there exists $s^R \in S^R(k)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ such that $s^R(\hat{m}) \neq b^R\left(\hat{a}^S\right)$. Let $\hat{\sigma}^R$ be a totally mixed strategy in $S^R(k-1)$ to which $\left(\hat{m}, \hat{a}^S\right)$ is a best response. Then there exists two Sender strategies $\left(m_1, a_1^S\right)$, $\left(m_2, a_2^S\right)$ in $S^S(k)$ where message m_1 belongs to F_1, message m_2 belongs to F_2, and Sender actions a_1^S and a_2^S both maximize*

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

over $\left\{a_{\phi(i)}^S : i > N-l\right\}$.

Proof. See the Appendix. □

Lemma 4.6. *Given natural numbers k and $N-l \neq q$ where $\left(m_{\phi(N-l)}, a_{\phi(q)}^S\right)$ survives the k^{th} round of deletion of weakly dominated strategies, if*

1. ***(Non-exclusiveness)**. a Sender strategy that uses a message in $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ and takes an action other than $a_{\phi(q)}^S$ also survives the k^{th} round of deletion of weakly dominated strategies, but*

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

then

1. *$\left(m_{\phi(N-l)}, a_{\phi(q)}^S\right)$ cannot be an action-strict best response to any $\sigma^R \in \Delta S^R(k-1)$*

which puts positive weights only on Receiver strategies constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, *and*

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

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

is a best response, there exists a Sender action $\hat{a}^S$ *not equal to* $a^S_{\phi(q)}$ *and a message* $\hat{m}$ *in* $M_{\phi(N-l)}$ *such that the Sender strategy* $\left(\hat{m}, \hat{a}^S\right)$ *is a best response to* σ^R *and survives the* k^{th} *round of deletion as well. In addition,*

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

Proof. See the Appendix. □

Lemma 4.7. *Given any Sender strategy* $\left(m, a^S\right)$ *surviving the* $(4k-1)^{th}$ *round of deletion, where* $m \in M_{\phi(N-k-1)}$, *any Sender strategy surviving the* $(4k-1)^{th}$ *round of deletion that uses a message in* $M^{SC}\left(m, b^R\left(a^S\right)\right)$ *must take the action* a^S. *That is,*

$$\begin{aligned} & S^S(4k-1) \cap \left(M^{SC}\left(m, b^R\left(a^S\right)\right) \times A^S\right) \\ = \quad & S^S(4k-1) \cap \left(M^{SC}\left(m, b^R\left(a^S\right)\right) \times \left\{a^S\right\}\right). \end{aligned}$$

Lemma 4.8. *Given any Sender strategy* $\left(m, a^S\right)$ *surviving the* $(4k-1)^{th}$ *round of deletion, where* $m \in M_{\phi(N-k-1)}$, *any Sender strategy surviving the* $(4k-1)^{th}$ *round of deletion that*

uses a message in $M^{cstr}\left(m, b^R\left(a^S\right)\right)$ *must take the action* a^S*. That is,*

$$\begin{aligned} & S^S(4k-1) \cap \left(M^{cstr}\left(m, b^R\left(a^S\right)\right) \times A^S\right) \\ = \; & S^S(4k-1) \cap \left(M^{cstr}\left(m, b^R\left(a^S\right)\right) \times \left\{a^S\right\}\right). \end{aligned}$$

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

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

Claim 4.1. *Given any* $s^R \in S^R\left(4\bar{k}\right)$ *non-constant on*

$$m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)},$$

and any $\left(m, a^S\right) \in S^S\left(4\bar{k}\right)$ *where* m *belongs to* $M_{\phi\left(N-\bar{k}-1\right)}$*, it has to be the case that* $s^R(m) = b^R\left(a^S\right)$*.*

Proof. By assumption,

$$\begin{aligned} & S^S\left(4\bar{k}-1\right) \cap \left(M^{cstr}\left(m, b^R\left(a^S\right)\right) \times A^S\right) \\ = \; & S^S\left(4\bar{k}-1\right) \cap \left(M^{cstr}\left(m, b^R\left(a^S\right)\right) \times \left\{a^S\right\}\right) \end{aligned}$$

for any Sender strategy $\left(m, a^S\right) \in S^S\left(4\bar{k}-1\right)$ where $m \in M_{\phi\left(N-\bar{k}-1\right)}$. Therefore, given any number of Sender strategies

$$\left(m_1, a_1^S\right), \left(m_2, a_2^S\right), ..., \left(m_n, a_n^S\right) \in S^S\left(4\bar{k}-1\right)$$

where all the messages belong to $M_{\phi\left(N-\bar{k}-1\right)}$ and no two actions are the same,

$$M^{cstr}\left(m_i, a_1^S\right) \cap M^{cstr}\left(m_j, a_2^S\right) = \emptyset$$

for any $i \neq j$. Thus, given any s^R non-constant on $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$, there exists $\psi\left(s^R\right)$ consistent with language such that

$$\psi\left(s^R\right)(m) = \begin{cases} b^R\left(a^S\right) & if\ m \in M_{\phi\left(N-\bar{k}-1\right)} \text{ and } \left(m, a^S\right) \in S^S\left(4\bar{k}-1\right) \\ s^R(m) & if\ m \notin M_{\phi\left(N-\bar{k}-1\right)} \end{cases}.$$

It can be easily seen that

$$u^R\left(\left(m, a^S\right), s^R\right) = u^R\left(\left(m, a^S\right), \psi\left(s^R\right)\right)$$

for $m \notin M_{\phi\left(N-\bar{k}-1\right)}$. If $m \in M_{\phi\left(N-\bar{k}-1\right)}$ and $\left(m, a^S\right) \in S^S\left(4\bar{k}-1\right)$, then

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

where strict inequality holds if $s^R(m) \neq b^R\left(a^S\right)$. Therefore, if s^R non-constant on $M_{\phi\left(N-\bar{k}-1\right)}$ survives the $(4k)^{th}$ round of deletion, it cannot be weakly dominated by $\psi\left(s^R\right)$. Thus it has to be the case that $s^R(m) = b^R\left(a^S\right)$ for every $\left(m, a^S\right) \in S^S\left(4\bar{k}-1\right)$ where $m \in M_{\phi\left(N-\bar{k}-1\right)}$. □

Claim 4.2. *If, after the* $\left(4\bar{k}-1\right)^{th}$ *round of deletion, two different actions are possible following messages in* $M_{\phi\left(N-\bar{k}-1\right)}$*, for some permutation* ϕ*, then there exists a Receiver strategy* s^R *non-constant on* $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$ *suriviving the* $\left(4\bar{k}\right)^{th}$ *round of deletion.*

Proof. If $S^S\left(4\bar{k}-1\right)$ contains $\left(m_1, a_1^S\right)$ and $\left(m_2, a_2^S\right)$ where $m_1, m_2 \in M_{\phi\left(N-\bar{k}-1\right)}$ and $a_1^S \neq a_2^S$, then

$$\begin{aligned} & S^S\left(4\bar{k}-1\right) \cap \left(M^{cstr}\left(m_1, a_1^S \cap M^{cstr}\left(m_2, a_2^S\right)\right) \times A^S\right) \\ = \; & S^S\left(4\bar{k}-1\right) \cap \left(M^{cstr}\left(m_1, a_1^S\right) \times A^S\right) \cap \left(M^{cstr}\left(m_2, a_2^S\right) \times A^S\right) \\ = \; & S^S\left(4\bar{k}-1\right) \cap \left(M^{cstr}\left(m_1, a_1^S\right) \times \left\{a_1^S\right\}\right) \cap \left(M^{cstr}\left(m_2, a_2^S\right) \times \left\{a_2^S\right\}\right) \\ = \; & \emptyset. \end{aligned}$$

Since $S^S\left(4\bar{k}-1\right) \neq \emptyset$, and $A^S \neq \emptyset$, it has to be the case that

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

Let $\sigma^S \in \Delta^+ S^S\left(4\bar{k}-1\right)$. Then for ε sufficiently small, any best response to

$$\frac{1-\varepsilon}{2}\left(m_1, a_1^S\right) + \frac{1-\varepsilon}{2}\left(m_2, a_2^S\right) + \varepsilon\sigma^S$$

must respond to message m_1 with action $b^R\left(a_1^S\right)$, message m_2 with action $b^R\left(a_2^S\right)$, and therefore is non-constant on $M_{\phi(N-k-1)}$. □

Claim 4.3. *Given any Sender strategy* $\left(m, a^S\right)$ *surviving the* $\left(4\bar{k}+1\right)^{th}$ *round of deletion, where* $m \in m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$*,* a^S *is also the only possible action surviving the*

$\left(4\bar{k}+1\right)^{th}$ *round of deletion following a message in* $M^{cstr}\left(m, b^R\left(a^S\right)\right)$. *That is,*

$$\begin{aligned} & S^S\left(4k+1\right) \cap \left(M^{cstr}\left(m, b^R\left(a^S\right)\right) \times A^S\right) \\ = \; & S^S\left(4k+1\right) \cap \left(M^{cstr}\left(m, b^R\left(a^S\right)\right) \times \left\{a^S\right\}\right). \end{aligned}$$

Proof. Let $\left(m, a^S\right)$ belong to $S^S\left(4k+1\right)$. If $m \in M_{\phi\left(N-\bar{k}-1\right)}$, then

$$\begin{aligned} & S^S\left(4k+1\right) \cap \left(M^{cstr}\left(m, b^R\left(a^S\right)\right) \times A^S\right) \\ \subset \; & S^S\left(4k-1\right) \cap \left(M^{cstr}\left(m, b^R\left(a^S\right)\right) \times A^S\right) \\ = \; & S^S\left(4k-1\right) \cap \left(M^{cstr}\left(m, b^R\left(a^S\right)\right) \times \left\{a^S\right\}\right) \end{aligned}$$

, and therefore

$$\begin{aligned} & S^S\left(4k+1\right) \cap \left(M^{cstr}\left(m, b^R\left(a^S\right)\right) \times A^S\right) \\ = \; & S^S\left(4k+1\right) \cap \left(M^{cstr}\left(m, b^R\left(a^S\right)\right) \times \left\{a^S\right\}\right). \end{aligned}$$

We worry only if there exists $\left(m_{\phi\left(N-\bar{k}-1\right)}, a^S_{\phi(q)}\right) \in S^S\left(4\bar{k}+1\right)$ where $q \neq N-\bar{k}-1$, and at least two actions following messages in $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$ are possible after the $\left(4\bar{k}+1\right)^{th}$ round of deletion. We know from the characterization of $S^S\left(1\right)$ that $q \geq N-\bar{k}-1$. Then from lemma 4.6, given any $\sigma^R_q \in \Delta S^R\left(4\bar{k}\right)$, there exists a Sender strategy $\left(\hat{m}, \hat{a}^S\right) \in S^S\left(4\bar{k}+1\right)$where and $\hat{m} \in M_{\phi\left(N-\bar{k}-1\right)}$ such that

$$\begin{aligned} & \chi\left(\sigma^R_q, m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}, \hat{a}^S\right) \\ \geq \; & \chi\left(\sigma^R_q, m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}, a^S_{\phi(q)}\right). \end{aligned}$$

But from claim 4.1, $s^R(\hat{m}) = b^R\left(\hat{a}^S\right)$ for every $s^R \in S^R\left(4\bar{k}\right)$ non-constant on $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$, and from claim 4.2, there exists $s^R \in S^R\left(4\bar{k}\right)$ non-constant on $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$. It follows that

$$\begin{aligned} & u^S\left(\left(\hat{m}, \hat{a}^S\right), \sigma_q^R\right) \\ = \; & \chi\left(\sigma_q^R, m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}, \hat{a}^S\right) \\ \geq \; & \chi\left(\sigma_q^R, m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}, a^S_{\phi(q)}\right) \\ > \; & u^R\left(\left(m_{\phi\left(N-\bar{k}-1\right)}, a^S_{\phi(q)}\right), \sigma_q^R\right), \end{aligned}$$

which contradicts the construction of σ_q^R. □

Claim 4.4. *Given any Receiver strategy s^R non-constant on $M_{\phi\left(N-\bar{k}-2\right)}$, surviving the $(4k+2)^{th}$ round of deletion, and any Sender strategy $\left(m, a^S\right)$ surviving the $(4k+1)^{th}$ round of deletion where $m \in m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$, $s^R(m) = b^R\left(a^S\right)$.*

Proof. The proof is analogous to that of claim 4.1. □

Claim 4.5. *Given any pair of Sender strategies surviving the $(4k+3)^{th}$ round of deletion, $\left(m_1, a_1^S\right), \left(m_2, a_2^S\right)$, where $m_1, m_2 \in M_{\phi(N-k-2)}$ and $a_1^S \neq a_2^S$, it has to be the case that*

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

Proof.

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

if m_1 and m_2 both belong to $m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$, or if m_1 and m_2 belong to different

parallel message subsets in $M_{\phi\left(N-\bar{k}-2\right)}$. We need only worry if m_1 and m_2 belong to the same parallel message set E in $M_{\phi\left(N-\bar{k}-2\right)}$. Suppose to the contrary that

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

Then from lemma 4.5, given $\sigma_1^R \in S^R\left(4\bar{k}+2\right)$, there exists $\left(\hat{m}, \hat{a}^S\right) \in S^S\left(4\bar{k}+3\right)$ where $\hat{m} \in m_{\phi\left(N-\bar{k}-1\right)} \cup M_{\phi\left(N-\bar{k}-1\right)}$ and

$$\begin{aligned} & \chi\left(\sigma_1^R, m_{\phi\left(N-\bar{k}-2\right)} \cup M_{\phi\left(N-\bar{k}-2\right)}, \hat{a}^S\right) \\ \geq\ & \chi\left(\sigma_1^R, m_{\phi\left(N-\bar{k}-2\right)} \cup M_{\phi\left(N-\bar{k}-2\right)}, a_1^S\right). \end{aligned}$$

But from the previous claim, $s^R\left(\hat{m}\right) = b^R\left(\hat{a}^S\right)$ for every $s^R \in S^R\left(4\bar{k}+2\right)$ non-constant on $M_{\phi\left(N-\bar{k}-2\right)}$. Therefore,

$$\begin{aligned} & u^S\left(\left(\hat{m}, \hat{a}^S\right), \sigma_1^R\right) \\ =\ & \chi\left(\sigma_1^R, m_{\phi\left(N-\bar{k}-2\right)} \cup M_{\phi\left(N-\bar{k}-2\right)}, \hat{a}^S\right) \\ \geq\ & \chi\left(\sigma_1^R, m_{\phi\left(N-\bar{k}-2\right)} \cup M_{\phi\left(N-\bar{k}-2\right)}, \hat{a}^S\right) \\ >\ & u^R\left(\left(m_1, a_1^S\right), \sigma_1^R\right) \end{aligned}$$

if $\sigma_1^R \in \Delta^+ S^R\left(4\bar{k}+2\right)$. This contradicts the construction that of σ_1^R. □

□

Proof of Proposition. Denote by $A^S_{Stackelberg}$ the set of Sender actions that maximize

$$g^S\left(a^S, b^R\left(a^S\right)\right).$$

Define

$$\Pi^S(k) \equiv \left\{u^S\left(s^S, s^R\right) : \left(s^S, s^R\right) \in S(k)\right\}.$$

Step 1 Given any k, any pair of Sender strategies $\left(m^*, a_*^S\right), \left(\hat{m}, \hat{a}^S\right)$ in $S^S(k)$ where $a_*^S \in A^S_{Stackelberg}$ but $\hat{a}^S \notin A^S_{Stackelberg}$, it cannot be the case that $\left(m^*, a_*^S\right)$ is weakly dominated by $\left(\hat{m}, \hat{a}^S\right)$ w.r.t. $S(k)$.

Proof. Since $\left(m^*, a_*^S\right) \in S^S(k)$, there exists $s_*^R \in S^R(k)$ where $s_*^R(m^*) = b^R\left(a_*^S\right)$. Thus,

$$\begin{aligned}
& u^S\left(\left(m^*, a_*^S\right), s_*^R\right) \\
= \; & u^S\left(a_*^S, b^R\left(a_*^S\right)\right) \\
> \; & u^S\left(\hat{a}^S, b^R\left(\hat{a}^S\right)\right) \\
\geq \; & u^S\left(\hat{a}^S, s_*^R(\hat{m})\right) \\
\geq \; & u^S\left(\left(\hat{m}, \hat{a}^S\right), s_*^R\right).
\end{aligned}$$

So $\left(m^*, a_*^S\right)$ is not weakly dominated by $\left(\hat{m}, \hat{a}^S\right)$. □

Step 2 If $\left(m_1, a_1^S\right), \left(m_2, a_2^S\right) \in S^S(\infty)$, then $g^S\left(a_1^S, b^R\left(a_1^S\right)\right) = g^S\left(a_2^S, b^R\left(a_2^S\right)\right)$. Therefore, $\min \Pi^S(\infty) = \max \Pi^S(\infty)$.

Proof. Suppose to the contrary that $g^S\left(a_1^S, b^R\left(a_1^S\right)\right) > g^S\left(a_2^S, g^R\left(a_2^S\right)\right)$. Lemma 4.7 combined with claim 4.1 tells us that $s^R(m) = b^R\left(a^S\right)$ for every $\left(m, a^S\right) \in S^S(\infty)$ and

$s^R \in S^R(\infty)$. Therefore,

$$\begin{aligned} & u^S\left(\left(m_1, a_1^S\right), s^R\right) \\ = \; & u^S\left(a_1^S, b^R\left(a_1^S\right)\right) \\ > \; & u^S\left(a_2^S, b^R\left(a_2^S\right)\right) \\ = \; & u^S\left(\left(m_2, a_2^S\right), s^R\right). \end{aligned}$$

Therefore $\left(m_2, a_2^S\right)$ is weakly dominated by $\left(m_1, a_1^S\right)$ w.r.t. $S(\infty)$. Contradiction! □

Step 3 $\max \Pi^S(\infty) = \max_{a^S} u^S\left(a^S, b^R\left(a^S\right)\right)$.

Proof. Suppose to the contrary that $\max \Pi(\infty) < \max_{a^S} g^S\left(a^S, b^R\left(a^S\right)\right)$. Then

$$M \times A^S_{Stackelberg} \cap S^S(\infty) = \emptyset.$$

Let $\left(\hat{m}, \hat{a}^S\right)$ be one of the last strategies in $M \times A^S_{Stackelberg}$ to be eliminated by the iterative process. Let K be such that $\left(\hat{m}, \hat{a}^S\right)$ belongs to $S^S(K)$ but not to $S^S(K+1)$. So $\left(\hat{m}, \hat{a}^S\right)$ is weakly dominated by some $\left(m^*, a_*^S\right) \in S^S(K+1)$. By construction of K, $a_*^S \notin A^S_{Stackelberg}$. But this contradicts the conclusion from 4.4.1. □

Based on the previous three steps, we conclude that

$$\begin{aligned} \min \Pi^S(\infty) &= \max \Pi^S(\infty) \\ &= \max_{a^S} g^S\left(a^S, b^R\left(a^S\right)\right). \end{aligned}$$

So the Sender is guaranteed her Stackelberg payoff. □

4.4.2 Games with Positive Spillovers

The self-signalling criterion implies that the Sender's preference over the Receiver's actions differ with her own intention. Our language assumption combined with iterative admissibility connects different messages with different preferences. This then separates one intention from the other and guarantees the Sender her Stackleberg payoff.

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

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

Proposition 4.2. *If the stage game is self-committing and the Sender's preference over the Receiver's actions is independent of her own action, then for every* $\left(a^S, a^R\right) \in A(\infty)$*, there exists* $\left(m, a^S\right) \in S^S(\infty)$ *and* $s^R \in S^R(\infty)$ *such that* $s^R(m) = a^R$.

Proof. Since $A^R(\infty) = A^R$, it is easy to see that $A^S(\infty) = A^S(1)$. Define

$$M(k) := supp\left(S^S(k)\right)|_M.$$

So $M(k)$ is the set of messages that are used in $S^S(k)$.

Step 1 First, we show that $S^S(1) = M(1) \times A^S(1)$. This implies that any message that is used in $S^S(1)$ could be uttered with any intention in $A^S(1)$. Suppose $\left(\hat{m}, \hat{a}^S\right) \in S^S(1)$. Then $\left(\hat{m}, \hat{a}^S\right)$ is not weakly dominated by $\left(m', \hat{a}^S\right)$ for any $m' \neq \hat{m}$. Write $\hat{m} = A_1 ... A_n$. Then there exists $\bar{a}_j \in A_j$ and $\underline{a}_j \in A_{j-1} \backslash A_j$ where $g^S\left(\hat{a}^S, \bar{a}_j\right) >$

$g^S\left(\hat{a}^S, \underline{a}_j\right)$, because otherwise, $\left(\hat{m}, \hat{a}^S\right)$ would be weakly dominated by every $m' \in M\left(A_1 ... A_{j-1}\left(A_{j-1} \backslash A_j\right)\right)$.

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

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

and

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

for $j = 2, ..., n$. It follows that

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

for every $m \in M\left(A_1 ... A_{j-1}\left(A_{j-1} \backslash A_j\right)\right)$. It follows that

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

for every a^S and every $m \in M\left(A_1 ... A_{j-1}\left(A_{j-1} \backslash A_j\right)\right)$. Define

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

Therefore,

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

for every $m \neq \hat{m}$, every a^S and every ε sufficiently small. Let a^R denote also the constant Receiver strategy that takes the action a^R upon receiving every message. Let $\alpha \in \Delta A^R$ also denote the Receiver strategy that puts weights $\alpha\left(a^R\right)$ on the constant strategy a^R. Then for ε sufficiently small,

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

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

Step 2 It remains to show that, if $S^R(k) = S_L^R$ and $S^S(k) = S^S(1)$, then $S^R(k+1) = S_L^R$ and $S^S(k+1) = S^S(1)$. Since $S^R(k) = S_L^R$, the set of Sender best responses to mixed strategies in $S^R(k)$ is equal to $S^S(1) \subset S^S(k)$. It follows that $S^S(k+1) = S^S(1)$. Now we will show that $S^R(k+1) = S_L^R$. Pick an arbitrary $\hat{s}^R \in S_L^R$. Define an equivalence relation $\sim$ on M as follows. m_1 is equivalent to m_2 if either m_1 belongs to $M^{cstr}\left(m_2, \hat{s}^R(m_2)\right)$ or m_2 belongs to $M^{cstr}\left(m_1, \hat{s}^R(m_1)\right)$. It is easy to verify that this definition forms an equivalence relation. We can partition $M(k)$ by $\sim$. For

every $m \in M(k)$, pick $h(m) \in \arg\max_{\substack{m' \sim m \\ m' \in M(k)}} M^{cstr}\left(m', \hat{s}^R(m')\right)$. Let $\langle m \rangle_k$ denote the set of m' in $M(k)$ that is equivalent to m. Then

$$M^{cstr}\left(h(m), \hat{s}^R(h(m))\right) = \langle m \rangle_k$$

for every $m \in M(k)$. By definition of this equivalence relation, and the construction that $\hat{s}^R \in S_L^R$, $\hat{s}^R$ must be constant on $\langle m \rangle_k$ for every $m \in M(k)$. Define

$$\hat{\sigma}^S := \sum_{m \in M(k)} \frac{1}{\sum_{m \in M(k)} |\langle m \rangle_k|} \left(h(m), \left(b^R\right)^{-1}\left(\hat{s}^R(m)\right)\right).$$

Then for ε sufficiently small, for every $\sigma^S \in \Delta^+ S^S(k)$, $\hat{s}^R$ is a best response to $(1-\varepsilon)\hat{\sigma}^S + \varepsilon\sigma^S$. To see this, note that by definition, $\hat{s}^R(m)$ is the best action to $\left(b^R\right)^{-1}\left(\hat{s}^R(m)\right)$. If $m' \in M(k)$ where $s^R(m') \neq \hat{s}^R(m')$, then $s^R(h(m')) \neq \hat{s}^R(h(m'))$ because $\hat{s}^R$ is a constant on $M^{cstr}\left(h(m'), \hat{s}^R(h(m'))\right)$. Then

$$u^R\left(\hat{\sigma}^s, s^R\right) < u^R\left(\hat{\sigma}^S, \hat{s}^R\right).$$

Since S_L^R is finite, and the inequality is strict for every s^R which is not equal to $\hat{s}^R$ at some message $m \in M(k)$, there exists ε small enough such that

$$u^R\left((1-\varepsilon)\hat{\sigma}^S + \varepsilon\sigma^S, s^R\right) < u^R\left((1-\varepsilon)\hat{\sigma}^S + \varepsilon\sigma^S, \hat{s}^R\right)$$

for every s^R which is not equal to $\hat{s}^R$ at some message $m \in M(k)$, and every $\sigma^S \in \Delta^+ S^S(1)$. We are done!

□

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

Table 4.8: leading example in Baliga Morris (2002)

4.5 Comparison with Baliga and Morris

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

The easiest way to see the comparison is to look at the leading example in Baliga and Morris (2002). The game is shown in table 4.8.

In this stage game, both action *Invest* and action *Not* are self-committing. So the recommendation "invest" and "not invest" are both self-committing. If $x < 0$, the stage-game is self-signalling, while the game exhibits positive spillovers if $x > 0$. To formally study the role of self-signalling, Baliga and Morris (2002) study the following incomplete information game where with probability $1 - p$ the Sender is of *Low Cost* and with probability $p > 0$ the Sender is of *High Cost*. The *Low Cost* type has the same payoff matrix as in the complete information game of table 4.8. However, the *High Cost* Sender has a dominant strategy to not invest. The Receiver's payoff depends only on the action taken by the Sender, not on the Sender's type. Therefore, the Receiver cares about the type of the

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

Table 4.9: Incomplete Information Investment Game

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

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

the self-signalling criterion.

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

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

Proposition 4.3 (Baliga and Morris (2002)). *If (1) for each $a^S \in A^S$, there exists a type $\tau\left(a^S\right) \in T$ such that a^S the dominant strategy for the Sender in the game $g^S\left(., t\right)$; and (2) for each action $a^R \in A^R$, there exists $a^S \in A^S$ such that $a^R = b^R\left(a^S\right)$, then there exists a full revelation perfect Bayesian equilibrium in the one-sided cheap talk game if and only if*

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

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

in the game G. Define

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

$\bar{d}$ is thus the maximum payoff difference for the Receiver. Expand the utility function $g^S : A^S \times A^R \to R$ into $\tilde{g}^S : A^S \times A^R \times T \to R$ as follows. $\tilde{g}^S\left(.,\tau\left(a_*^S\right)\right) = g\left(.\right)$, and for every $a^S \neq a_*^S$, $\tilde{g}^S\left(a^S,a'^R,\tau\left(a^S\right)\right) = g^S\left(a^S,a'^R\right)$ for every $a'^R \in A^R$, and $\tilde{g}^S\left(a'^S,a'^R,\tau\left(a^S\right)\right) = g^S\left(a'^S,a'^R\right) - 2\bar{d}$, for every $a'^S \neq a^S$ and every $a'^R \in A^R$. Denote the one-sided cheap talk extension game of g with language by G_L, and the one-sided cheap talk extension game of $\tilde{g}$ by $\tilde{G}$. Then proposition 4.3 implies that $\tilde{G}$ has a full revelation prefect Bayesian equilibrium if and only if the complete information stage game g is self-signalling and every action a^S for the Sender is self-committing. In this equilibrium, the type $\tau\left(a_*^S\right)$, whose payoff matrix is g, gets her Stackelberg payoff. Our positive result equivalently states that if every a^S in g is self-committing and g is strongly self-signalling, the unique iterative admissible outcome in G_L gives the Sender her Stackelberg payoff. The negative result in Baliga and Morris (2002) says that there is no communication in any equilibrium of $\tilde{G}$ if g exhibits binary action positive spillovers. Equilibrium outcomes of $\tilde{G}$ in such games depend on the common prior over T. If there is no common prior, and we allow any prior over T, we can span every rationalizable outcome. Our negative result relaxes the condition to any finite games with positive spillovers, and states that every rationalizable outcome is consistent with iterative admissibility in G_L.

4.6 Conclusion

By adjusting our language assumption to games where there is no natural order in actions and applying our general framework, we formalize the idea of self-committing and self-signalling within the framework of complete information games. We show that, if the stage game is self-committing and strongly self-signalling, every iterative admissible outcome in the language game gives the Sender her Stackelberg payoff. On the other hand, if the stage game is self-committing but the Sender's preference for the Receiver's actions does not depend on her intended action, every rationalizable stage game outcome is also an iteratively admissible outcome in the language game.

4.7 Appendix

4.7.1 Proof for lemma 4.4

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

$$(1-\varepsilon)\left(m_1, a_1^S\right) + \frac{\varepsilon(1-\varepsilon)}{2}\sigma_2^S|_{F_2} + \frac{\varepsilon(1-\varepsilon)}{2}\sigma_3^S|_{F_3} + \varepsilon^2\sigma^S|_{M\setminus(F_2\cup F_3)} \tag{4.1}$$

which responds to message m_1 with action a_1^S, is equal to s_i^R on F_i for $i = 2, 3$, and is equal to Receiver strategy s^R outside of $F_1 \cup F_2 \cup F_3$. Since the Sender strategy in line 4.1 is totally mixed on $S^S(j-1)$, any Receiver strategy as described above which is a best

response to expression 4.1 belongs to $S^R(j)$. To see this, notice that

$$u^S\left(\left(m_1, a_1^S\right), b^R\left(a_1^S\right)\right) > u^S\left(\left(m_1, a_1^S\right), s^R\right)$$

for any Receiver strategy that responds to message m_1 with an action not equal to $b^R\left(a_1^S\right)$.

4.7.2 Proof for lemma 4.5

It is easy to see that the statement holds for every integer l between 3 and N, for $k = 1$. Suppose the statement is true for every integer l between 3 and N, for $k = 1, ..., \bar{k}$. Suppose $S^S\left(\bar{k}+1\right)$ contains a Sender strategy $\left(\hat{m}, \hat{a}^S\right)$ where $\hat{m}$ belongs to a message bundle E in $M_{\phi(N-l)}$ for some $l \in \{3, 4, ...N\}$ and there exists $s'^R \in S^R\left(\bar{k}\right)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ such that $s'^R(\hat{m}) \neq b^R\left(\hat{a}^S\right)$. Let $A^{\max}$ denote the set of Sender actions in $\left\{a_{\phi(i)}^S : i > N-l\right\}$ that maximize

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

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

Claim 4.6. *$S^R\left(\bar{k}\right)$ contains a Receiver strategy s^R non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$.*

Proof. Suppose not, then

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

Therefore, $\left(m_1, a_1^S\right)$ is a best response to $\hat{\sigma}^R$. Since $\hat{\sigma}^R$ is a totally mixed Receiver strategy in $S^R\left(\bar{k}\right)$, it follows that $\left(m_1, a_1^S\right)$ belongs to $S^S\left(\bar{k}+1\right)$. We have thus arrived at a contradiction. □

Claim 4.7. *Given any $\left(\tilde{m}, \tilde{a}^S\right) \in S^S\left(\bar{k}\right)$ where $\tilde{a}^S \in A^{\max}$ and $\tilde{m} \in E \cup F_1 \cup F_2$, there exists $s^R_{wrong,(\tilde{m},\tilde{a}^S)} \in S^R\left(\bar{k}\right)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ such that $s^R_{wrong,(\tilde{m},\tilde{a}^S)}\left(\tilde{m}\right) \neq b^R\left(\tilde{a}^S\right)$. Given any $\left(\tilde{m}, \tilde{a}^S\right) \in S^S\left(\bar{k}\right)$ where $\tilde{a}^S \in A^{\max}$ and $\tilde{m} \in E \cup F_2$, there exists $\psi_{d,(\tilde{m},\tilde{a}^S)} : S^R\left(\bar{k}\right) \to S^R\left(\bar{k}\right)$ such that $\psi_d\left(s^R\right) = s^R$ for every Receiver strategy s^R constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, while for every s^R non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, $\psi_d\left(s^R\right)$ is equal to s^R outside of $E \cup F_1 \cup F_2$, responds to message m_1 with action $b^R\left(a_1^S\right)$, and to message $\tilde{m}$ with an action other than $b^R\left(\tilde{a}^S\right)$.*

Proof. From claim 4.7, there exists $s^R_{wrong} \in S^R\left(\bar{k}\right)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ such that $s^R_{wrong,(\tilde{m},\tilde{a}^S)}\left(\tilde{m}\right) \neq b^R\left(\tilde{a}^S\right)$. The proof for this claim is broken down into two cases.

First, suppose $|A^{\max}| = 1$. By construction, both $\tilde{a}^S$ and a_1^S belong to $A^{\max}$. Since $|A^{\max}| = 1$, $\tilde{a}^S = a_1^S$. Therefore, $s^R_{wrong}\left(\tilde{m}\right) \neq b^R\left(a_1^S\right)$. The claim follows immediately by

lemma 4.4.

Now suppose otherwise. If $s^R_{wrong,(\tilde{m},\tilde{a}^S)} \neq b^R\left(a_1^S\right)$, then again the claim immediately follows from lemma . Otherwise, first consider the situation where $\tilde{m} \in F_2$ and $\hat{a}^S = a_1^S$. In that case, $s'^R(\hat{m}) \neq b^R\left(a_1^S\right)$. The claim again follows from lemma 4.4. If $\tilde{m} \in F_2$ and $\hat{a}^S \neq a_1^S$, then let $\hat{s}^R$ be a Receiver strategy in $S^R\left(\bar{k}\right)$ which is a best response to $\frac{1}{2}\left(m_1, a_1^S\right) + \frac{1}{2}\left(\hat{m}, \hat{a}^S\right)$. It is easy to see that $\hat{s}^R$ is non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, and $\hat{s}^R$ is not a constant of $b^R\left(a_1^S\right)$ on E since $\hat{s}^R(\hat{m}) = b^R\left(\hat{a}^S\right) \neq b^R\left(a_1^S\right)$. Again the claim follows from lemma 4.4. If $\tilde{m} \in E$, and $a_2^S = a_1^S$, then $s^R_{wrong,\left(m_2,a_2^S\right)}(m_2) \neq b^R\left(a_1^S\right)$, then we have $s^R_{wrong,\left(m_2,a_2^S\right)}$ and $s^R_{wrong,(\tilde{m},\tilde{a}^S)}$ satisfy the conditions of lemma 4.4 and the claim follows. Otherwise, $\tilde{m} \in E$ and $a_2^S \neq a_1^S$. We know that $S^R\left(\bar{k}\right)$ contains a Receiver strategy s_2^R that is a best response to $\frac{1}{2}\left(m_1, a_1^S\right) + \frac{1}{2}\left(m_2, a_2^S\right)$. It has to be the case that s_2^R is not a constant of $b^R\left(a_1^S\right)$ on F_2 since $s_2^R(m_2) = b^R\left(a_2^S\right)$. Then s_2^R and $s^R_{wrong,(\tilde{m},\tilde{a}^S)}$ together satisfy the conditions of lemma 4.4 and the claim follows. □

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

$$\left\{\psi_{d,(\tilde{m},\tilde{a}^S)} : \left(\tilde{m}, \tilde{a}^S\right) \in E \cup F_2 \times A^{\max} \cap S^S\left(\bar{k}+1\right)\right\}.$$

Note that $T_d\left(s^R\right)(m_1) = b^R\left(a_1^S\right)$ for every s^R non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. Then

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

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

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

$$\begin{aligned} & u^S\left(\left(m, a^S\right), T_d\left(\hat{\sigma}^R\right)\right) \\ \leq\ & \chi\left(\hat{\sigma}^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a^S\right) \qquad (4.2) \\ \leq\ & \chi\left(\hat{\sigma}^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a_1^S\right) \qquad (4.3) \\ =\ & u^S\left(\left(m_1, a_1^S\right), T_d\left(\hat{\sigma}^R\right)\right). \end{aligned}$$

If $a^S \notin A^{\max}$, then strictly inequality holds for inequality 4.3. If $a^S \in A^{\max}$ and $\left(m, a^S\right) \in S^S\left(\bar{k}+1\right)$, then strictly inequality holds on inequality 4.2, because from claim 4.8 and the construct that $\hat{\sigma}^R$ is totally mixed on $S^R\left(\bar{k}\right)$, $\hat{\sigma}^R$ puts positive probability on Receiver strategies non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, and from the definition of T_d, $T_d\left(\hat{\sigma}^R\right)$ puts positive probability on Receiver strategies non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ that respond to message m with an action other than $b^R\left(a^S\right)$. If $m \in F_1$ but $a^S \notin A^{\max}$, then it follows that

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

Therefore, if $\left(m_*, a_*^S\right)$ is a best response to $T_d\left(\hat{\sigma}^R\right) \in \Delta S^R\left(\bar{k}\right)$, it has to be the case that either message $m_* \in F_1$ and Sender action $a_*^S \in A^{\max}$ or $\left(m, a^S\right) \notin S^S\left(\bar{k}+1\right)$. Since at least one Sender best response to $T_d\left(\hat{\sigma}^R\right)$ survives the $\bar{k}^{th}$ round of deletion of weakly dominated strategies, one that survives must be such that the message belongs to F_1 and the action belongs to $A^{\max}$. We have thus completed the proof of the lemma.

4.7.3 Proof for lemma 4.6

Proof. Both statements are true for $k = 1$, for all $N - l \neq q$. From the characterization we also know that $N - l \leq q$ for any $\left(m_{\phi(N-l)}, a_{\phi(q)}^S\right) \in S^S(k)$ where $k \geq 1$. Suppose both statements are true for every $N - l \neq q$, for $k = 1, ..., \bar{k}$. Let $\left(m_{\phi(N-l)}, a_{\phi(q)}^S\right)$ be a Sender strategy that survives the $\left(\bar{k}+1\right)^{th}$ round of deletion where $N - l \neq q$. From the characterization of $S^S(1)$, we know that $q > N - l$. Suppose $\left(m_{\phi(N-l)}, a_{\phi(q)}^S\right)$ satisfies both the *Non-exclusiveness* condition and the *Always Incorrect* condition.

Proof of Statement 1

Suppose to the contrary that the Sender strategy $\left(m_{\phi(N-l)}, a_{\phi(q)}^S\right)$ is an action-strict best response to a mixed Receiver strategy $\sigma_{flat}^R \in \Delta S^R\left(\bar{k}-1\right)$ which puts positive weights only on Receiver strategies constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. By the assumption that both statements are true for every $k \leq \bar{k}$, we induce that there exists $\hat{m}$ in $M_{\phi(N-l)}$ such that $\left(\hat{m}, a_{\phi(q)}^S\right)$ survives the $\bar{k}^{th}$ round of deletion. By the assumption that the *Non-exclusiveness* condition holds, $S^S\left(\bar{k}+1\right)$ contains a Sender strategy $\left(\tilde{m}, a_{\phi(r)}^S\right)$ where $\tilde{m} \in M_{\phi(N-l)}$ and $r \neq q$.

Claim 4.8. *There exists Receiver strategies in* $S^R\left(\bar{k}\right)$ *non-constant on the message bundle* $m_{\phi(N-l)} \cup M_{\phi(N-l)}$.

Proof. If not, then every message in the message bundle $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ is equivalent to each other. However, $\left(\hat{m}, a^S_{\phi(q)}\right)$ is eliminated at the $\left(\bar{k}+1\right)^{th}$ round while $\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right)$ survives the $\left(\bar{k}+1\right)^{th}$ round of deletion, which implies that message $\hat{m} \in M_{\phi(N-l)}$ is not equivalent to message $m_{\phi(N-l)}$. We have arrived at a contradiction. □

Claim 4.9. $S^S\left(\bar{k}\right)$ *contains a Sender strategy* $\left(\hat{m}_2, a^S_{\phi(q)}\right)$ *where* $\hat{m}_2 \in M_{\phi(N-l)}$ *and there exists a language-based Receiver strategy that is a best response to both* $\left(\hat{m}_2, a^S_{\phi(q)}\right)$ *and* $\left(\tilde{m}, a^S_{\phi(r)}\right)$. *That is,*

$$M^{cstr}\left(\hat{m}_2, b^R\left(a^S_{\phi(q)}\right)\right) \cap M^{cstr}\left(\tilde{m}, a^S_{\phi(r)}\right) = \emptyset.$$

Proof. If

$$M^{cstr}\left(\hat{m}, b^R\left(a^S_{\phi(q)}\right)\right) \cap M^{cstr}\left(\tilde{m}, a^S_{\phi(r)}\right) = \emptyset, \tag{4.4}$$

then we are done by the construction of $\hat{m}$. Otherwise, any Receiver strategy that is a best response to $\left(\tilde{m}, a^S_{\phi(r)}\right)$ responds to message $\hat{m}$ with an action different from $b^R\left(a^S_{\phi(q)}\right)$. Suppose that the claim does not hold. Let j be the last round after which this statement is still true, and let $\left(\hat{m}_2, a^S_{\phi(q)}\right)$ be a Sender strategy in $S^S\left(j\right)$ where

$$M^{cstr}\left(\hat{m}_2, a^S_{\phi(q)}\right) \cap M^{cstr}\left(\tilde{m}, a^S_{\phi(r)}\right) = \emptyset.$$

Let $\hat{\sigma}^R$ be a totally mixed strategy in $S^R\left(\bar{k}-1\right)$ to which $\left(\hat{m}, a^S_{\phi(q)}\right)$ is a best response. By lemma 4.3, there exists a mapping $\psi_d : S^R\left(j\right) \to S^R\left(j\right)$ such that $\psi_d\left(s^R\right) = s^R$ for

every Receiver strategy s^R constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, while for every s^R non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, $\psi_d\left(s^R\right)$ is equal to s^R outside of B, but takes on action $b^R\left(a^S_{\phi(q)}\right)$ after receiving message $\hat{m}_2$ and action $a^S_{\phi(r)}$ after receiving message $\tilde{m}_2$.

Since $\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right)$ is an action-strict best response to σ^R_{flat}, for ε sufficiently small, any Sender best respone to

$$(1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\psi_d\left(\hat{\sigma}^R\right)$$

involves taking action $a^S_{\phi(q)}$. By the definition of ψ_d, message $\hat{m}_2$ either yields the same Receiver action as any other message in the bundle $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, or $\hat{m}_2$ yields Receiver action $b^R\left(a^S_{\phi(q)}\right)$. By the self-signalling condition, the Sender prefers the Receiver action $b^R\left(a^S_{\phi(q)}\right)$ the most given that she intends to take action $a^S_{\phi(q)}$. From claim 4.2 and the construction that $\hat{\sigma}^R$ is totally mixed in $S^R\left(\bar{k}-1\right)$, $\hat{\sigma}^R$ puts positive probability on Receiver strategies non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. Therefore,

$$\begin{aligned}
& u^S\left(\left(\hat{m}_2, a^S_{\phi(q)}\right), (1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\psi_d\left(\hat{\sigma}^R\right)\right) \\
> \; & u^S\left(\left(\hat{m}, a^S_{\phi(q)}\right), \sigma^R_{flat}\right) \\
\geq \; & u^S\left(\left(m, a^S_{\phi(q)}\right), \sigma^R_{flat}\right) \\
= \; & u^S\left(\left(m, a^S_{\phi(q)}\right), (1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\psi_d\left(\hat{\sigma}^R\right)\right)
\end{aligned}$$

for every $m \notin m_{\phi(N-l)} \cup M_{\phi(N-l)}$. Therefore, given any $\left(m^*, a^S_{\phi(q)}\right) \in S^S(j+1)$ which is a best response to $(1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\psi^d\left(\hat{\sigma}^R\right)$, m^* belongs to $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ and it has to be the case that $\psi^d\left(s^R\right)(m^*) = b^R\left(a^S_{\phi(q)}\right)$ for every $s^R \in S^R(j)$ nonconstant on

$m_{\phi(N-l)} \cup M_{\phi(N-l)}$. It follows that

$$M^{cstr}\left(m^*, b^R\left(a^S_{\phi(q)}\right)\right) \cap M^{cstr}\left(\tilde{m}, b^R\left(a^S_{\phi(r)}\right)\right) = \emptyset.$$

we have thus reached a contradiction. □

Claim 4.10. *The Always Incorrect condition cannot hold.*

Proof. Following Claim 4.9, we know from lemma 4.3 that there exists $\psi_d : S^R\left(\bar{k}\right) \to S^R\left(\bar{k}\right)$ such that $\psi_d\left(s^R\right) = s^R$ if s^R is constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, while given a Receiver strategy s^R nonconstant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, $\psi_d\left(s^R\right)$ takes action $b^R\left(a^S_{\phi(q)}\right)$ after receiving message $\hat{m}_2$ and action $b^R\left(a^S_{\phi(r)}\right)$ after receiving message $\tilde{m}$. By the language assumptions, $s^R\left(m_{\phi(N-l)}\right) = a^R_{\phi(N-l)}$ for every Receiver strategy s^R non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. It follows that

$$\begin{aligned} \psi_d\left(s^R\right)\left(m_{\phi(N-l)}\right) &= s^R\left(m_{\phi(N-l)}\right) \\ &\neq a^R_{\phi(N-l)} \end{aligned} \tag{4.5}$$

for every Receiver strategy s^R non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. By construction, the Sender strategy $\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right)$ belongs to $S^S\left(\bar{k}+1\right)$. Therefore, there exists a totally mixed Receiver strategy $\sigma^{*R} \in \Delta^+ S^R\left(\bar{k}\right)$ to which $\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right)$ is a best response. From claim 4.8, the construction that σ^{*R} is totally mixed, and the construction that $\psi_d\left(s^R\right)\left(\hat{m}_2\right) = b^R\left(a^S_{\phi(q)}\right)$ for any Receiver strategy s^R nonconstant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$,

it follows that

$$\begin{aligned} & u^S\left(\left(\hat{m}_2, a^S_{\phi(q)}\right), (1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\psi_d\left(\sigma^{*R}\right)\right) \\ > \; & u^S\left(\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right), (1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\sigma^{*R}\right) \\ \geq \; & u^S\left(\left(m, a^S_{\phi(q)}\right), (1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\sigma^{*R}\right) \\ = \; & u^S\left(\left(m, a^S_{\phi(q)}\right), (1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\psi_d\left(\sigma^{*R}\right)\right) \end{aligned} \tag{4.6}$$

for every message m outside of the message bundle $m_{\phi(N-l)} \cup M_{\phi(N-l)}$.

Since $\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right)$ is an action-strict best response to $\sigma^R_{flat} \in \Delta S^R\left(\bar{k}\right)$, for ε sufficiently small, any Sender best response to

$$(1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\psi_d\left(\sigma^{*R}\right)$$

involves taking action $a^S_{\phi(q)}$. Inequality 4.5 implies that

$$\begin{aligned} & u^S\left(\left(\hat{m}_2, a^S_{\phi(q)}\right), (1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\psi_d\left(\sigma^{*R}\right)\right) \\ > \; & u^S\left(\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right), (1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\psi_d\left(\sigma^{*R}\right)\right). \end{aligned}$$

It follows that every Sender best response to

$$(1-\varepsilon)\,\sigma^R_{flat} + \varepsilon\psi_d\left(\sigma^{*R}\right)$$

involves taking action $a^S_{\phi(q)}$ and sending a message in $M_{\phi(N-l)}$. Therefore, the *Always Incorrect* condition does not hold.

We have thus arrived at contradiction. The proof for statement 1 is then complete. □

Proof for Statement 2

Since $\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right) \in S^S\left(\bar{k}+1\right)$, there exists a totally mixed Receiver strategy $\hat{\sigma}^R$ in $S^R\left(\bar{k}\right)$ to which $\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right)$ is a best response. We just proved that statement 1 holds for $k = \bar{k}+1$. Therefore, $\hat{\sigma}^R$ must put positive probability on Receiver strategies that are non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. Suppose to the contrary that there exists no Sender strategy $\left(m_*, a^S_*\right) \in S^S\left(\bar{k}+1\right)$ where $m_* \in M_{\phi(N-l)}$ and

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

Let j be the last round of deletion after which a Sender strategy $\left(m, a^S\right)$ remains where message m belongs to $M_{\phi(N-l)}$ and a^S is either equal to $a^S_{\phi(q)}$ or maximizes

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

over $\left\{a^S_{\phi(i)} : i > N-l\right\}$. Let $\left(\hat{m}, \hat{a}^S\right)$ be such a Sender strategy in $S^S(j)$. Let E denote the message bundle in $M_{\phi(N-l)}$ that contains message $\hat{m}$.

Claim 4.11 (Non-exclusiveness of set E). *Given every $\left(\tilde{m}, \tilde{a}^S\right) \in S^S(j)$ where $\tilde{m} \in M_{\phi(N-l)}$ and either $\tilde{a}^S$ is equal to $a^S_{\phi(q)}$ or $\tilde{a}^S$ maximizes*

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

over $\left\{a^S_{\phi(i)} : i > N-l\right\}$, there exists a Receiver strategy $\tilde{s}^R \in S^R(j)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ where $\tilde{s}^R(\tilde{m}) \neq b^R\left(\tilde{a}^S\right)$.

Proof. Suppose this is not true for some $\left(\tilde{m}, \tilde{a}^S\right)$. Then every Receiver strategy in $S^R(j)$ that is non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ responds to message $\hat{m}_i$ with action $b^R\left(a^S_{\phi(r_i)}\right)$. Therefore,

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

The last inequality holds strictly because $\hat{\sigma}^R$ puts positive probability on Receiver strategies non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, and message $m_{\phi(N-l)}$ yields an action other than $b^R\left(a^S_{\phi(q)}\right)$ under such Receiver strategies. This contradicts the construction that $\left(m_{\phi(N-l)}, a^S_{\phi(q)}\right)$ is a best response to $\hat{\sigma}^R$.

It follows that there exists $s^R_{wrong,(\hat{m},\hat{a}^S)} \in S^R(j)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ such that $s^R_{wrong,(\hat{m},\hat{a}^S)}(\hat{m}) \neq b^R\left(\hat{a}^S\right)$. From the characterization of $S^S(1)$, it has to be the case that $l \geq 3$. Therefore, there are at least three parallel message bundles in $M_{\phi(N-l)}$. Let F_1 and F_2 be two parallel message bundles in $M_{\phi(N-l)}$ different from E. From lemma 4.5, $S^S(j)$ contains $\left(m_1, a^S_1\right)$ and $\left(m_2, a^S_2\right)$ where $m_1 \in F_1$, $m_2 \in F_2$ and both a^S_1 and a^S_2 maximize $\chi\left(\hat{\sigma}^R, m_{\phi(N-l)} \cup M_{\phi(N-l)}, a^S\right)$ over $\left\{a^S_{\phi(i)} : i > N-l\right\}$. From claim 4.11, for $i = 1, 2$, there exists $s^R_{wrong,\left(m_i,a_i^S\right)} \in S^R(j)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ such that $s^R_{wrong,\left(m_i,a_i^S\right)}(m_i) \neq b^R\left(a^S_i\right)$. Then either $a^S_i \neq \hat{a}^S$, and thus there exists $s^R_i \in S^R(j)$ non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ such that $s^R_i(m_i) = b^R\left(a^S_i\right)$ and thus $s^R_i|_{F_i}$ is not equivalent to a constant of $b^R\left(\hat{a}^S\right)$, or $s^R_{wrong,\left(m_i,a_i^S\right)}(m_i) \neq b^R\left(\hat{a}^S\right)$, and $s^R_{wrong,\left(m_i,a_i^S\right)}|_{F_i}$

is not equivalent to a constant of $b^R\left(\hat{a}^S\right)$ since

$$\begin{aligned} & u^R\left(\left(m_i, a_i^S\right), s^R_{wrong,\left(m_i,a_i^S\right)}\right) \\ = \; & g^R\left(a_i^S, s^R_{wrong,\left(m_i,a_i^S\right)}\left(m_i\right)\right) \\ \neq \; & g^R\left(a_i^S, b^R\left(\hat{a}^S\right)\right). \end{aligned}$$

It follows from lemma 4.4 that there exists $\psi_d : S^R(j) \to S^R(j)$ such that $\psi_d\left(s^R\right) = s^R$ for Receiver strategies constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, and for Receiver strategies nonconstant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, $\psi_d\left(s^R\right)$ is equal to s^R outside of $m_{\phi(N-l)} \cup M_{\phi(N-l)}$, while $\psi_d\left(s^R\right)(\hat{m}) = b^R\left(\hat{a}^S\right)$ for s^R non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$. $\psi_d\left(\hat{\sigma}^R\right)$ puts positive probability on Receiver strategies non-constant on $m_{\phi(N-l)} \cup M_{\phi(N-l)}$ because $\hat{\sigma}^R$ does. Therefore,

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

for every Sender strategy $\left(m, a^S\right)$ where $m \notin M_{\phi(N-l)}$. Therefore, if $\left(m_*, a^S_*\right)$ is a best

response to $\psi_d\left(\hat{\sigma}^R\right)$, it has to be the case that $m_* \in M_{\phi(N-l)}$ and a_*^S maximizes

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

over $\left\{a^S_{\phi(i)} : i > N-l\right\}$. Since at least one such Sender strategy survives the $(j+1)^{th}$ round of deletion, this contradicts the construction of j. We have thus completed the proof for statement 2. □

□

Bibliography

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

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

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

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

Brandenburger, A., and H.J. Keisler, "An Impossibility Theorem on Beliefs in Games," mimeo, 2004. Available at www.stern.nyu.edu/~abranden.

Crawford, V., and J. Sobel, "Strategic Information Transmission," Econometrica, 50(6), 1982, 1431-1451.

Farrell, J. "Meaning and Credibility in Cheap-Talk Games," Games and Economic Behavior, 5(4), 1993, 514-531.

Fudenberg, D. and J. Tirole, *Game Theory*, The MIT Press, 1991.

Kim, Y.-G. and J. Sobel, "An Evolutionary Approach to Pre-play Communication," Econometrica, 63(5), 1995, 1181-1193.

Kreps, D., and R. Wilson, "Sequential Equilibria," Econometrica, 50(4), 1982, 863-894.

Pearce, D.G. "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52(4), 1984, 1029-1050.

Rabin, M. "Communication between Rational Agents," Journal of Economic Theory, 51, 1990, 144-179.

Warneryd, K. "Cheap Talk, Coordination, and Evolutionary Stability," Games and Economic Behavior, 5, 1993, 532-546.

Zapater, I. "Credible Proposals in Communication Games," Journal of Economic Theory, 72, 1997, 173-197.