Persuasive Selection in Signaling Games*

Haoyuan Zeng†

November 1, 2025

Latest version here.

Abstract

This paper introduces a novel criterion, persuasiveness, to select equilibria in signaling games. In response to the Stiglitz critique, persuasiveness focuses on the comparison across equilibria. An equilibrium is more persuasive than an alternative if the set of types of the sender who prefer the alternative would sequentially deviate to the former once other types have done so—that is, if an unraveling occurs. Persuasiveness has strong selective power: it uniquely selects an equilibrium outcome in monotone signaling games. Moreover, in non-monotone signaling games, persuasiveness refines predictions beyond existing selection criteria. Notably, it can also select equilibria in cheap-talk games, where standard equilibrium refinements for signaling games have no selective power.

Keywords: Equilibrium Selection, Persuasiveness, Signaling Games, Stiglitz Critique
JEL Codes: C70, C72, D82

*First draft: Oct 2025. I thank my advisor, Marek Pycia, for his guidance and support throughout this project. For their comments, I thank Christian Ewerhart, Jindi Huang, Nick Netzer, Armin Schmutzler, Joel Watson, and talk audiences at the University of Zurich, the 9th Swiss Theory Day. All errors are my own.

†University of Zurich. Email: haoyuan.zeng@econ.uzh.ch.

1 Introduction

Many economic interactions can be modeled as a signaling game, where an informed sender sends a message to an uninformed receiver. The receiver responds by taking an action that is payoff-relevant to both players. The seminal paper on signaling games is Spence (1973)'s job market signaling model. Since then, a vast body of literature has applied signaling games across a wide range of fields, including advertising, bargaining, finance, industrial organization, and reputation.[1]

It is well-known that signaling games often lead to many sequential equilibria (Kreps and Wilson, 1982). The multiplicity of equilibria limits the usefulness of the model in analyzing the underlying economic problem, as it fails to yield precise predictions regarding the outcome of strategic interactions. Moreover, for some problems at least, many equilibria seem implausible due to the beliefs associated with messages off the equilibrium path, i.e., off-path beliefs.

A number of equilibrium refinements have been proposed to address the issue of equilibrium multiplicity in signaling games. The classic approach to refining equilibria in signaling games is to formalize plausible restrictions on off-path beliefs, such as the intuitive criterion (Cho and Kreps, 1987) and the divinity criterion (Banks and Sobel, 1987), which are based on the concepts of strategic stability and forward induction (Kohlberg and Mertens, 1986). The idea is that if sending an off-path message can be interpreted as a signal of certain types of the sender who would be better off by sending this message than following the proposed equilibrium, then this equilibrium fails to pass the equilibrium selection criterion. It is well recognized that this type of reasoning is imperfect.

> *...the justification for these restrictions are that they are more "intuitive"; they go with stories about how players might reason at off-path information sets. I put scare quotes around "intuitive," because what is and isn't intuitive is largely subjective; it is your job to judge which of these restrictions appeals to your intuition and which do not.* (Kreps, 2023, p. 647)

In particular, it has been criticized by Joseph Stiglitz (Cho and Kreps, 1987, p. 202) for logical inconsistencies stemming from restricting off-path beliefs while holding on-path beliefs fixed. Such inconsistencies lead to the fact that the intuitive criterion always selects the Pareto-dominant separating equilibrium outcome (the Riley outcome (Riley, 1979)) in a

[1]For a comprehensive survey of the signaling games literature, see Riley (2001) and Sobel (2009). Applications of signaling games span several fields: in advertising, see Nelson (1974); in bargaining, see Fudenberg and Tirole (1983) and Sobel and Takahashi (1983); in finance, see Leland and Pyle (1977) and John and Williams (1985); in industrial organization, see Milgrom and Roberts (1982); and in reputation, see Barro and Gordon (1983).

two-type Spencian model, regardless of how unlikely it is that the worker is of low type. In other words, even when asymmetric information is nearly absent from the firm's perspective—because the low-type worker is extremely rare—the high-type worker nevertheless separates by attaining a high level of education. This separating outcome is substantially worse than the pooling equilibrium outcome, in which the firm would treat them almost as if they were high-type without requiring any costly education. We elaborate on why the intuitive criterion eliminates the pooling equilibrium in Example 1 below, where we discuss the Stiglitz critique in detail.

The Stiglitz critique points out that the meaning of messages (beliefs)—whether on-path or off-path—should always be interpreted in equilibrium. As a result, imposing restrictions on off-path beliefs generally also affects on-path beliefs. When both beliefs are consistent with each other, they constitute equilibrium beliefs in an alternative equilibrium. Thus, equilibrium selection reduces to a comparison across alternative equilibria. Based on this idea, Mailath, Okuno-Fujiwara and Postlewaite (1993) introduce the concept of undefeated equilibrium, which is immune to the Stiglitz critique. Nevertheless, in monotone signaling games, such as the job market signaling model (Spence, 1973), their criterion typically fails to select a unique equilibrium outcome.

This paper introduces a novel criterion to select equilibria in signaling games, which is called *persuasiveness*. In response to the Stiglitz critique, persuasiveness focuses on the comparison across equilibria. Because different types of the sender may obtain different payoffs in distinct equilibria, they may prefer different equilibria. Accordingly, it is reasonable for the receiver to expect that each type of the sender would like to play the equilibrium which gives them a higher payoff. Persuasiveness formalizes this logic of forward induction by considering how the receiver interprets messages across different equilibria.

To illustrate, consider the two-type Spencian model introduced earlier. When the low-type worker is sufficiently unlikely, both types of the worker prefer the pooling equilibrium to the separating equilibrium, because it yields a higher equilibrium payoff for each type of the worker. Hence, it is reasonable for the firm to expect that both types of the worker would like to play the pooling equilibrium rather than the separating equilibrium. Upon observing a zero education level, the firm therefore finds it more *persuasive* to interpret this message according to the pooling equilibrium—believing that both types choose zero education—than according to the separating equilibrium, in which only the low-type worker does so. In this sense, the pooling equilibrium is more *persuasive* than the separating equilibrium.

More generally, let σ be the putative equilibrium. The problem of equilibrium selection can be conceptualized as the receiver posing the following question to themselves: is there a message $\overline{m}$ on-path of an alternative equilibrium $\overline{\sigma}$ for which it is more *persuasive* to

interpret $\overline{m}$ in $\overline{\sigma}$ than what σ prescribes? Interpreting $\overline{m}$ in the context of $\overline{\sigma}$ follows directly from the Stiglitz critique, which emphasizes that the meaning of a message should always be interpreted in equilibrium. In the two-type Spencian model, where the low-type worker is sufficiently unlikely, σ corresponds to the separating equilibrium, $\overline{\sigma}$ to the pooling equilibrium, and $\overline{m} = 0$ to the zero education level.

In order for $\overline{\sigma}$ to serve as a challenger to σ, there must exist at least one type of the sender who strictly prefers $\overline{\sigma}$ to σ—that is, whose equilibrium payoff in $\overline{\sigma}$ is strictly higher than in σ. In the absence of such a type, every type of the sender who sends $\overline{m}$ in $\overline{\sigma}$ would instead prefer σ to $\overline{\sigma}$, implying that the receiver should not expect $\overline{m}$ to be sent in $\overline{\sigma}$. When such a type exists, the receiver reasons about the set of types of the sender who send $\overline{m}$ in $\overline{\sigma}$ by dividing them into two groups. The first group consists of the set of types of the sender who prefer the challenger $\overline{\sigma}$ to the putative equilibrium σ, while the second group consists of the set of types of the sender who strictly prefer σ to $\overline{\sigma}$.

When the second group is empty, $\overline{\sigma}$ is more *persuasive* than σ in the sense that every type of the sender who sends $\overline{m}$ in $\overline{\sigma}$ prefers $\overline{\sigma}$ to σ, and the receiver should expect them to do so. In the two-type Spencian model, where the low-type worker is sufficiently unlikely, the pooling equilibrium $\overline{\sigma}$ can challenge the separating equilibrium σ with an alternative interpretation of the zero education level $\overline{m} = 0$—namely, that this message $\overline{m}$ is sent by both types rather than solely by the low type. In this scenario, both types of the worker belong to the first group, while the second group is empty. The pooling equilibrium is therefore more *persuasive* than the separating equilibrium, as both types of the worker prefer the pooling equilibrium to the separating equilibrium.

When the second group is non-empty (see Example 2 below), we check whether the interpretation of $\overline{m}$ in $\overline{\sigma}$ rather than in σ provides a rationale for why every type of the sender in the second group would ultimately prefer to play the challenger $\overline{\sigma}$ rather than the putative equilibrium σ, even though they initially prefer σ. Consider the simple case where the second group contains only one type t'. When the receiver observes a message m' sent by type t' in σ, the receiver expects that this message is not sent by the set of types of the sender in the first group, because they would prefer to send $\overline{m}$ in $\overline{\sigma}$ instead. If, relative to the equilibrium payoff that type t' would obtain in the challenger $\overline{\sigma}$, type t' is worse off when sending the message m' in σ—given that the receiver best responds to m' under the conditional belief that excludes all types in the first group—then type t' will also have an incentive to deviate to $\overline{\sigma}$, even though they initially prefer σ. In this case, the second group *unravels* as a result of the first group's deviation.

When the second group contains multiple types, we examine them sequentially according to a specified ranking. For each type t' in this group, we compare their equilibrium payoff

in the challenger $\overline{\sigma}$ with the payoff they would obtain by sending some message $\tilde{m}'$ in σ, assuming the receiver best responds to $\tilde{m}'$ under the conditional belief that excludes all types in the first group, as well as all types ranked above t'. If, under this belief, type, type t' is worse off in σ than in $\overline{\sigma}$, then type t' will also have an incentive to deviate to $\overline{\sigma}$. In this case, we say that the interpretation of $\overline{m}$ in $\overline{\sigma}$ rather than in σ triggers an *unraveling* (Definition 2).

The unraveling starts with the highest-ranked type, who finds it better off deviating to $\overline{\sigma}$ than playing σ if the receiver believes only the second group would play σ. Conditional on higher-ranked types having deviated, the unraveling then proceeds to the highest-ranked undeviated type, who similarly prefers deviating to $\overline{\sigma}$ if the receiver believes only the remaining undeviated types in the second group would play σ. This iterative argument continues until all types in the group have deviated. The reasoning is similar to the classical analysis of voluntary disclosure in Grossman and Hart (1980), Grossman (1981), Milgrom (1981), and Verrecchia (1983).

Unraveling provides the receiver with a rationale for why every type of the sender who sends $\overline{m}$ in $\overline{\sigma}$ would ultimately prefer to play $\overline{\sigma}$ rather than σ, regardless of their initial preference. It is thus more persuasive to interpret $\overline{m}$ in $\overline{\sigma}$ rather than in σ. Note that we have a trivial unraveling when the second group is empty. Hence, we say that $\overline{\sigma}$ is *more persuasive* than σ if there exists a message $\overline{m}$ on the equilibrium path of $\overline{\sigma}$ such that the interpretation of $\overline{m}$ in $\overline{\sigma}$ rather than in σ triggers an unraveling.

An equilibrium $\overline{\sigma}$ is *most persuasive* if $\overline{\sigma}$ is more persuasive than any other equilibrium σ that is not payoff-equivalent for the sender. In other words, the most persuasive equilibrium $\overline{\sigma}$ can always challenge any other equilibrium with a new interpretation of some message such that every type of the sender who sends that message in $\overline{\sigma}$ would like to deviate to that message and play $\overline{\sigma}$. If the most persuasive equilibrium is unique up to payoff equivalence for the sender, then no other equilibrium can challenge the most persuasive equilibrium with new interpretations of any messages. Hence, when there exists a unique most persuasive equilibrium, the interpretations of all messages in the game are determined, in the sense that no other equilibrium can provide a more persuasive interpretation of any message.

Unlike previous equilibrium refinements, our selection criterion emphasizes how the receiver interprets messages rather than how the sender signals through off-path messages. In particular, it does not require the receiver to detect that the sender is deviating from σ by sending an off-path message. Instead, by introspection, the receiver finds it more persuasive to interpret $\overline{m}$ in $\overline{\sigma}$ rather than in σ. For instance, in the two-type Spencian model, the zero education level is on-path of the separating equilibrium. By introspection, the firm finds it more persuasive to interpret this message as being sent by both types when the

low-type worker is sufficiently unlikely. Persuasiveness therefore has selective power even when $\overline{m}$ is on the equilibrium path of σ (see also Example 3 below). The logic of forward induction underlying persuasiveness is that the receiver expects the sender would like to play the equilibrium which gives them a higher payoff—a natural assumption given that the sender moves first.

Persuasiveness has strong selective power. In monotone signaling games, such as the job market signaling model (Spence, 1973), it uniquely selects the equilibrium outcome that provides the highest type of the sender with the maximum equilibrium payoff (lexicographically maximum outcome). Next, we illustrate by examples that persuasiveness has stronger selective power than other equilibrium refinements in some non-monotone signaling games (see Table 4 for a summary of the comparison). We also discuss the limitations of persuasiveness, noting that cyclicality can emerge in the absence of a unique most persuasive equilibrium. We explain why it may be reasonable to consider the *least persuasive* equilibrium in some cases. Lastly, we demonstrate that persuasiveness can have good selective power even in cheap-talk games, where standard equilibrium refinements for signaling games have no selective power.

The paper is organized as follows. Section 2 introduces the setup. Section 3 discusses the Stiglitz critique and formally defines persuasiveness. Section 4 studies persuasiveness in monotone signaling games. Section 5 discusses the extensions and limitations of our selection criterion. Section 6 reviews the related literature. Section 7 concludes the paper. All proofs are relegated to Appendix A. In Appendix B, we offer an intuitive explanation of some equilibrium refinements in the previous literature.

2 Setup

The signaling game G is described as follows. There is a sender (S) and a receiver (R). It is common knowledge that Nature draws the type t of the sender from a non-empty set of types T according to a prior probability distribution $p \in \Delta(T)$ with full support. The sender privately observes t and chooses a message m from a non-empty set of messages M. The receiver only observes the message m and chooses an action a from a non-empty set of actions A. Both the sender's and the receiver's payoffs depend on the message, the action, and the sender's type, i.e., $u_S, u_R : T \times M \times A \to \mathbb{R}$.

A strategy for the sender σ_S maps types to distributions over messages, i.e., $\sigma_S : T \to \Delta(M)$. Let $\sigma_S(m|t)$ denote the probability of the type t sender choosing the message m. Upon receiving the sender's message, the receiver updates their belief μ over the sender's types, i.e., $\mu : M \to \Delta(T)$. Let $\mu(t|m)$ denote the receiver's posterior belief about the

sender being type t after receiving the message m. A strategy for the receiver σ_R maps messages to distributions over actions, i.e., $\sigma_R : M \to \Delta(A)$. Let $\sigma_R(a|m)$ denote the probability of action a being played upon receiving the message m. Let $\text{BR}(m, \mu)$ be the set of actions that are best responses to the message m given the belief μ. That is,

$$\text{BR}(m, \mu) = \arg\max_{a \in A} \sum_{t \in T} u_R(t, m, a)\, \mu(t|m).$$

Let $\text{supp}(\cdot)$ be the support of a function, i.e., the set of points at which the function is non-zero. For example, if $m \in \text{supp}(\sigma_S(t))$, then $\sigma_S(m|t) > 0$.

A sequential equilibrium σ is a strategy-belief profile

$$\sigma = (\sigma_S(m|t), \sigma_R(a|m), \mu(t|m))_{t \in T, m \in M}$$

that satisfies sequential rationality and consistency.[2]

Definition 1. $\sigma = (\sigma_S, \sigma_R, \mu)$ is a *sequential equilibrium* if,

1. Sequential Rationality: $\forall t \in T$,

$$\text{supp}(\sigma_S(t)) \subseteq \arg\max_{m \in M} \sum_{a \in A} u_S(t, m, a)\, \sigma_R(a|m),$$

   where $\text{supp}(\sigma_R(m)) \subseteq \text{BR}(m, \mu)\ \ \forall m \in M$.

2. Consistency: $\forall t \in T$, $\forall m \in M$,

$$\mu(t|m) = \frac{\sigma_S(m|t)\, p(t)}{\sum_{t' \in T} \sigma_S(m|t')\, p(t')}$$

   conditional on $\sum_{t' \in T} \sigma_S(m|t')\, p(t') > 0$.

Let $\text{SE}(G)$ denote the set of sequential equilibria in the game G. Let $\text{PSE}(G)$ denote the set of pure-strategy sequential equilibria in the game G. Let $u_S(t, \sigma)$ denote the expected payoff of the type t sender in the equilibrium σ.

A message m is off the equilibrium path of σ if $\sum_{t \in T} \sigma_S(m|t)\, p(t) = 0$, i.e., no type of the sender sends m with positive probability in σ. We also call m an off-path message (with respect to σ). The belief $\mu(\cdot|m)$ of the receiver after observing m is called an off-path belief if m is off the equilibrium path of σ.

[2]Since the signaling game is a two-period game with observed actions and independent types, any perfect Bayesian equilibrium is also a sequential equilibrium (Fudenberg and Tirole, 1991).

3 Persuasive Selection

In this section, we start with the intuitive criterion (Cho and Kreps, 1987), one of the most widely used equilibrium refinements in signaling games. Using Example 1, we illustrate how it is subject to the Stiglitz critique—a limitation that also applies to related selection criteria (Banks and Sobel, 1987; Grossman and Perry, 1986). Then, we introduce the concept of persuasiveness, which is immune to the Stiglitz critique.

The intuitive criterion is based on the idea of forward induction (Kohlberg and Mertens, 1986): if sending an off-path message can be interpreted as a signal of certain types of the sender who would be better off by sending this message than following the proposed equilibrium, then this equilibrium fails to pass the intuitive criterion. Verifying whether an equilibrium σ fails the intuitive criterion involves the following two steps.[3]

- Step 1: Which types of the sender *could benefit* by sending an off-path message m?

  We denote the set of such types as D. Formally,

  $$D = \left\{ t \in T | \, u_S\left(t, \sigma\right) \leq \max_{a \in \mathrm{BR}(m, \Delta(T))} u_S\left(t, m, a\right) \right\},$$

  where $\mathrm{BR}\left(m, \Delta\left(T\right)\right) = \cup_{\mu \in \Delta(T)} \mathrm{BR}\left(m, \mu\right)$.

- Step 2: If deviations only come from the set of types of the sender identified in Step 1, is the *lowest* payoff from deviating higher than their equilibrium payoff for some type of the sender?

  Formally, if there exists $t \in D$ such that

  $$\min_{a \in \mathrm{BR}(m, \Delta(D))} u_S\left(t, m, a\right) > u_S\left(t, \sigma\right),$$

  then this equilibrium σ fails the intuitive criterion.

The reader is referred to Cho and Kreps (1987) for a detailed discussion of the intuitive criterion. The authors are well aware of the subtleties involved.

> *"Despite the name we have given it, the Intuitive Criterion is not completely intuitive."* (Cho and Kreps, 1987, p. 202)

Here, we highlight that Step 1 performs a *best-case* scenario analysis for the sender, while Step 2 performs a *worst-case* scenario analysis. In Step 1, we choose the *best* action for the

[3]This explanation of the intuitive criterion follows from Munoz-Garcia and Espinola-Arredondo (2011).

sender as long as the receiver's action is a best response to some belief—equivalently, we select the *most favorable* receiver's belief for the sender. In Step 2, by contrast, we choose the *worst* action for the sender—equivalently, we select the *least favorable* receiver's belief for the sender. Consequently, beliefs jump from one extreme to the other when moving from Step 1 to Step 2. This asymmetry in the treatment of beliefs underlies the Stiglitz critique, as illustrated in Example 1 below. As we will see later, it also accounts for the intuitive criterion's weak selective power among pooling equilibria (see below Example 5).

3.1 The Stiglitz Critique

We consider a simple two-type version of Spence (1973)'s job market signaling model.

Example 1 (Two-Type Spencian Game). There are two types of a worker (sender), $T = \{t_L, t_H\}$, where $t_L = 1$ and $t_H = 2$. We call t_L the low-type worker, and t_H the high-type worker. The prior probability that the worker is low-type is $p \in (0,1)$. The worker chooses an education level $m \in M = [0, \infty)$. The cost of acquiring education is $c(t,m) = \frac{m}{t}$. A firm (receiver) observes the education level of the worker and offers a wage of $a \in A = \mathbb{R}_+$ in a competitive market. In equilibrium, the firm always earns zero profit, i.e., $\sigma_R(m) = \mathbb{E}_\mu[t|\, m]$.[4] The payoff of a worker of type t who chooses an education level m and receives a wage of a is given by $u_S(t,m,a) = a - c(t,m)$.

It is well-known that this game has a Pareto-dominating separating equilibrium, i.e., the Riley equilibrium σ^{Riley} (Riley, 1979), in which the low-type worker chooses $m = 0$ and the high-type worker chooses $m = 1$. The firm offers a wage of $a = 1$ to a worker with $m = 0$ and $a = 2$ to a worker with $m = 1$. The payoffs of the low-type and the high-type workers in this equilibrium are $u_S\left(t_L, \sigma^{\text{Riley}}\right) = 1$ and $u_S\left(t_H, \sigma^{\text{Riley}}\right) = 1.5$, respectively.

This game also has a pooling equilibrium σ^{Pooling}, in which both types of workers choose $m = 0$ and the firm offers an expected wage of $a = \mathbb{E}_{\mu^{\text{Pooling}}}[t|\, m = 0] = 2 - p$. The payoffs of the low-type and the high-type workers in this equilibrium are $u_S\left(t_L, \sigma^{\text{Pooling}}\right) = u_S\left(t_H, \sigma^{\text{Pooling}}\right) = 2 - p$.[5]

For both equilibria, after an off-path message, the firm believes that the worker is low-type with probability one. We can summarize the equilibrium strategies and payoffs of the low-type and high-type workers in the two equilibria in Table 1.

The intuitive criterion uniquely selects the Riley outcome in σ^{Riley} irrespective of p (Cho

[4] Strictly speaking, we should model this as a game with two firms, who would then engage in a Bertrand competition for the worker. A single firm with the payoff function $u_R(t,m,a) = -(t-a)^2$ yields the similar behavior.

[5] There exist other equilibria, but they are not relevant for our discussion here because they are generally deemed implausible by any equilibrium refinement.

	t_L	t_H		t_L	t_H
$\sigma_S^{\text{Riley}}(t)$	0	1	$u_S\left(t, \sigma^{\text{Riley}}\right)$	1	1.5
$\sigma_S^{\text{Pooling}}(t)$	0	0	$u_S\left(t, \sigma^{\text{Pooling}}\right)$	$2-p$	$2-p$

Table 1: Equilibrium Strategies and Payoffs of the Worker in Example 1

and Kreps, 1987).[6] Notice that when p is close to zero, i.e., the low-type worker is very unlikely, the pooling outcome in σ^{Pooling} Pareto-dominates the Riley outcome in σ^{Riley} for each type of the worker. Intuitively, when p is close to zero, the firm offers a wage close to 2 in the pooling equilibrium. The high-type worker does not want to incur a cost to separate themselves from the low-type worker, because it is almost the same as if there were no low-type worker at all. In particular, when $p = 0$, the pooling outcome degenerates to the equilibrium outcome when there is no low-type worker.[7] It seems counterintuitive that the pooling equilibrium fails the intuitive criterion, even though its outcome Pareto-dominates the Riley outcome for each type of the worker. This raises the question of why the intuitive criterion rejects the pooling equilibrium when p approaches zero.

To see this, consider an off-path education level $m' \in (p, 2p)$. In Step 1 of the intuitive criterion, the firm would at most offer a wage of 2 after observing m'. The payoff of the low-type worker if deviating to m' is at most $2 - m' < 2 - p$, while the payoff of the high-type worker if deviating to m' is at most $2 - \frac{m'}{2} > 2 - p$. Then, only the high-type worker could benefit by deviating to m'. In Step 2, since the firm believes that only the high-type worker would choose m', the firm would offer a wage of 2 and the high-type worker would benefit from this deviation. Hence, the pooling equilibrium fails the intuitive criterion.

The Stiglitz critique points out that the above reasoning is flawed. The two steps of the intuitive criterion are usually motivated as the high-type worker making an implicit "speech" to the firm by deviating to m'. However, this "speech" induces an inconsistent "story" when taking into account the low-type worker. Following the logic of the intuitive criterion, the firm would now believe that the worker is low-type with probability one after observing $m = 0$ because the high-type worker would have deviated to m'. Then, the low-type worker would not keep choosing $m = 0$ as in the pooling equilibrium because the firm would only

[6]The intuitive criterion uniquely selects the Riley outcome instead of the Riley equilibrium σ^{Riley} itself. There can be multiple Riley equilibria which produce exactly the same Riley outcome on path. However, they can differ on the off-path beliefs, which cannot be uniquely pinned down in general.

[7]In contrast, the Riley outcome stays the same as long as $p > 0$. When $p = 0$, the equilibrium payoff of the high-type worker jumps from 1.5 to 2. There is generally a discontinuity in the Riley outcome as one of the prior probabilities goes to zero for any finite type space (Mailath, Okuno-Fujiwara and Postlewaite, 1993).

offer a wage of 1 after observing $m = 0$. Instead, the low-type worker has an incentive to deviate to m' and and imitate the high-type worker, as this deviation yields a payoff of $2 - m' > 1$ when p is close to zero. If the firm anticipates the low-type worker's response after the high-type worker's deviation, the firm would offer an expected wage of $2 - p$ instead of 2. As a result, the high-type worker would not want to deviate to m', because they incur a cost by choosing $m' > 0$ but the wage is the same as before in the pooling equilibrium, which invalidates the reason why the pooling equilibrium fails the intuitive criterion.

These logical inconsistencies in the intuitive criterion stem from restricting off-path beliefs while holding on-path beliefs fixed. In Example 1, the intuitive criterion postulates that the off-path belief of the firm after observing m' is that the worker is high-type with probability one. However, it also implicitly assumes that the on-path belief of the firm after observing $m = 0$ is that the worker is high-type with probability $1 - p$. Apparently, these two beliefs cannot hold simultaneously. This implies that when we impose restrictions on off-path beliefs, consistency requires further adjustments on on-path beliefs. When these beliefs are consistent with each other, we are effectively looking at another equilibrium. Hence, to address the Stiglitz critique, we should always interpret the meaning of messages (beliefs after observing $m = 0$ or m')—whether on-path or off-path—in equilibrium. Equilibrium selection concerns identifying the most appropriate interpretation of a given message among multiple equilibria.

In Example 1, when the firm interprets m' in equilibrium, the firm can at most offer an expected wage of $2 - p$.[8] Hence, the high-type worker would not benefit from deviating to m', and strictly prefers to play the pooling equilibrium. Given this, how should we conceptualize the selection between the pooling equilibrium and the Riley equilibrium?

Note that the low-type worker always chooses $m = 0$ in both equilibria. Then, the high-type worker is effectively facing the decision of whether to separate from the low-type worker by choosing $m = 1$ or to pool with the low-type worker by choosing $m = 0$. When $p > 0.5$, the high-type worker strictly prefers separation to pooling. When $p \leq 0.5$, the high-type worker weakly prefers pooling to separation.[9] Since the worker moves first, it is reasonable for the firm to expect the high-type worker to go with the choice that gives them a higher payoff. As a result, when $p > 0.5$, the firm should interpret $m = 0$ in the Riley equilibrium rather than in the pooling equilibrium; when $p \leq 0.5$, the firm should interpret $m = 0$ in the pooling equilibrium rather than in the Riley equilibrium. Since $m = 0$ is on-path of both equilibria, the reasoning is not about how the high-type worker

[8]This wage is consistent with another pooling equilibrium at m'.

[9]When $p = 0.5$, the high-type worker is indifferent between separation and pooling, but the low-type worker always strictly prefers pooling to separation. In such a case, we select the weakly Pareto-dominating equilibrium, i.e., the pooling equilibrium.

signals through off-path education levels as in the intuitive criterion but rather how the firm expects what the high-type worker would do. The interpretation of $m = 0$ changes with the equilibrium.

The above reasoning motivates a selection criterion that neither selects the pooling outcome nor the Riley outcome irrespective of p. Instead, we select the equilibrium outcome that gives the high-type worker a higher payoff, which we will call the lexicographically maximum outcome (lex max outcome) later (Definition 7). We call the equilibrium that produces this outcome the lexicographically maximum sequential equilibrium (LMSE). In Example 1, the LMSE is the Riley equilibrium when $p > 0.5$ and the pooling equilibrium when $p \leq 0.5$. To address the Stiglitz critique, the selection criterion should build on how the receiver interprets messages in different equilibria and finding a good interpretation among those, which will be formalized as the concept of *persuasiveness* now.

3.2 Persuasiveness

To illustrate the logic of persuasiveness, we consider a three-type version of Spence (1973)'s job market signaling model used in Mailath, Okuno-Fujiwara and Postlewaite (1993).

Example 2 (Three-Type Spencian Game). We follow the setup of Example 1 except that there are three types of a worker, $T = \{t_L, t_M, t_H\}$, where $t_L = 1$, $t_M = 2$, and $t_H = 3$. We call t_L the low-type worker, t_M the medium-type worker, and t_H the high-type worker. The prior probabilities are $p(t_L) = 0.35$, $p(t_M) = 0.20$, and $p(t_H) = 0.45$.

We focus on the following equilibria with different information revealed, and summarize in Table 2 below the equilibrium strategies and payoffs of the low-type, medium-type, and high-type workers in these equilibria, presented in the same format as Example 1.[10] For instance, in the equilibrium σ^1, the low-type and medium-type workers choose $m = 0$, while the high-type worker chooses $m = 3.27$, which is the minimum level of education the high-type worker has to choose in order to separate themselves from the low-type and medium-type workers. The firm offers an expected wage of $a = \mathbb{E}_{\mu^1}[t|\, m = 0] = \frac{0.35\times 1 + 0.2\times 2}{0.35+0.2} = 1.36$ to a worker with $m = 0$ and $a = 3$ to a worker with $m = 3.27$. The payoffs of the low-type, medium-type, and high-type workers in this equilibrium are $u_S(t_L, \sigma^1) = u_S(t_M, \sigma^1) = 1.36$, and $u_S(t_H, \sigma^1) = 1.91$, respectively. The other equilibria can be similarly interpreted. In particular, the equilibrium $\overline{\sigma}$ is the LMSE, where the high-type worker attains the highest equilibrium payoff.

Note that the intuitive criterion has limited selective power when there are more than two types. The D1 criterion is applied more often instead, because it uniquely selects the Riley

[10] To make the comparison easier, we round numbers to two decimal places when necessary.

	t_L	t_M	t_H		t_L	t_M	t_H
$\sigma_S^1(t)$	0	0	3.27	$u_S(t,\sigma^1)$	1.36	1.36	1.91
$\sigma_S^{\text{Riley}}(t)$	0	1	3	$u_S\left(t,\sigma^{\text{Riley}}\right)$	1	1.5	2
$\sigma_S^{\text{Pooling}}(t)$	0	0	0	$u_S\left(t,\sigma^{\text{Pooling}}\right)$	2.1	2.1	2.1
$\overline{\sigma}_S(t)$	0	1.1	1.1	$u_S(t,\overline{\sigma})$	1	1.85	2.13

Table 2: Equilibrium Strategies and Payoffs of the Worker in Example 2

outcome (Cho and Sobel, 1990).[11] Because the D1 criterion is also subject to the Stiglitz critique for the same reason as the intuitive criterion, we see in this example again that the Riley outcome is uniquely selected even when it is Pareto-dominated by the pooling outcome in σ^{Pooling} for each type of the worker. However, we will not argue that the pooling outcome should be selected instead of the Riley outcome in this case as in Example 1. Instead, we will show that the LMSE $\overline{\sigma}$ is *more persuasive* than both the Riley equilibrium σ^{Riley} and the pooling equilibrium σ^{Pooling}, and hence the lex max outcome should be selected.

It is easy to see that the LMSE $\overline{\sigma}$ Pareto-dominates the Riley equilibrium σ^{Riley} for the medium-type and high-type workers. Given σ^{Riley}, we consider what happens when the medium-type and high-type workers both deviate to choose $\overline{m} = 1.1$. When the firm interprets $\overline{m}$ in $\overline{\sigma}$, the firm would offer the corresponding equilibrium wage in $\overline{\sigma}$, and both types of workers would indeed be better off by deviating to $\overline{m}$ than following the Riley equilibrium. In this sense, we say that $\overline{\sigma}$ is *more persuasive* than σ^{Riley}.

However, the above reasoning does not work when comparing the LMSE $\overline{\sigma}$ with the pooling equilibrium σ^{Pooling}, because the high-type worker prefers $\overline{\sigma}$ to σ^{Pooling}, while the medium-type worker prefers σ^{Pooling} to $\overline{\sigma}$. To argue that $\overline{\sigma}$ is still *more persuasive* than σ^{Pooling}, we first consider what could happen to the medium-type worker when they do not deviate to $\overline{m}$ and instead continue to send $m' = 0$, as in σ^{Pooling}, while the high-type worker deviates to $\overline{m}$.

As in Example 1, the firm expects that the worker would like to play the equilibrium which gives them a higher payoff. Therefore, upon observing $m' = 0$, the firm would no longer interprets this message as part of the pooling equilibrium, as the high-type worker has deviated to $\overline{m}$. The firm would infer that a worker choosing $m' = 0$ must be either medium-type or low-type—precisely as in σ^1. Consequently, the medium-type worker would obtain the same payoff as in σ^1 (1.36), which is lower than the payoff obtained in $\overline{\sigma}$ (1.85)

[11] See Munoz-Garcia and Espinola-Arredondo (2011) for a detailed explanation of why the intuitive criterion cannot uniquely select the Riley outcome.

when following the high-type worker and deviating to $\overline{m}$. Hence, the medium-type worker has no incentive to keep sending $m' = 0$, since doing so would only reveal that they are not high-type. In other words, although the medium-type worker initially prefers σ^{Pooling} to $\overline{\sigma}$, they ultimately find it optimal to deviate to $\overline{m}$ once the high-type worker has done so. We summarize this reasoning by saying that the interpretation of $\overline{m}$ in $\overline{\sigma}$ rather than in σ^{Pooling} triggers an *unraveling*.

Conversely, we consider what would happen to the high-type worker when they deviate to $\overline{m}$, while the medium-type worker continues to send $m' = 0$, as in σ^{Pooling}. Since the high-type worker obtains the highest equilibrium payoff by sending $\overline{m}$ in $\overline{\sigma}$, the firm should believe that the worker could be high-type after observing $\overline{m}$. When the firm interprets $\overline{m}$ in $\overline{\sigma}$, the worker could also be medium-type. However, the high-type worker has an incentive to deviate to $\overline{m}$ regardless of whether the firm believes the medium-type worker would also deviate. In particular, if the firm believes that only the high-type worker would deviate to $\overline{m}$, the firm would offer a wage higher than that in $\overline{\sigma}$, making the high-type worker even better off. Hence, the high-type worker's deviation to $\overline{m}$ is unaffected by the medium-type worker's choice. There will be no unraveling for the high-type worker.

When the firm expects that the worker would like to play the equilibrium that give them a higher payoff, the high-type worker would like to deviate to $\overline{m}$ and play $\overline{\sigma}$ because there is no unraveling, while the medium-type worker would ultimately deviate to $\overline{m}$ and deviate to $\overline{\sigma}$ because of unraveling. Taken together, unraveling provides a consistent story for the firm, explaining why both the medium-type and high-type workers would like to deviate to $\overline{m}$ and play $\overline{\sigma}$ instead of σ^{Pooling} irrespective of their initial preferences for $\overline{\sigma}$ or σ^{Pooling}. In this sense, we say that $\overline{\sigma}$ is *more persuasive* than σ^{Pooling}.

Now we formalize the concept of persuasiveness in any signaling game G. In response to the Stiglitz critique, persuasiveness builds on how the receiver interprets messages in different equilibria. Consider two equilibria $\overline{\sigma} = (\overline{\sigma}_S, \overline{\sigma}_R, \overline{\mu}), \sigma = (\sigma_S, \sigma_R, \mu) \in \text{SE}(G)$. We can view $\overline{\sigma}$ and the set $\{\sigma^1, \sigma^{\text{Riley}}, \sigma^{\text{Pooling}}\}$ in Example 2 as specific instances of $\overline{\sigma}$ and $\sigma \in \{\sigma^1, \sigma^{\text{Riley}}, \sigma^{\text{Pooling}}\}$. In order for $\overline{\sigma}$ to serve as a challenger to the putative equilibrium σ, we start with a message $\overline{m}$ that is on-path of the challenger $\overline{\sigma}$ such that there exists a type $\bar{t}$ who obtains a strictly higher payoff in $\overline{\sigma}$ than in σ. If no such message exists, every type of the sender will prefer σ to $\overline{\sigma}$. Then, the receiver should not expect that $\overline{\sigma}$ is played instead of σ. In Example 2, we have $\overline{m} = 1.1$, and $\bar{t} = t_H$.

Following the analysis in Example 2, we first divide the set of types of the sender who sends $\overline{m}$ in $\overline{\sigma}$ into two groups based on their preferences for $\overline{\sigma}$ or σ.[12] We allow mixed

[12]When some type of the sender is indifferent, we put them in the group that prefers $\overline{\sigma}$, which makes persuasiveness more selective (in a reasonable way). For instance, in Example 1, when $p = 0.5$, the high-type worker is indifferent between separation and pooling. Still, the pooling equilibrium is

strategies and define the two groups as follows:

$$T_{\overline{m}}^{\overline{\sigma}\geq\sigma} = \{t \in T | \, \overline{m} \in \text{supp}\left(\overline{\sigma}_S\left(t\right)\right), u_S\left(t, \overline{\sigma}\right) \geq u_S\left(t, \sigma\right)\}$$
$$T_{\overline{m}}^{\overline{\sigma}<\sigma} = \{t \in T | \, \overline{m} \in \text{supp}\left(\overline{\sigma}_S\left(t\right)\right), u_S\left(t, \overline{\sigma}\right) < u_S\left(t, \sigma\right)\}.$$

In Example 2, when comparing $\overline{\sigma}$ with either σ^1 or σ^{Riley}, we have $T_{\overline{m}}^{\overline{\sigma}\geq\sigma^1} = T_{\overline{m}}^{\overline{\sigma}\geq\sigma^1} = \{t_H, t_M\}$, and $T_{\overline{m}}^{\overline{\sigma}<\sigma^1} = T_{\overline{m}}^{\overline{\sigma}<\sigma^{\text{Riley}}} = \emptyset$. When comparing $\overline{\sigma}$ to σ^{Pooling}, we have $T_{\overline{m}}^{\overline{\sigma}\geq\sigma^{\text{Pooling}}} = \{t_H\}$ and $T_{\overline{m}}^{\overline{\sigma}<\sigma^{\text{Pooling}}} = \{t_M\}$. Building on the preceding discussion, we characterize the unraveling dynamics in Example 2 as follows.

Definition 2. The interpretation of $\overline{m}$ in $\overline{\sigma}$ rather than in σ triggers an *unraveling*, if (1) there exists $\bar{t} \in T_{\overline{m}}^{\overline{\sigma}\geq\sigma}$ such that $u_S\left(\bar{t}, \overline{\sigma}\right) > u_S\left(\bar{t}, \sigma\right)$; and (2) there exists a ranking of types $f : T_{\overline{m}}^{\overline{\sigma}<\sigma} \to \mathbb{R}$ such that for all $t' \in T_{\overline{m}}^{\overline{\sigma}<\sigma}$, and all $m' \in \text{supp}\left(\sigma_S\left(t'\right)\right)$,

$$u_S\left(t', \overline{\sigma}\right) \geq \max_{a' \in \text{BR}(m', \mu')} u_S\left(t', m', a'\right), \tag{1}$$

where

$$\mu'\left(t | \, m'\right) = \begin{cases} \frac{\mu(t|m')}{\sum_{\hat{t} \in U_{m'}^{\sigma>\overline{\sigma}}} \mu\left(\hat{t}|m'\right)} & \text{if } t \in U_{m'}^{\sigma>\overline{\sigma}}, \\ 0 & \text{otherwise.} \end{cases}$$
$$U_{m'}^{\sigma>\overline{\sigma}} = T_{m'}^{\sigma} \setminus \left(F_{\overline{m}}^{\overline{\sigma}<\sigma}\left(t'\right) \cup T_{\overline{m}}^{\overline{\sigma}\geq\sigma}\right)$$
$$T_{m'}^{\sigma} = \{t \in T | \, m' \in \text{supp}\left(\sigma_S\left(t\right)\right)\}$$
$$F_{\overline{m}}^{\overline{\sigma}<\sigma}\left(t'\right) = \left\{t \in T_{\overline{m}}^{\overline{\sigma}<\sigma} \middle| \, f\left(t\right) > f\left(t'\right)\right\}.$$

Note that if $T_{\overline{m}}^{\overline{\sigma}<\sigma} = \emptyset$, the comparison between $\overline{\sigma}$ and σ leads to a trivial unraveling. This corresponds to the situation in Example 2 when comparing $\overline{\sigma}$ with either σ^1 or σ^{Riley}. In such cases, all types who send $\overline{m}$ in $\overline{\sigma}$ prefer $\overline{\sigma}$ to σ, and hence they all have an incentive to deviate to $\overline{m}$ and play $\overline{\sigma}$ instead of σ. The more substantive case arises when $T_{\overline{m}}^{\overline{\sigma}<\sigma} \neq \emptyset$, as in Example 2 when comparing $\overline{\sigma}$ to σ^{Pooling}. In this case, there exists a type t' who sends $\overline{m}$ in $\overline{\sigma}$ but strictly prefers σ to $\overline{\sigma}$, i.e., $t' \in T_{\overline{m}}^{\sigma<\overline{\sigma}}$.

Unraveling occurs when any such type t', despite strictly preferring σ to $\overline{\sigma}$, nonetheless has an incentive to deviate to $\overline{m}$ and play $\overline{\sigma}$ once certain "other" types of the sender have already deviated to $\overline{m}$. These "other" types consist of (i) types who prefer $\overline{\sigma}$ to σ, i.e., $T_{\overline{m}}^{\overline{\sigma}\geq\sigma}$, and (ii) types ranked higher than t', i.e., $F_{\overline{m}}^{\overline{\sigma}<\sigma}\left(t'\right)$. If the sender of type t' adheres to σ and sends m', their payoff is at most equal to the equilibrium payoff in $\overline{\sigma}$ if the receiver believes

more persuasive than the Riley equilibrium because the low-type worker strictly prefers pooling.

that only types other than those "other" types would still play σ, forming a conditional belief μ' derived from μ by excluding those "other" types after observing m'. As a result, they would like to deviate to $\overline{\sigma}$. Unraveling proceeds sequentially from the highest-ranked type to the lowest-ranked type. The logic is similar to the classical analysis of voluntary disclosure in Grossman and Hart (1980), Milgrom (1981) and Verrecchia (1983). Here, lower-ranked types voluntarily deviate to $\overline{\sigma}$ once all higher-ranked types have already done so.

For instance, in Example 2, the firm forms a belief such that $\mu^{\text{Pooling}'}(t_H|\,\tilde{m}'=0)=0$ when the medium-type worker adheres to the pooling equilibrium. Under this belief, the medium-type worker's payoff (1.36) is lower than the equilibrium payoff in $\overline{\sigma}$ (1.85). As a result, the medium-type worker would prefer to deviate to $\overline{\sigma}$, even though they initially prefer σ^{Pooling} to $\overline{\sigma}$. Hence, the interpretation of $\overline{m}$ in $\overline{\sigma}$ rather than in σ^{Pooling} triggers an unraveling.

Unraveling provides a consistent story for the receiver, explaining why all types who send $\overline{m}$ in $\overline{\sigma}$ would like to deviate to $\overline{m}$ and play $\overline{\sigma}$ instead of σ irrespective of their initial preferences for $\overline{\sigma}$ or σ. Hence, the receiver should interpret $\overline{m}$ in $\overline{\sigma}$ instead of in σ, and we therefore say that $\overline{\sigma}$ is more persuasive than σ.

Definition 3. $\overline{\sigma}\in \text{SE}(G)$ is *more persuasive* than $\sigma\in\text{SE}(G)$, if there exists a message $\overline{m}$ on-path of $\overline{\sigma}$ such that the interpretation of $\overline{m}$ in $\overline{\sigma}$ rather than in σ triggers an unraveling.

Like the intuitive criterion and the D1 criterion, persuasiveness cannot distinguish between equilibria which are payoff-equivalent for the sender.[13] Hence, we define the most persuasive equilibrium as follows.

Definition 4. Two equilibria $\overline{\sigma},\sigma\in\text{SE}(G)$ are *payoff-equivalent* for the sender if $u_S(t,\overline{\sigma})=u_S(t,\sigma)$ for all $t\in T$.

Definition 5. $\overline{\sigma}\in\text{SE}(G)$ is *most persuasive* if it is more persuasive than any other equilibrium $\sigma\in\text{SE}(G)$ that is not payoff-equivalent for the sender.

If $\overline{\sigma}$ is most persuasive, then for any other equilibrium σ, we can alway find a message $\overline{m}$ such that every type who sends $\overline{m}$ in $\overline{\sigma}$ would like to play $\overline{\sigma}$ instead of σ because of unraveling. In Example 2, we argue that the LMSE is more persuasive than both the Riley equilibrium and the pooling equilibrium. More generally, one can establish that the LMSE is the unique most persuasive equilibrium among all equilibria in monotone signaling games; we formalize this result in the next section. In some non-monotone signaling games, however, a most persuasive equilibrium may fail to exist or be unique (see Section 5.2).

[13] Equilibria which differ only in the receiver's off-path beliefs are payoff-equivalent for the sender. There might also exist rare cases where two equilibria differ in the sender's strategies but still generate the same payoff.

4 Monotone Signaling Games

We study a class of monotone signaling games that satisfy the following assumptions, which includes Spence (1973)'s job market signaling model as an application.

Assumption 1. *Continuity and Concavity:*

- $T = \{1, 2, \ldots, n\}$ *is finite.*
- M *and* A *are closed intervals of* $\mathbb{R}$*.*
- u_S *and* u_R *are continuous in* m *and* a*.*
- u_R *is strictly concave in* a*.*

Assumption 1 ensures that $\mathrm{BR}(m, \mu)$, i.e., the receiver's best response correspondence after observing the message m under the belief μ, is always well-defined. In particular, it is a single-valued continuous function in m.

Assumption 2. *Monotonicity:*

- *If* $a' > a$*, then* $u_S(t, m, a') > u_S(t, m, a)$ *for all* t *and* m*.*
- $\frac{\partial u_R}{\partial a}$ *is a strictly increasing function of* t*.*

Assumption 2 describes the sender's and the receiver's preferences. The sender prefers a higher action from the receiver. The receiver prefers to take a higher action when they believe the sender is of a higher type. In the job market signaling model, this means that the worker prefers a higher wage from the firm, and the firm prefers to offer a higher wage when they believe the worker is of a higher type (more productive).

Assumption 3. *Single-Crossing:*

If $m < m'$ *and* $t < t'$*, then*
$u_S(t, m, a) \leq u_S(t, m', a')$ *implies that* $u_S(t', m, a) < u_S(t', m', a')$*.*

Assumption 3 is the Spence-Mirrlees single-crossing condition, which guarantees that the indifference curves of different types of the sender through a fixed message-action pair intersect only once. It captures the idea that higher messages are less costly for higher types to send than for lower types. In the job market signaling model, this means that the cost of acquiring higher education levels is decreasing in the worker's type.

For the next assumption, we introduce additional notation. For any non-empty subset K of T, the K-conditional belief $p_K \in \Delta(T)$ is defined as:

$$p_K(t) = \begin{cases} \frac{p(t)}{\sum_{t' \in K} p(t')} & \text{if } t \in K, \\ 0 & \text{otherwise.} \end{cases}$$

To simplify the notation, we write the receiver's best response to the message m under the K-conditional belief p_K as $\mathrm{BR}(m, K)$, i.e., $\mathrm{BR}(m, K) = \mathrm{BR}(m, p_K)$.

Assumption 4. *Low-Cost and High-Cost Messages: Let* $m^l = \min\{m \in M\}$ *and* $m^h = \max\{m \in M\}$.

- $\forall t \in T$, $\forall m \in M$, $u_S\left(t, m^l, BR\left(m^l, \{1\}\right)\right) \geq u_S(t, m, BR(m, \{1\}))$.
- $\forall t \in T \setminus \{n\}$, $u_S\left(t, m^h, BR\left(m^h, \{n\}\right)\right) < u_S\left(t, m^l, BR\left(m^l, \{1\}\right)\right)$.

Assumption 4 describes the message space. It states that the lowest message m^l is the cheapest message for all types of the sender to send, and the highest message m^h is the most expensive message such that no type of the sender, except possibly the highest type, would want to send it. In the job market signaling model, the worker incurs no cost for having zero education, while the level of education can go to infinity, which is too costly for every one.

Definition 6. A *monotone signaling game* G_{S} is a signaling game that satisfies Assumptions 1-4.

Similar assumptions (including A5 below) have been applied in many general treatments of this class of games (Riley, 1979; Cho and Sobel, 1990; Mailath, Okuno-Fujiwara and Postlewaite, 1993). Given that the message and action spaces are intervals, pure-strategy equilibria always exist. We denote the set of pure-strategy equilibria in the monotone signaling game G_{S} as $\mathrm{PSE}(G_{\mathrm{S}})$. Our analysis focuses on pure-strategy equilibria in G_{S}, and we begin by formally defining the LMSE.

Definition 7. $\overline{\sigma} \in \mathrm{PSE}(G_{\mathrm{S}})$ *lexicographically dominates* (lex-dominates) $\sigma \in \mathrm{PSE}(G_{\mathrm{S}})$, if there exists $\bar{t} \in T$ such that:

- $u_S(\bar{t}, \overline{\sigma}) > u_S(\bar{t}, \sigma)$.
- $u_S(t, \overline{\sigma}) \geq u_S(t, \sigma) \ \ \forall t > \bar{t}$.

$\overline{\sigma} \in \mathrm{PSE}(G_{\mathrm{S}})$ is a *lexicographically maximum sequential equilibrium* (LMSE) if there exists no other equilibrium $\sigma \in \mathrm{PSE}(G_{\mathrm{S}})$ that lex-dominates $\overline{\sigma}$. The equilibrium outcome of a LMSE is the *lexicographically maximum* outcome (lex max outcome). A LMSE exists

because pure-strategy equilibria exist in G_S (Mailath, Okuno-Fujiwara and Postlewaite, 1993).

We now present our main results, which formalize the intuitions illustrated in the preceding examples and establish, step by step, that the lex max outcome is generically the unique most persuasive equilibrium outcome.

Theorem 1. *In any game G_S, the LMSE is most persuasive.*

Proof Sketch: We need to show that the LMSE is more persuasive than any other equilibrium that is not payoff-equivalent for the sender. Given any other equilibrium σ, we first identify the message $\overline{m}$ on the equilibrium path of the LMSE that can be used to show that the interpretation of this message in the LMSE rather than in σ can trigger an unraveling. We pin down this message $\overline{m}$ as the equilibrium message sent by the highest type who strictly prefers the LMSE to σ. Next, we show that there exists a cutoff type among the set of types who send $\overline{m}$ in the LMSE such that all types weakly above this cutoff type prefer the LMSE to σ, while all types strictly below this cutoff type strictly prefer σ to the LMSE. To show the existence of an unraveling, we show that when the unraveling condition (1) is violated at some type, we can construct another equilibrium that lex-dominates the LMSE, which is a contradiction. See Appendix A.1 for the details of the proof.

Theorem 1 shows that for any alternative equilibrium that is not payoff-equivalent to the LMSE, the LMSE can always challenge this equilibrium with a new interpretation of some message. Under this new interpretation, all types who send this message in the LMSE would like to deviate to this message and play the LMSE instead of this alternative equilibrium because of unraveling. A natural question is whether an alternative equilibrium could similarly challenge the LMSE. The following result demonstrates that no such equilibrium exists.

Theorem 2. *In any game G_S, the most persuasive equilibrium is unique up to payoff equivalence for the sender.*

Proof Sketch: We prove by contradiction. Suppose there exist two most persuasive equilibria, the LMSE and $\hat{\sigma}$ that are not payoff-equivalent for the sender. Then, we can find a message $\hat{m}$ on the equilibrium path of $\hat{\sigma}$ such that the interpretation of this message in $\hat{\sigma}$ rather than in the LMSE triggers an unraveling. In particular, there exists a type i who sends $\hat{m}$ in $\hat{\sigma}$ and strictly prefers $\hat{\sigma}$ to the LMSE. However, there also exist types above i who send $\hat{m}$ in $\hat{\sigma}$ and strictly prefer the LMSE to $\hat{\sigma}$, which means we do not have the cutoff structure we see in the proof of Theorem 1. Then, we show that the unraveling condition (1) always fails when it comes to the highest-ranked type among the types identified above

irrespective of the ranking function, which is a contradiction. See Appendix A.2 for the details of the proof.

Theorems 2 establishes that no alternative equilibrium can challenge the LMSE by providing a new interpretation of any message. Moreover, Theorems 1 and 2 together imply that the interpretations of all messages in the game are determined by the LMSE, in the sense that no other equilibrium offers a more persuasive interpretation of any message.

We characterize the sender's equilibrium payoff in the most persuasive equilibrium. However, it does not guarantee a unique outcome. In order to ensure uniqueness, we assume the following as in Cho and Sobel (1990) and Mailath, Okuno-Fujiwara and Postlewaite (1993).

Assumption 5. *$u_S(t, m, BR(m, p))$ is strictly quasi-concave in m for all $t \in T$ and $p \in \Delta(T)$.*

Theorem 3. *Assume A5. Generically in the space of prior $p \in \Delta(T)$, the lex max outcome is the unique most persuasive equilibrium outcome in the game G_S.*

The genericity of the result implies that the measure of the priors under which uniqueness fails is zero. See Appendix A.3 for the details of the proof.

Under assumptions similar to A1-A5, Cho and Sobel (1990) uniquely select the Riley outcome. When taking into account the Stiglitz critique, persuasiveness uniquely selects the lex max outcome. Cho and Sobel (1990) introduce the following example to show that their result does not hold when the message space is discrete in monotone signaling games.

Example 3 (Discrete Spencian Game). We follow the setup of Example 1. There are two types of a worker, $T = \{t_L, t_H\}$, where $t_L = \frac{2}{3}$ and $t_H = 1$. The prior probabilities are $p(t_L) = p(t_H) = 0.5$. The set of education levels is discrete, $M = \{m_0, m_1\}$, where $m_0 = 0$ and $m_1 = 1$. The firm is the same as in Example 1. The payoff functions of workers are given by $u_S(t_L, m, a) = a - \frac{m}{2}$ and $u_S(t_H, m, a) = a - \frac{m}{4}$.

There are essentially three equilibria in this game, and we summarize in Table 3 below the equilibrium strategies and payoffs of the low-type and high-type workers in these equilibria, presented in the same format as Example 1.[14] Here, we allow mixed strategies. $\sigma_S^1(t_H) = \frac{1}{3}m_0 + \frac{2}{3}m_1$ implies that the high-type worker chooses m_0 with probability $\frac{1}{3}$ and m_1 with probability $\frac{2}{3}$.

All equilibria pass the D1 criterion (and the intuitive criterion), while the most persuasive equilibrium is the LMSE σ^{Pooling}, which Pareto-dominates the other two equilibria. In this

[14]The term "essentially" indicates that any other equilibria can differ only with respect to off-path beliefs.

	t_L	t_H
$\sigma_S^{\text{Riley}}(t)$	m_0	m_1
$\sigma_S^1(t)$	m_0	$\frac{1}{3}m_0 + \frac{2}{3}m_1$
$\sigma_S^{\text{Pooling}}(t)$	m_0	m_0

	t_L	t_H
$u_S\left(t, \sigma^{\text{Riley}}\right)$	$\frac{2}{3}$	$\frac{3}{4}$
$u_S\left(t, \sigma^1\right)$	$\frac{3}{4}$	$\frac{3}{4}$
$u_S\left(t, \sigma^{\text{Pooling}}\right)$	$\frac{5}{6}$	$\frac{5}{6}$

Table 3: Equilibrium Strategies and Payoffs of the Worker in Example 3

example, the high-type worker finds it too costly to separate from the low-type worker. However, it is not possible for the higher-type worker to signal their preference through an off-path message when applying the D1 criterion (or the intuitive criterion), because m_0 is on-path of every equilibrium. In contrast, persuasiveness builds on how the receiver interprets messages in different equilibria, regardless of whether these messages (m_0) are on-path or off-path of either σ^{Riley} or σ^1. In this example, the firm expects that the high-type worker would like to deviate to m_0 and play σ^{Pooling} instead of any other equilibria because both types of the worker strictly benefit, which is captured by the fact that the LMSE σ^{Pooling} is most persuasive. Cho and Sobel (1990) also introduce another example in their paper to show that their result does not hold when the action space is discrete, while we can still find a unique most persuasive equilibrium, i.e., the LMSE. Both examples suggest that persuasiveness has more selection power than the D1 criterion (and the intuitive criterion) when the message space or action space is discrete in monotone signaling games.[15]

5 Discussion

So far, we have focused on monotone signaling games.[16] We now show the applicability of persuasiveness to more general signaling games by looking at examples that are discussed in the previous literature. In Section 5.1, we show that persuasiveness can have more

[15]Extending Theorems 1 and 2 to monotone signaling games with compact message and action spaces is non-trivial. We need to allow mixed-strategy equilibria to ensure the existence of equilibrium. Persuasiveness relies on the comparison of the payoffs of different types of the sender across different equilibria. The main challenge lies in bounding the payoff that a particular type of sender could obtain under certain adjustments to the belief, as specified in Definition 2.

[16]To the author's best knowledge, monotone signaling games are the only class of signaling games where the previous selection criteria have strong selection power. For general signaling games, although there exist equilibria that pass the intuitive criterion and the D1 criterion, the set of such equilibria is difficult to characterize. The set of equilibrium outcomes that pass these criteria is typically non-singleton and could be potentially large (see Example 5 below). The applicability of these criteria is less clear in general signaling games. Hence, this paper focuses on monotone signaling games, and illustrates the applicability of persuasiveness to general signaling games by examples.

selection power than other criteria in some non-monotone signaling games. In Section 5.2, we turn to other examples in which no existing criteria have strong selection power. These examples illustrate that cyclicality arises when there does not exist a unique most persuasive equilibrium outcome. We argue that persuasiveness offers insights into equilibrium selection. In such cases, it may be reasonable to consider the *least* persuasive equilibrium. In Section 5.3, we show how persuasiveness can be applied to cheap-talk games, where other criteria designed for signaling games have no selection power.

5.1 Non-Monotone Signaling Games

We first look at the famous Beer-Quiche example introduced by Cho and Kreps (1987) to motivate the intuitive criterion.

Example 4 (Beer-Quiche Game). There are two types of the sender, $T = \{t_w, t_s\}$, where t_w is a wimp type and t_s is a surly type. The prior probabilities are $p(t_w) = 0.1$ and $p(t_s) = 0.9$. The sender can choose to have either beer or quiche for breakfast, i.e., $M = \{\text{beer}, \text{quiche}\}$. The receiver can choose to either challenge the sender to a duel or not, i.e., $A = \{\text{duel}, \text{don't}\}$. The payoffs are given in Figure 1 below, where the first entry in each pair is the sender's payoff and the second entry is the receiver's payoff. For instance, if the sender is of type t_w, has beer for breakfast, and the receiver chooses to duel, then the sender's payoff is 0 and the receiver's payoff is 1.

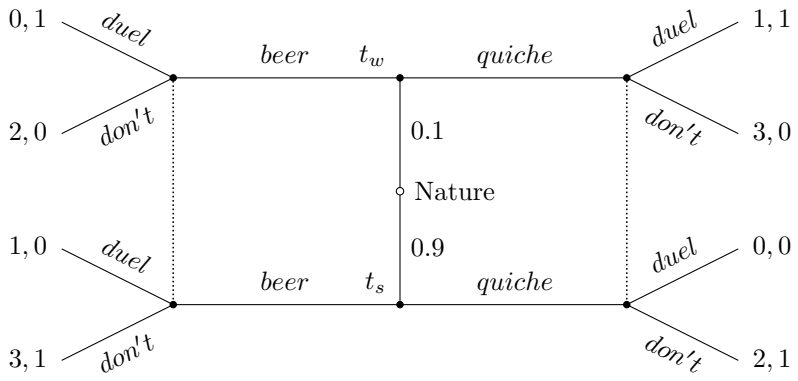

Figure 1: Beer-Quiche Game

There are essentially two pooling equilibria σ^{Beer} and σ^{Quiche} in this game. In σ^{Beer}, both types of the sender choose beer, and the receiver chooses to duel after observing quiche and not to duel after observing beer. In σ^{Quiche}, both types of the sender choose quiche, and the receiver chooses not to duel after observing quiche and to duel after observing beer. Both the intuitive criterion and the D1 criterion selects σ^{Beer}. It is easy to check that σ^{Beer}

is more persuasive than σ^{Quiche}, because the type t_s prefers σ^{Beer}, and the type t_w would also like to choose beer because of unraveling.

In Example 4, persuasiveness selects the same equilibrium outcome as both the intuitive criterion and the D1 criterion. However, this is not always the case in any non-monotone signaling game. In Example 5 below, which is introduced by Cho and Kreps (1987) to discuss their limitations, we show that persuasiveness can have more selection power than both the intuitive criterion and the D1 criterion.

Example 5 (Hiding Game). There are two types of the sender, $T = \{t_1, t_2\}$. The prior probabilities are $p(t_1) = p(t_2) = 0.5$. The sender selects either message m_1 or message m_2, i.e., $M = \{m_1, m_2\}$. The game terminates immediately following the choice of m_1, without any subsequent action from the receiver. Only if m_2 is chosen does the receiver have the opportunity to select an action from the set $A = \{a_1, a_2, a_3\}$. The payoffs are given in Figure 2 below, where the first entry in each pair is the sender's payoff and the second entry is the receiver's payoff. For instance, if the sender is of type t_1, chooses m_2, and the receiver chooses a_1, then the sender's payoff is -1 and the receiver's payoff is 3.

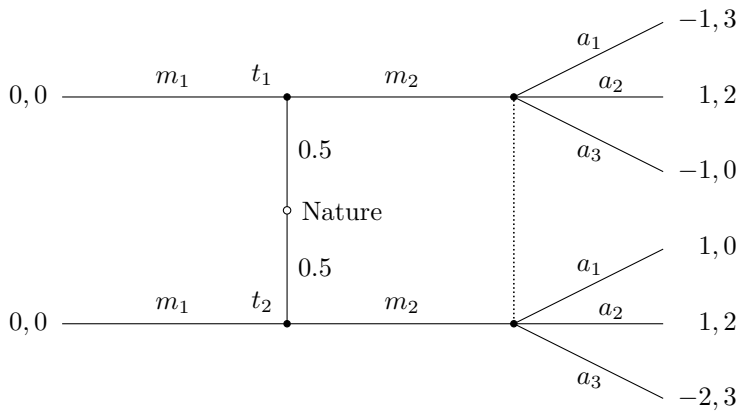

Figure 2: Hiding Game

In this game, both types of the sender try to hide their types from the receiver, as we can see from the payoff structure in Figure 2. There are essentially two pooling equilibria σ^{m_1} and σ^{m_2} in this game. In σ^{m_1}, both types of the sender choose m_1, and the receiver chooses a_3 after observing m_2. In σ^{m_2}, both types of the sender choose m_2, and the receiver chooses a_2 after observing m_2. Notice that σ^{m_2} strictly Pareto-dominates σ^{m_1}, as it yields higher payoffs for both the sender and the receiver. However, both equilibria pass the intuitive criterion and the D1 criterion. This is because both criteria emphasize how *one* type of the sender would like to *signal* to the receiver through an off-path message, but they do not

account for situations in which *two* types of the sender would like to *hide* from the receiver only by *jointly* sending an off-path message.[17] In contrast, persuasiveness selects σ^{m_2} as the most persuasive equilibrium, because both types of the sender would like to jointly deviate to m_2 and play σ^{m_2} instead of σ^{m_1}, which is captured by a trivial unraveling, i.e., the unraveling condition (1) is vacuously satisfied in this example. Note that persuasiveness not only considers the incentive of the sender to reveal their type as in monotone signaling games, but also takes into account the incentive of the sender to hide their types by pooling together as in this example.

In Table 4 below, we summarize the selection results of some equilibrium refinements in the examples discussed so far. In Appendix B, we offer an intuitive explanation for those criteria and show how they apply to the examples. Please refer to the original papers for the details of each selection criterion.

Table 4: Comparison of Equilibrium Refinements in Examples

	Intuitive & D1	G-P	Undefeated	Persuasive
Ex. 1 (Two-Type)	σ^{Riley}	None	LMSE	LMSE
Ex. 2 (Three-Type)	σ^{Riley}	None	LMSE, Pooling	LMSE
Ex. 3 (Discrete)	All	All	All	LMSE
Ex. 4 (Beer-Quiche)	σ^{Beer}	σ^{Beer}	$\sigma^{\text{Beer}}, \sigma^{\text{Quiche}}$	σ^{Beer}
Ex. 5 (Hiding)	$\sigma^{m_1}, \sigma^{m_2}$	σ^{m_2}	σ^{m_2}	σ^{m_2}

1. The intuitive criterion cannot select σ^{Riley} in Ex. 2, because there are more than two types of the sender. Otherwise, the intuitive criterion and the D1 criterion select the same equilibria in all other examples.
2. G-P: Perfect Sequential Equilibrium (Grossman and Perry, 1986).
3. Undefeated Equilibrium (Mailath, Okuno-Fujiwara and Postlewaite, 1993).

The table shows that persuasiveness uniquely selects the most persuasive outcome across all examples, while other criteria fail to yield a unique prediction in certain cases. It suggests that persuasiveness has strong selection power even beyond monotone signaling games. Note that all examples discussed so far have a unique most persuasive equilibrium outcome. However, this is not always the case in general signaling games, as we discuss next.

[17] If we test σ^{m_1} by applying the two steps of the intuitive criterion as described at the start of Section 3, then we get $D = \{t_1, t_2\}$ in Step 1 because both types of the sender could benefit by deviating to m_2. However, in Step 2, no type of the sender can profitably deviate to m_2 under the least favorable belief of the receiver.

5.2 Limitations

In this section, we examine two examples in which no existing criteria have strong selection power. They also do not admit a unique most persuasive equilibrium outcome. The first example, attributed to Kreps, is introduced by Grossman and Perry (1986) to highlight the coordination problem between the sender and the receiver. The second example is proposed by Mailath, Okuno-Fujiwara and Postlewaite (1993) to illustrate the difficulty of imposing plausible restrictions on off-path beliefs. When applying persuasiveness to both examples, the issues translate into the non-uniqueness and non-existence of the most persuasive equilibrium, respectively. However, we argue that persuasiveness is a useful concept in these examples by providing insights into equilibrium selection.

Example 6 (Coordination Game). There are two types of the sender, $T = \{t_1, t_2\}$. The prior probabilities are $p(t_1) = p(t_2) = 0.5$. The sender selects either message m_1 or message m_2, i.e., $M = \{m_1, m_2\}$. The receiver can choose to take either action a_1 or action a_2, i.e., $A = \{a_1, a_2\}$. The payoffs are given in Figure 3 below.

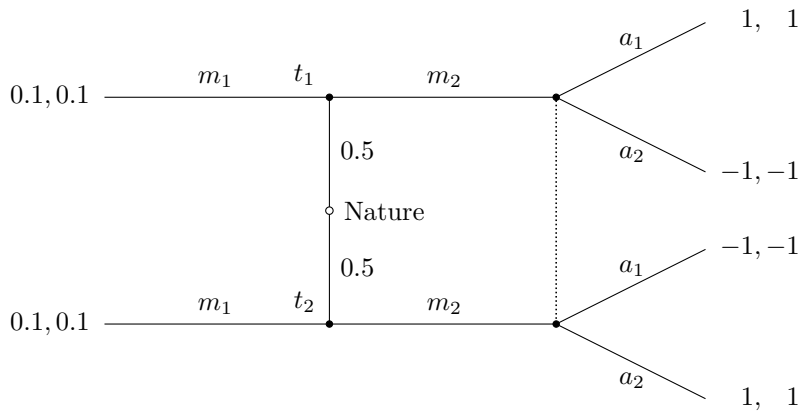

Figure 3: Coordination Game

There are essentially three equilibria in this game, and we summarize in Table 5 below the equilibrium strategies and payoffs of both types of the sender in these equilibria, presented in the same format as Example 1.

The first two equilibria, σ^1 and σ^2, are separating equilibria in which the receiver takes different actions that best respond to different types of the sender in each equilibrium. The third equilibrium, σ^{Pooling}, is a pooling equilibrium in which both types of the sender send m_1, and the receiver randomizes between the two actions with equal probabilities after observing m_2. All equilibria pass the intuitive criterion and the D1 criterion.[18]

[18] σ^1 and σ^2 are both undefeated equilibria and perfect sequential equilibria.

	t_1	t_2
$\sigma_S^1(t)$	m_2	m_1
$\sigma_S^2(t)$	m_1	m_2
$\sigma_S^{\text{Pooling}}(t)$	m_1	m_1

	t_1	t_2
$u_S(t, \sigma^1)$	1	0.1
$u_S(t, \sigma^2)$	0.1	1
$u_S\left(t, \sigma^{\text{Pooling}}\right)$	0.1	0.1

Table 5: Equilibrium Strategies and Payoffs of the Sender in Example 6

The coordination problem in this example arises because, a priori, there is no compelling reason for the sender and receiver to coordinate on either σ^1 or σ^2. Indeed, if coordination were achieved, both types of the sender would strictly prefer to send m_2. Notice that if the sender could make a "speech" to the receiver, the "speech" would be used as a coordination device to select either σ^1 or σ^2, which creates a cheap-talk equilibrium in which both types of the sender obtain a payoff of 1 (Reny, 2025). However, in the absence of such a "speech," the sender and receiver cannot coordinate on either equilibrium.[19] In this example, it becomes more reasonable to expect that the pooling equilibrium σ^{Pooling} would be played, because the receiver would like to randomize in order to avoid miscoordination after observing m_2. Consequently, σ^{Pooling} can be interpreted as a "safe" equilibrium outcome.

When applying persuasiveness to this example, the coordination problem is captured by the fact that σ^1 is more persuasive than σ^2, while at the same time σ^2 is more persuasive than σ^1. Although both equilibria more persuasive than the pooling equilibriumσ^{Pooling}, they exhibit a cyclical relationship with one another. Hence, there does not exist a unique most persuasive equilibrium outcome. Non-uniqueness highlights the existence of multiple interpretations of m_2 in either σ^1 or σ^2. In this example, the coordination problem translates into multiple interpretations of the same message due to the non-uniqueness of the most persuasive equilibrium. As we discussed above, the least persuasive equilibrium σ^{Pooling} is more appealing, which faces no issue of multiple interpretations of m_2.

Definition 8. $\underline{\sigma} \in \text{SE}(G)$ is *least persuasive* if any other equilibrium $\sigma \in \text{SE}(G)$, that is not payoff-equivalent for the sender, is more persuasive than $\underline{\sigma}$.

The above example illustrates the coordination problem that arises when the interests of the sender and receiver are perfectly aligned. We now turn to a contrasting example in which their interests are misaligned. In this case, no most persuasive equilibrium exists. The interpretations of certain messages in some equilibria are ambiguous due to the presence of

[19]It also shows that we cannot always rely on the implicit "speech" to interpret equilibrium selection criteria such as the intuitive criterion and the D1 criterion.

alternative, more persuasive equilibria. Consequently, upon observing such a message, the receiver may retain a reasonable doubt regarding the sender's type.

Example 7 (Reasonable Doubt Game). There are three types of the sender, $T = \{t_1, t_2, t_3\}$. The prior probabilities are $p(t_1) = p(t_2) = p(t_3) = \frac{1}{3}$. The sender selects one message from $M = \{m_1, m_2, m_3, m_4\}$. The game ends immediately after m_4, otherwise the receiver can choose one action from $A = \{a_1, a_2, a_3\}$. The payoffs are given in Figure 4 below. There are essentially four equilibria in this game, and we summarize in Table 6 below the equilibrium strategies and payoffs of the sender in all equilibria.

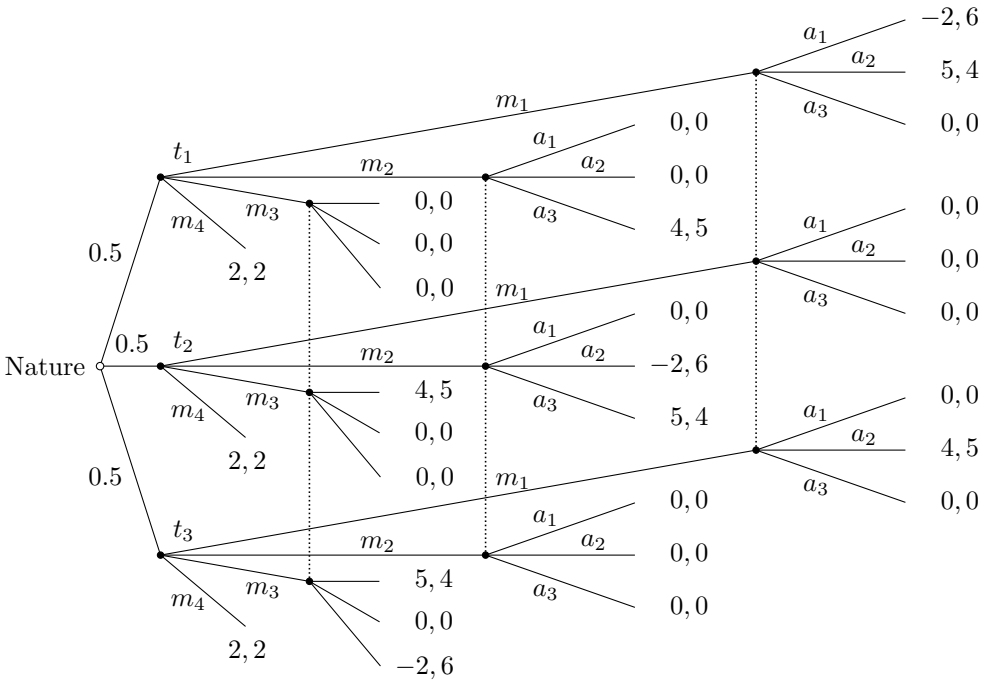

Figure 4: Reasonable Doubt Game

	t_1	t_2	t_3
$\sigma_S^1(t)$	m_1	m_4	m_1
$\sigma_S^2(t)$	m_2	m_2	m_4
$\sigma_S^3(t)$	m_4	m_3	m_3
$\sigma_S^{\text{Pooling}}(t)$	m_4	m_4	m_4

	t_1	t_2	t_3
$u_S(t, \sigma^1)$	5	2	4
$u_S(t, \sigma^2)$	4	5	2
$u_S(t, \sigma^3)$	2	4	5
$u_S\left(t, \sigma^{\text{Pooling}}\right)$	2	2	2

Table 6: Equilibrium Strategies and Payoffs of the Sender in Example 7

All equilibria pass the intuitive criterion and the D1 criterion.[20] Notice that the sender of type t_i attains the highest equilibrium payoff in σ^i for each $i = 1, 2, 3$. However, for type t_i to obtain this payoff, type t_{i-1} must also be willing to send message m_i; yet, in σ^{i-1}, type t_{i-1} could achieve a higher equilibrium payoff by instead sending m_{i-1}. Hence, when the receiver observes m_i, they have a reasonable doubt that the sender might be of type t_i, because only type t_i could achieve the highest equilibrium payoff by sending m_i. This reasonable doubt could lead the receiver to take action a_i after observing m_i instead of action a_{i+1} in σ^i. Then, type t_i would obtain a payoff of -2 instead of 5 by sending m_i.[21] In this example, it becomes more reasonable to expect that the pooling equilibrium σ^{Pooling} would be played because of the existence of such reasonable doubts. Consequently, σ^{Pooling} can be interpreted as a "safe" equilibrium outcome.

When applying persuasiveness to this example, those reasonable doubts are captured by the fact that σ^{i-1} is more persuasive than σ^i, which leads to a cycle. They are all more persuasive than σ^{Pooling}. Hence, there does not exist a most persuasive equilibrium. Non-existence casts doubt on the interpretation of m_i in the equilibrium σ^i because there exists another equilibrium σ^{i-1} that is more persuasive than σ^i. As we discussed above, the least persuasive equilibrium σ^{Pooling} is more appealing, which is free from such reasonable doubts.

Neither the coordination game nor the reasonable doubt game admits a unique most persuasive equilibrium outcome. This non-uniqueness leads to multiple interpretations of the same message in different equilibria, while non-existence casts doubt on the interpretations of certain messages in some equilibria. In both examples, it appears more reasonable to expect that the least persuasive equilibrium would be played. This is not a coincidence. The term "least persuasive" can be somewhat misleading when a unique most persuasive equilibrium does not exist, as the argument that all other equilibria are more persuasive relies on the interpretations of messages that are subject to either multiple readings or reasonable doubts. Hence, the term "least persuasive" is a manifestation of the fact that this equilibrium is free from both issues and can be considered "safe."

In general, when a game does not admit a unique most persuasive equilibrium outcome, it is unclear which equilibrium will prevail. Nonetheless, the notion of persuasiveness remains useful, as it offers guidance for reasoning about equilibrium selection in such contexts. As demonstrated by the two examples, there are situations in which it is reasonable to focus on the least persuasive equilibrium.

[20] No equilibrium is undefeated equilibrium or perfect sequential equilibrium.

[21] When $i = 4$, we relabel $i = 1$. When $i = 0$, we relabel $i = 3$.

5.3 Cheap-Talk Games

Although we focus on (costly) signaling games in this paper, persuasiveness can also be applied to costless signaling games, i.e., cheap-talk games. Building on the setup of signaling games in Section 2, cheap-talk games can be readily modeled by assuming that the set of messages M is infinite and that the payoff functions of both the sender and the receiver are independent of the message sent. We denote cheap-talk games as G_{CT}. Note that the standard equilibrium refinements for signaling games, such as the intuitive criterion and the D1 criterion, have no selective power in cheap-talk games. This is because one can always support any equilibrium outcome with a "noisy" equilibrium in which all messages are sent with positive probability on the equilibrium path, so arguments that put plausible restrictions on off-path beliefs have no power to refine (Farrell, 1993, Section 3). In contrast, persuasiveness emphasizes how the receiver interprets messages in equilibrium, which can select equilibria even when every message is on-path (see Example 3 above).

We illustrate the selection power of persuasiveness in cheap-talk games with the following examples introduced in Farrell (1993).

Example 8 (Cheap-Talk Games). There are two types of the sender, $T = \{t_1, t_2\}$. The prior probabilities are $p(t_1) = p$ and $p(t_2) = 1 - p$. The receiver has three different actions: $a(t_1)$ and $a(t_2)$ are best responses for the receiver when they are sufficiently confident that the sender is of type t_1 or t_2 respectively, and $a(T)$ is the best response when the receiver has (close enough to) the prior probabilities in mind.[22] We consider the following three different cheap-talk games. The payoffs of the sender are given in Table 7 below.

	t_1	t_2
$a(t_1)$	3	0
$a(t_2)$	0	3
$a(T)$	2	2

G^1_{CT}: I Will Tell You

	t_1	t_2
$a(t_1)$	1	0
$a(t_2)$	0	1
$a(T)$	2	2

G^2_{CT}: I Won't Tell You

	t_1	t_2
$a(t_1)$	2	1
$a(t_2)$	-1	0
$a(T)$	0	2

G^3_{CT}: I Can't Tell You

Table 7: Payoffs of the Sender in Example 8

In both G^1_{CT} and G^2_{CT}, there are two equilibria: a babbling equilibrium where no information is transmitted and an informative equilibrium where the sender perfectly reveals their type. In G^3_{CT}, there is a unique babbling equilibrium.

[22] It is easy to construct payoff functions of the receiver that lead to Table 7. We omit these constructions, as they are not essential to our analysis.

We call G^1_{CT} the "I Will Tell You" game, because it is in both the sender's and receiver's interests to coordinate as in Example 6. Hence, the informative equilibrium is more appealing in G^1_{CT}. In contrast, we call G^2_{CT} the "I Won't Tell You" game, because it is against both types of the sender's interests to reveal their types. Hence, the babbling equilibrium is more appealing in G^2_{CT}. In both games, the more appealing equilibrium is the most persuasive equilibrium. In G^3_{CT}, the babbling equilibrium is the unique equilibrium, which is vacuously the most persuasive equilibrium. We call it the "I Can't Tell You" game, because even if the sender of type t_1 wants to tell the receiver that they are of type t_1, they are unable to do so credibly as the sender of type t_2 always imitates type t_1.

Persuasiveness uniquely selects the most persuasive equilibrium in the above three cheap-talk games. It is related to Farrell (1993)'s notion of neologism-proofness. However, this criterion is too demanding that the unique babbling equilibrium in G^3_{CT} is not neologism-proof. It captures the incentive of the sender of type t_1 to reveal their type, but ignores the incentive of the sender of type t_2 to mimic type t_1 in an equilibrium—it does not account for whether the sender of type t_1 can credibly reveal their type in an equilibrium.[23] In contrast, persuasiveness considers both incentives and admits the unique babbling equilibrium in G^3_{CT}.

Unfortunately, persuasiveness cannot uniquely select the most informative equilibrium in the Crawford-Sobel model (Crawford and Sobel, 1982). Chen, Kartik and Sobel (2008) propose an alternative criterion, NITS, which can uniquely select the most informative equilibrium under certain conditions. However, it is not clear how to extend their criterion to the current example or to more general cheap-talk games.[24]

6 Related Literature

The literature on equilibrium selection is mostly based on the logic of forward induction. There are two main approaches. One is the axiomatic approach that begins with strategic stability initiated by Kohlberg and Mertens (1986), which looks for equilibria that satisfy a list of axiomatic desiderata. Subsequent research seeks to refine and redefine the concept of stability (Mertens, 1989, 1991; van Damme, 1989; Hillas, 1990; Dilmé, 2024). Two recent papers that examine forward induction from a decision-theoretic and axiomatic approach are Govindan and Wilson (2009, 2012). The other approach is the belief-based refinement,

[23] It can also be seen as the Stiglitz critique in cheap-talk games (Rabin, 1990, p. 163).

[24] NITS is specifically designed for the Crawford-Sobel model. It requires a general notion of lowest type. It is not clear how to define the lowest type in general cheap-talk games such as the current example.

which is more directly motivated by putting plausible restrictions on off-path beliefs.[25] The two most well-known members of this family is the intuition criterion (Cho and Kreps, 1987) and the divinity criterion (Banks and Sobel, 1987). Cho (1987) extends the intuitive criterion to general extensive-form games. Grossman and Perry (1986) refine the intuitive criterion. However, the G-P criterion suffers the non-existence problem in the job market signaling model (Spence, 1973). Recent studies formalize the implicit "speech" used by Cho and Kreps (1987) to motivate the intuitive criterion by adding cheap talk to signaling games (Clark and Fudenberg, 2021; Reny, 2025). Adding cheap talk to signaling games can refine equilibria; however, it may also select an equilibrium that does not exist in signaling games without cheap talk, as illustrated in Example 6.[26]

There is also a strand of literature that studies equilibrium selection in a dynamic model from an evolutionary and learning perspective.[27] Building on the work of Kandori, Mailath and Rob (1993) and Young (1993), Nöldeke and Samuelson (1997) show that in a two-type Spencian game, the lex max outcome, which is uniquely selected by persuasiveness, is always contained in the unique recurrent set, while the Riley outcome may not be.[28]

The Stiglitz critique (Cho and Kreps, 1987, p. 203) calls into question the general sorts of arguments used in these criteria to refine equilibria (Mailath, 1988), when they uniquely select the Riley outcome in Example 1. In response to that, Mailath, Okuno-Fujiwara and Postlewaite (1993) propose the notion of undefeated equilibrium.[29] Persuasiveness is related to, yet distinct from, the undefeated equilibrium. Specifically, we do not require that every type of the sender who sends a message in a more persuasive equilibrium must prefer this equilibrium to the putative equilibrium (see Appendix B for details).[30] The unraveling condition (1) ensures that the set of types who initially strictly prefer the putative equilibrium to the more persuasive one would eventually deviate. This distinction is crucial for establishing the uniqueness of the most persuasive equilibrium in monotone signaling

[25]The two approaches are connected. Both the intuitive criterion and the divinity criterion are weaker than NWBR (Never a Weak Best Response), which is implied by strategic stability.

[26]For the equilibrium refinements in cheap-talk games, see Farrell (1993); Rabin (1990); Matthews, Okuno-Fujiwara and Postlewaite (1991); Matthews and Postlewaite (1994); Zapater (1997); Chen, Kartik and Sobel (2008); Sémirat and Forges (2023); Gordon, Kartik, Lo, Olszewski and Sobel (2024).

[27]See also Rabin and Sobel (1996), Umbhauer (1997) and Clark and Fudenberg (2021).

[28]They use the term "Hellwig equilibrium" to refer to the best pooling equilibrium for the high-type sender (Hellwig, 1987). When the Riley equilibrium lex-dominates the Hellwig equilibrium, the Riley outcome is the unique outcome in the recurrent set. When the Hellwig equilibrium lex-dominates the Riley equilibrium, the Hellwig outcome is always contained in the unique recurrent set, while the Riley outcome may not be. In both cases, the lex outcome is always selected.

[29]Umbhauer (1991) develops the consistent forward induction criterion which is similar to the undefeated equilibrium.

[30]We also do not require this message in the more persuasive equilibrium to be off-path of the putative equilibrium.

games (Theorem 2), given that the undefeated equilibrium is typically not unique (see Table 4 above).

The Stiglitz critique is similar to the motivation of Wilson equilibrium (Wilson, 1977) in insurance markets with adverse selection (Rothschild and Stiglitz, 1976). Hellwig (1987, Section 3) poses the question of why different equilibrium outcomes are selected in different contexts depending on which side—the informed or uninformed player—moves first, even though the underlying intuitive principles guiding these selections appear similar. Specifically, the Riley outcome is selected (by the intuitive criterion) when the customer proposes contracts, whereas the Wilson outcome arises when the insurance company does. This paper addresses the Stiglitz critique and provides an answer to this question. It shows that persuasiveness selects the same outcome in a signaling game where the customer moves first as the Wilson outcome in a screening game where the insurance company moves first, since both correspond to the lex max outcome.[31] Hence, the discrepancy in equilibrium selection in different contexts disappears, when we apply persuasiveness to signaling games.

The unraveling logic behind persuasiveness is related to the literature on information disclosure. See Grossman and Hart (1980), Grossman (1981), Milgrom (1981), Verrecchia (1983), and Madarász and Pycia (2025). The iterative elimination of types of the sender from playing the putative equilibrium connects to the literature on the iterative elimination of strictly dominated strategies (Abreu and Matsushima, 1992; Kapon, Del Carpio and Chassang, 2024).[32]

7 Concluding Remarks

This paper introduces a novel criterion, called persuasiveness, to select equilibria in signaling games. Persuasiveness is immune to the Stiglitz critique. It builds on how the receiver interprets messages in different equilibria, regardless of whether these messages are on-path or off-path. An equilibrium is more persuasive than an alternative one if there exists a message on its equilibrium path such that every type of the sender who sends that message in the equilibrium would like to deviate from the alternative equilibrium to that message because of unraveling. An equilibrium is most persuasive if it is more persuasive than any

[31] To ensure the consistency between the signaling and screening games, we rule out cross-subsidization (Miyazaki, 1977), because the customer cannot propose a menus of contracts in a signaling game. For a game theoretical foundation of the equilibrium selection in insurance markets with cross-subsidization, see Mimra and Wambach (2011) and Netzer and Scheuer (2014).

[32] It is also related to the literature on the use of divide-and-conquer mechanisms to implement desirable social outcomes under all rationalizable strategy profiles. See also Segal and Whinston (2000); Spiegler (2000); Segal (2003); Winter (2004); Bó (2007); Eliaz and Spiegler (2015); Halac, Kremer and Winter (2020); Halac, Lipnowski and Rappoport (2021).

other equilibrium that is not payoff-equivalent for the sender. When there exists a unique most persuasive equilibrium, the interpretations of all messages in the game are determined, in the sense that no other equilibrium can provide a more persuasive interpretation of any message.

Persuasiveness has strong selective power. In monotone signaling games, it uniquely selects the most persuasive equilibrium outcome—the lex max outcome. In some non-monotone signaling games, it has stronger selective power than other existing equilibrium refinements. Furthermore, persuasiveness can have good selective power in cheap-talk games, where standard equilibrium refinements for signaling games have no selective power.

We conclude by posing three questions for future research. First, when a unique most persuasive equilibrium does not exist, how should equilibrium selection be approached? As illustrated by the two examples in Section 5.2, it may be plausible to expect that the least persuasive equilibrium prevails. Whether this observation generalizes to a broader class of signaling games, however, remains unclear. Second, can the concept of persuasiveness be extended to more general multi-stage games with multiple players? Intuitively, such an extension would require applying persuasiveness at each equilibrium history rather than to the equilibrium as a whole. It remains unclear whether this extension can be formulated in a concise manner and how it would relate to existing refinements for general extensive-form games (Cho, 1987; Govindan and Wilson, 2009). Third, can we identify a criterion that uniquely selects an equilibrium outcome in both the Spencian signaling games and Crawford-Sobel cheap-talk games, the two canonical models of communication under asymmetric information? Existing selection criteria for signaling games typically do not apply to cheap-talk games, and vice versa. Persuasiveness suggests the potential for a unified criterion, as it has selective power in both frameworks; nevertheless, it fails to uniquely select the most informative equilibrium in Crawford and Sobel (1982). Whether a unified criterion can be established that uniquely selects an equilibrium outcome in both classes of games remains an open question.

References

Abreu, Dilip and Hitoshi Matsushima (1992) "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica*, 60 (5), 993–1008, 10.2307/2951536. 32

Banks, Jeffrey S. and Joel Sobel (1987) "Equilibrium Selection in Signaling Games," *Econometrica*, 55 (3), 647–661, 10.2307/1913604. 2, 8, 31, 52

Barro, Robert J. and David B. Gordon (1983) "Rules, Discretion and Reputation in a Model of Monetary Policy," *Journal of Monetary Economics*, 12 (1), 101–121, 10.1016/0304-3932(83)90051-X. 2

Bó, Ernesto Dal (2007) "Bribing Voters," *American Journal of Political Science*, 51 (4), 789–803, https://www.jstor.org/stable/4620100. 32

Chen, Ying, Navin Kartik, and Joel Sobel (2008) "Selecting Cheap-Talk Equilibria," *Econometrica*, 76 (1), 117–136, 10.1111/j.0012-9682.2008.00819.x. 30, 31

Cho, In-Koo (1987) "A Refinement of Sequential Equilibrium," *Econometrica*, 55 (6), 1367–1389, 10.2307/1913561. 31, 33

Cho, In-Koo and David M. Kreps (1987) "Signaling Games and Stable Equilibria," *The Quarterly Journal of Economics*, 102 (2), 179–221, 10.2307/1885060. 2, 8, 9, 22, 23, 31, 51

Cho, In-Koo and Joel Sobel (1990) "Strategic Stability and Uniqueness in Signaling Games," *Journal of Economic Theory*, 50 (2), 381–413, 10.1016/0022-0531(90)90009-9. 13, 18, 20, 21

Clark, Daniel and Drew Fudenberg (2021) "Justified Communication Equilibrium," *American Economic Review*, 111 (9), 3004–3034, 10.1257/aer.20201692. 31

Crawford, Vincent P. and Joel Sobel (1982) "Strategic Information Transmission," *Econometrica*, 50 (6), 1431–1451, 10.2307/1913390. 30, 33

Dilmé, Francesc (2024) "Sequentially Stable Outcomes," *Econometrica*, 92 (4), 1097–1134, 10.3982/ECTA21402. 30

Eliaz, Kfir and Ran Spiegler (2015) "X-Games," *Games and Economic Behavior*, 89, 93–100, 10.1016/j.geb.2014.12.005. 32

Farrell, Joseph (1993) "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior*, 5 (4), 514–531, 10.1006/game.1993.1029. 29, 30, 31

Fudenberg, Drew and Jean Tirole (1983) "Sequential Bargaining with Incomplete Information," *The Review of Economic Studies*, 50 (2), 221–247, 10.2307/2297414. 2

——— (1991) "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 53 (2), 236–260, 10.1016/0022-0531(91)90155-W. 7

Gordon, Sidartha, Navin Kartik, Melody Pei-Yu Lo, Wojciech Olszewski, and Joel Sobel (2024) "Effective Communication in Cheap-Talk Games," *Working Papers* (hal-04743271), https://ideas.repec.org//p/hal/wpaper/hal-04743271.html. 31

Govindan, Srihari and Robert Wilson (2009) "On Forward Induction," *Econometrica*, 77 (1), 1–28, 10.3982/ECTA6956. 30, 33

——— (2012) "Axiomatic Equilibrium Selection for Generic Two-Player Games," *Econometrica*, 80 (4), 1639–1699, 10.3982/ECTA9579. 30

Grossman, S. J. and O. D. Hart (1980) "Disclosure Laws and Takeover Bids," *The Journal of Finance*, 35 (2), 323–334, 10.1111/j.1540-6261.1980.tb02161.x. 5, 16, 32

Grossman, Sanford J. (1981) "The Informational Role of Warranties and Private Disclosure about Product Quality," *The Journal of Law & Economics*, 24 (3), 461–483, https://www.jstor.org/stable/725273. 5, 32

Grossman, Sanford J and Motty Perry (1986) "Perfect Sequential Equilibrium," *Journal of Economic Theory*, 39 (1), 97–119, 10.1016/0022-0531(86)90022-0. 8, 24, 25, 31, 52

Halac, Marina, Ilan Kremer, and Eyal Winter (2020) "Raising Capital from Heterogeneous Investors," *American Economic Review*, 110 (3), 889–921, 10.1257/aer.20190234. 32

Halac, Marina, Elliot Lipnowski, and Daniel Rappoport (2021) "Rank Uncertainty in Organizations," *American Economic Review*, 111 (3), 757–786, 10.1257/aer.20200555. 32

Hellwig, Martin (1987) "Some Recent Developments in the Theory of Competition in Markets with Adverse Selection ∗," *European Economic Review*, 31 (1), 319–325, 10.1016/0014-2921(87)90046-8. 31, 32

Hillas, John (1990) "On the Definition of the Strategic Stability of Equilibria," *Econometrica*, 58 (6), 1365–1390, 10.2307/2938320. 30

John, Kose and Joseph Williams (1985) "Dividends, Dilution, and Taxes: A Signalling Equilibrium," *The Journal of Finance*, 40 (4), 1053–1070, 10.2307/2328394. 2

Kandori, Michihiro, George J. Mailath, and Rafael Rob (1993) "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica*, 61 (1), 29–56, 10.2307/2951777. 31

Kapon, Samuel, Lucia Del Carpio, and Sylvain Chassang (2024) "Using Divide-and-Conquer to Improve Tax Collection," *The Quarterly Journal of Economics*, 139 (4), 2475–2523, 10.1093/qje/qjae018. 32

Kohlberg, Elon and Jean-Francois Mertens (1986) "On the Strategic Stability of Equilibria," *Econometrica*, 54 (5), 1003–1037, 10.2307/1912320. 2, 8, 30

Kreps, David M. (2023) *Microeconomic Foundations II: Imperfect Competition, Information, and Strategic Interaction*, Princeton (N.J.): Princeton University Press. 2

Kreps, David M. and Robert Wilson (1982) "Sequential Equilibria," *Econometrica*, 50 (4), 863, 10.2307/1912767. 2

Leland, Hayne E. and David H. Pyle (1977) "Informational Asymmetries, Financial Structure, and Financial Intermediation," *The Journal of Finance*, 32 (2), 371–387, 10.2307/2326770. 2

Madarász, Kristóf and Marek Pycia (2025) "Cost over Content: Information Choice in Trade," *Working Paper*, https://drive.google.com/file/d/1nJFYdaGG2uKDacno4tmPg6Idl_uLiNYl/view?usp=embed_facebook. 32

Mailath, George J (1988) "A Reformulation of a Criticism of The Intuitive Criterion and Forward Induction," *Mimeo*. 31

Mailath, George J., Masahiro Okuno-Fujiwara, and Andrew Postlewaite (1993) "Belief-Based Refinements in Signalling Games," *Journal of Economic Theory*, 60 (2), 241–276, 10.1006/jeth.1993.1043. 3, 10, 12, 18, 19, 20, 24, 25, 31, 40, 53

Matthews, Steven A, Masahiro Okuno-Fujiwara, and Andrew Postlewaite (1991) "Refining Cheap-Talk Equilibria," *Journal of Economic Theory*, 55 (2), 247–273, 10.1016/0022-0531(91)90040-B. 31

Matthews, Steven A. and Andrew Postlewaite (1994) "On Modeling Cheap Talk in Bayesian Games," in Ledyard, John O. ed. *The Economics of Informational Decentralization: Complexity, Efficiency, and Stability: Essays in Honor of Stanley Reiter*, 347–366, Boston, MA: Springer US, 10.1007/978-1-4615-2261-4_13. 31

Mertens, Jean-François (1989) "Stable Equilibria: A Reformulation Part I. Definition and Basic Properties," *Mathematics of Operations Research*, 14 (4), 575–625, https://www.jstor.org/stable/3689732. 30

——— (1991) "Stable Equilibria: A Reformulation. Part II. Discussion of the Definition, and Further Results," *Mathematics of Operations Research*, 16 (4), 694–753, https://www.jstor.org/stable/3689907. 30

Milgrom, Paul R. (1981) "Good News and Bad News: Representation Theorems and Applications," *The Bell Journal of Economics*, 12 (2), 380–391, 10.2307/3003562. 5, 16, 32

Milgrom, Paul and John Roberts (1982) "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis," *Econometrica*, 50 (2), 443–459, 10.2307/1912637. 2

Mimra, Wanda and Achim Wambach (2011) "A Game-Theoretic Foundation for the Wilson Equilibrium in Competitive Insurance Markets with Adverse Selection," April, 10.2139/ssrn.1808672. 32

Miyazaki, Hajime (1977) "The Rat Race and Internal Labor Markets," *The Bell Journal of Economics*, 8 (2), 394–418, 10.2307/3003294. 32

Munoz-Garcia, Felix and Ana Espinola-Arredondo (2011) "The Intuitive and Divinity Criterion: Interpretation and Step-by-Step Examples," *Journal of Industrial Organization Education*, 5 (1), 1–20, 10.2202/1935-5041.1024. 8, 13, 52

Nelson, Phillip (1974) "Advertising as Information," *Journal of Political Economy*, 82 (4), 729–754, https://www.jstor.org/stable/1837143. 2

Netzer, Nick and Florian Scheuer (2014) "A Game Theoretic Foundation of Competitive Equilibria with Adverse Selection," *International Economic Review*, 55 (2), 399–422, 10.1111/iere.12054. 32

Nöldeke, Georg and Larry Samuelson (1997) "A Dynamic Model of Equilibrium Selection in Signaling Markets," *Journal of Economic Theory*, 73 (1), 118–156, 10.1006/jeth.1996.2239. 31

Rabin, Matthew (1990) "Communication between Rational Agents," *Journal of Economic Theory*, 51 (1), 144–170, 10.1016/0022-0531(90)90055-O. 30, 31

Rabin, Matthew and Joel Sobel (1996) "Deviations, Dynamics, and Equilibrium Refinements," *Journal of Economic Theory*, 68 (1), 1–25, 10.1006/jeth.1996.0001. 31

Reny, Philip J (2025) "Natural Language Equilibrium I: Off-Path Conventions," *Working Paper*. 26, 31

Riley, John G. (1979) "Informational Equilibrium," *Econometrica*, 47 (2), 331–359, 10.2307/1914187. 2, 9, 18

——— (2001) "Silver Signals: Twenty-Five Years of Screening and Signaling," *Journal of Economic Literature*, 39 (2), 432–478, 10.1257/jel.39.2.432. 2

Rothschild, Michael and Joseph Stiglitz (1976) "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *The Quarterly Journal of Economics*, 90 (4), 629–649, 10.2307/1885326. 32

Segal, Ilya (2003) "Coordination and Discrimination in Contracting with Externalities: Divide and Conquer?" *Journal of Economic Theory*, 113 (2), 147–181, 10.1016/S0022-0531(03)00114-5. 32

Segal, Ilya R. and Michael D. Whinston (2000) "Naked Exclusion: Comment," *The American Economic Review*, 90 (1), 296–309, https://www.jstor.org/stable/117295. 32

Sémirat, Stéphan and Françoise Forges (2023) "Forward-Neologism-Proof Equilibrium and Better Response Dynamics," *Working Paper*, https://hal.science/hal-04189188. 31

Sobel, Joel (2009) "Signaling Games," in *Encyclopedia of Complexity and Systems Science*, 8125–8139: Springer, New York, NY, 10.1007/978-0-387-30440-3_481. 2

Sobel, Joel and Ichiro Takahashi (1983) "A Multistage Model of Bargaining," *The Review of Economic Studies*, 50 (3), 411–426, 10.2307/2297673. 2

Spence, Michael (1973) "Job Market Signaling," *The Quarterly Journal of Economics*, 87 (3), 355–374, 10.2307/1882010. 2, 3, 6, 9, 12, 17, 31

Spiegler, Ran (2000) "Extracting Interaction-Created Surplus," *Games and Economic Behavior*, 30 (1), 142–162, 10.1006/game.1999.0713. 32

Umbhauer, Gisèle (1991) "Forward Induction, Consistency and Rationality of ε-Perturbations," *Working Paper*. 31

——— (1997) "Forward Induction, Evolutionary Processes and Behavior Evolution," *Journal of Evolutionary Economics*, 7 (4), 415–433, 10.1007/s001910050051. 31

van Damme, Eric (1989) "Stable Equilibria and Forward Induction," *Journal of Economic Theory*, 48 (2), 476–496, 10.1016/0022-0531(89)90038-0. 30

Verrecchia, Robert E. (1983) "Discretionary Disclosure," *Journal of Accounting and Economics*, 5, 179–194, 10.1016/0165-4101(83)90011-3. 5, 16, 32

Wilson, Charles (1977) "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory*, 16 (2), 167–207, 10.1016/0022-0531(77)90004-7. 32

Winter, Eyal (2004) "Incentives and Discrimination," *American Economic Review*, 94 (3), 764–773, 10.1257/0002828041464434. 32

Young, H. Peyton (1993) "The Evolution of Conventions," *Econometrica*, 61 (1), 57–84, 10.2307/2951778. 31

Zapater, Iñigo (1997) "Credible Proposals in Communication Games," *Journal of Economic Theory*, 72 (1), 173–197, 10.1006/jeth.1996.2206. 31

A Proofs Omitted from the Main Text

We first introduce some lemmas that will be used in the proofs of main theorems later.

Lemma 1. *Let $\sigma \in PSE(G_S)$. If $t < t'$, then $\sigma_S(t) \leq \sigma_S(t')$.*

Proof. Suppose not. Let $m = \sigma_S(t)$ and $m' = \sigma_S(t')$. Let $a = \sigma_R(m)$ and $a' = \sigma_R(m')$. By the equilibrium condition of σ, $u_S(t, m, a) \geq u_S(t, m', a')$ and $u_S(t', m', a') \geq u_S(t', m, a)$. By the single-crossing condition (A3), $m > m'$ implies that $u_S(t', m, a) > u_S(t', m', a')$, which is a contradiction. □

Lemma 2. *Reverse Single-Crossing:*

If $m < m'$ and $t < t'$, then
$u_S(t', m, a) \geq u_S(t', m', a')$ implies that $u_S(t, m, a) > u_S(t, m', a')$.

Proof. Suppose not. By the single-crossing condition (A3), $u_S(t, m, a) \leq u_S(t, m', a')$ implies that $u_S(t', m, a) < u_S(t', m', a')$, which is a contradiction. □

For the next lemma, we introduce an additional notation. We study the game truncated from G_S by restricting the sender's types to be a subset of the original set. Let

$$T^j = \{1, 2, \ldots, j\} \quad \text{and} \quad p^j(t) = p_{T^j}(t).$$

The truncated game G_S^j is defined as the original G_S except that the sender's type space is T^j and the prior distribution p^j is the T^j-conditional belief p_{T^j}. Given any equilibrium $\sigma \in \text{PSE}(G_S)$, we can construct a j-truncated strategy σ^j of the truncated game G_S^j by simply deleting those types higher than j. As long as no type higher than j is pooling with j in the equilibrium σ, we have a j-truncated equilibrium $\sigma^j \in \text{PSE}\left(G_S^j\right)$.

Corollary (Mailath, Okuno-Fujiwara and Postlewaite (1993)). *Suppose* $\tilde{\sigma} \in PSE(G_S)$ *and* $\hat{\sigma}^j \in PSE\left(G_S^j\right)$ *for some* $j < n$. *Suppose further that* $u_S\left(j, \hat{\sigma}^j\right) \geq u_S\left(j, \tilde{\sigma}\right)$. *Then, there exists* $\sigma \in PSE(G_S)$ *such that:*

$$u_S(t, \sigma) \geq u_S\left(t, \hat{\sigma}^j\right) \quad \forall t \leq j,$$
$$u_S(t, \sigma) \geq u_S(t, \tilde{\sigma}) \quad \forall t > j.$$

Lemma 3. *Let* $\overline{\sigma} \in PSE(G_S)$ *be the LMSE. Then, for any* $j \in T$, *and any equilibrium* $\sigma^j \in PSE\left(G_S^j\right)$, *we have* $u_S\left(j, \overline{\sigma}\right) \geq u_S\left(j, \sigma^j\right)$.

Proof. Suppose not. If $j = n$, then $u_S(n, \overline{\sigma}) < u_S(n, \sigma^n)$, which contradicts the definition of $\overline{\sigma}$ as the LMSE. If $j < n$, we have $u_S\left(j, \overline{\sigma}\right) < u_S\left(j, \sigma^j\right)$. Then, by the previous corollary, there exists a new equilibrium $\tilde{\sigma} \in \mathrm{PSE}\left(G_S\right)$ such that

$$u_S(t, \tilde{\sigma}) \geq u_S\left(t, \sigma^j\right) \quad \forall t \leq j,$$
$$u_S(t, \tilde{\sigma}) \geq u_S(t, \overline{\sigma}) \quad \forall t > j.$$

In particular, $u_S\left(j, \tilde{\sigma}\right) \geq u_S\left(j, \sigma^j\right) > u_S\left(j, \overline{\sigma}\right)$. Hence, $\tilde{\sigma}$ lex-dominates $\overline{\sigma}$, which contradicts the definition of $\overline{\sigma}$ as the LMSE. □

A.1 Proof of Theorem 1

We denote the LMSE by $\overline{\sigma} \in \mathrm{PSE}\left(G_S\right)$. Consider any other equilibrium $\sigma \in \mathrm{PSE}\left(G_S\right)$, in which $u_S(t, \sigma) \neq u_S(t, \overline{\sigma})$ for some $t \in T$. To show that $\overline{\sigma}$ is most persuasive, we need to prove that $\overline{\sigma}$ is more persuasive than σ. We first identify the message $\overline{m}$, whose interpretation in $\overline{\sigma}$ rather than in σ can trigger an unraveling. We define $\overline{m}$ as follows:

$$\begin{aligned} \bar{t} &= \max\left\{t \in T \middle| u_S(t, \overline{\sigma}) > u_S(t, \sigma)\right\}, \\ \overline{m} &= \overline{\sigma}_S\left(\bar{t}\right), \\ T_{\overline{m}}^{\overline{\sigma}} &= \left\{t \in T \middle| \overline{\sigma}_S(t) = \overline{m}\right\}. \end{aligned}$$

We start by showing that there exists a cutoff type $t^* \in T_{\overline{m}}^{\overline{\sigma}}$ such that for any $t \in T_{\overline{m}}^{\overline{\sigma}}$ and $t \geq t^*$, we have $t \in T_{\overline{m}}^{\overline{\sigma} \geq \sigma}$, while for any $t \in T_{\overline{m}}^{\overline{\sigma}}$ and $t < t^*$, we have $t \in T_{\overline{m}}^{\overline{\sigma} < \sigma}$. By the definition of $\overline{\sigma}$ and $\bar{t}$, we have $u_S(t, \overline{\sigma}) = u_S(t, \sigma)$ for all $t > \bar{t}$. Hence, if $t > \bar{t}$ and $t \in T_{\overline{m}}^{\overline{\sigma}}$, we have $t \in T_{\overline{m}}^{\overline{\sigma} \geq \sigma}$. Suppose now that $t' < \bar{t}$ and $t' \in T_{\overline{m}}^{\overline{\sigma} \geq \sigma}$, we claim that for any $t'' > t'$ and $t'' < \bar{t}$, we have t'' strictly prefers $\overline{\sigma}$ to σ. Otherwise, we have

$$u_S\left(t', \overline{m}, \overline{\sigma}_R(\overline{m})\right) = u_S\left(t', \overline{\sigma}\right) \qquad \geq \qquad u_S\left(t', \sigma\right) = u_S\left(t', m', \sigma_R\left(m'\right)\right) \tag{2}$$

$$u_S\left(t'', \overline{m}, \overline{\sigma}_R\left(\overline{m}\right)\right) = u_S\left(t'', \overline{\sigma}\right) \quad \leq \quad u_S\left(t'', \sigma\right) = u_S\left(t'', m'', \sigma_R\left(m''\right)\right) \tag{3}$$

$$u_S\left(\bar{t}, \overline{m}, \overline{\sigma}_R\left(\overline{m}\right)\right) = u_S\left(\bar{t}, \overline{\sigma}\right) \quad > \quad u_S\left(\bar{t}, \sigma\right) = u_S\left(\bar{t}, m, \sigma_R\left(m\right)\right) \tag{4}$$

First, $\overline{m} \geq m''$, otherwise the single-crossing condition (A3) and (3) imply that

$$u_S\left(\bar{t}, \overline{m}, \overline{\sigma}_R\left(\overline{m}\right)\right) < u_S\left(\bar{t}, m'', \sigma_R\left(m''\right)\right) \leq u_S\left(\bar{t}, m, \sigma_R\left(m\right)\right),$$

which contradicts (4). Second, $\overline{m} \leq m''$, otherwise the reverse single-crossing condition (Lemma 2) and (3) imply that

$$u_S\left(t', \overline{m}, \overline{\sigma}_R\left(\overline{m}\right)\right) < u_S\left(t', m'', \sigma_R\left(m''\right)\right) \leq u_S\left(t', m', \sigma_R\left(m'\right)\right),$$

which contradicts (2). Then, we have $\overline{m} = m''$. By monotonicity (A2), (3) implies that

$$u_S\left(\bar{t}, \overline{m}, \overline{\sigma}_R\left(\overline{m}\right)\right) \leq u_S\left(\bar{t}, \overline{m}, \sigma_R\left(\overline{m}\right)\right) \leq u_S\left(\bar{t}, m, \sigma_R\left(m\right)\right),$$

which contradicts (4).

Hence, there exists a cutoff type $t^* \in T_{\overline{m}}^{\overline{\sigma}}$ such that for any $t \in T_{\overline{m}}^{\overline{\sigma}}$ and $t \geq t^*$, we have $t \in T_{\overline{m}}^{\overline{\sigma} \geq \sigma}$, while for any $t \in T_{\overline{m}}^{\overline{\sigma}}$ and $t < t^*$, we have $t \in T_{\overline{m}}^{\overline{\sigma} < \sigma}$.

Next, we show that the interpretation of $\overline{m}$ in σ rather than in σ can trigger an unraveling under the original ranking, i.e., when f is the identity function. We prove by contradiction. Suppose not and the unraveling condition (1) is violated at some type $j \in T_{\overline{m}}^{\overline{\sigma} < \sigma}$ at some message m_j. Observe that $j < n$. Following the notation in Definition 2, we have

$$u_S\left(j, \overline{\sigma}\right) < u_S\left(j, m_j, \mathrm{BR}\left(m_j, \mu'\right)\right) = u_S\left(j, m_j, \mathrm{BR}\left(m_j, T_{m_j}^{\sigma} \cap \{t \leq j\}\right)\right).$$

The equality follows from the fact that for any $t > j$, we have either $t \in T_{\overline{m}}^{\overline{\sigma} \geq \sigma}$ or $t \in F_{\overline{m}}^{\overline{\sigma} < \sigma}\left(j\right)$. Then, we look at the j-truncated game G_{S}^{j} where the sender's type space T^j is $\{1, 2, \ldots, j\}$ and the prior distribution p^j is the T^j-conditional belief p_{T^j}. We claim that there exists an j-truncated equilibrium $\sigma^j \in \mathrm{PSE}\left(G_{\mathrm{S}}^{j}\right)$ such that

$$u_S\left(j, \sigma^j\right) \geq u_S\left(j, m_j, \mathrm{BR}\left(m_j, \mu'\right)\right) > u_S\left(j, \overline{\sigma}\right),$$

which leads to a contradiction with the fact that $\overline{\sigma}$ is the LMSE. To show the claim, we consider the following two cases.

Case 1: Type 1 does not pool with type j in the equilibrium σ. Let $k^* \in T^j$ be the highest type who does not pool with type j in the equilibrium σ. Then $1 \leq k^* < j$ and a k^*-truncated equilibrium σ^{k^*} can be derived from σ by deleting those types higher than

k^* and keeping everything else unchanged. Let m_k^k denote the message sent by type k at the k-truncated equilibrium σ^k. By Lemma 1, we have $m_{k^*}^{k^*} < m_j$. For any $k^* \leq k < j$, let $I_k = \left\{ t \in T^k \middle| \sigma_S^k(t) = m_k^k \right\}$ and $J_k = \{k+1, \ldots, j\}$. Then, $\sigma_R^{k^*}\left(m_{k^*}^{k^*}\right) = \mathrm{BR}\left(m_{k^*}^{k^*}, I_{k^*}\right)$ and $\mathrm{BR}(m_j, \mu') = \mathrm{BR}(m_j, J_{k^*})$. We use induction to show that given the existence of a k-truncated equilibrium σ^k such that

$$u_S\left(t, \sigma^k\right) \geq u_S\left(t, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right) \quad \forall t \in T^k$$

and $m_k^k < m_j$, we can either directly construct a j-truncated equilibrium σ^j that satisfies our claim or indirectly construct a $k+1$-truncated equilibrium σ^{k+1} such that

$$u_S\left(t, \sigma^{k+1}\right) \geq u_S\left(t, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right) \quad \forall t \in T^{k+1}$$

and $m_{k+1}^{k+1} < m_j$. When $k+1 = j$, we also prove our claim. Given the definition of σ^{k^*} and the equilibrium condition of σ, we know that

$$u_S\left(t, \sigma^{k^*}\right) = u_S(t, \sigma) \geq u_S\left(t, m_j, \sigma_R\left(m_j\right)\right) \geq u_S\left(t, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right) \quad \forall t \in T^{k^*}.$$

For any $k^* \leq k < j$, we consider type $k+1$. By definition, type $k+1$ pools with type j at σ, i.e., $\sigma_S(k+1) = m_j$. There are two cases:

- Case 1.1: $u_S\left(k+1, m_k^k, \mathrm{BR}\left(m_k^k, I_k\right)\right) < u_S\left(k+1, m_j, \mathrm{BR}\left(m_j, J_k\right)\right)$. Now we construct a j-truncated equilibrium σ^j based on σ^k by pooling types higher than k together. Since $k \geq k^*$ and J_k-conditional belief first-order stochastic dominates J_{k^*}-conditional belief, we have

  $$u_S\left(k+1, m_j, \mathrm{BR}\left(m_j, J_k\right)\right) \geq u_S\left(k+1, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right).$$

  By the equilibrium condition of σ and A4, we have

  $$\begin{aligned} & u_S(k+1, \sigma) = u_S\left(k+1, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right) \\ \geq & u_S\left(k+1, m^l, \sigma_R\left(m^l\right)\right) \geq u_S\left(k+1, m^l, \mathrm{BR}\left(m^l, \{1\}\right)\right) \\ > & u_S\left(k+1, m^h, \mathrm{BR}\left(m^h, \{n\}\right)\right) \geq u_S\left(k+1, m^h, \mathrm{BR}\left(m^h, J_k\right)\right) \end{aligned}$$

  The above inequalities are due to the fact that the $\{n\}$-conditional belief first-order stochastic dominates J_k-conditional belief, and $\{1\}$-conditional belief is the worst belief. By the same argument, we have

  $$u_S\left(k+1, m_k^k, \mathrm{BR}\left(m_k^k, I_k\right)\right) > u_S\left(k+1, m^h, \mathrm{BR}\left(m^h, J_k\right)\right).$$

Because $u_S(k+1, m, \mathrm{BR}(m, J_k))$ is continuous in m, there exists a message $m_j^j \in \left[m_j, m^h\right)$ such that

$$\begin{aligned} & u_S\left(k+1, m_j^j, \mathrm{BR}\left(m_j^j, J_k\right)\right) \\ =& \max\left\{u_S\left(k+1, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right), u_S\left(k+1, m_k^k, \mathrm{BR}\left(m_k^k, I_k\right)\right)\right\}, \end{aligned}$$

where $m_j^j = m_j$ if and only if $k = k^*$.

The strategy-belief profile σ^j in the j-truncated game G_S^j is constructed based on the k-truncated equilibrium σ^k as follows:

- $\forall t < k+1$, $\sigma_S^j(t) = \sigma_S^k(t)$;
- $\forall t \in J_k$, $\sigma_S^j(t) = m_j^j$;
- $\forall m \leq m_k^k$, $\mu^j(\cdot \mid m) = \mu^k(\cdot \mid m)$ and $\sigma_R^j(m) = \sigma_R^k(m)$.
- $\mu^j\left(\cdot \mid m_j^j\right) = p_{J_k}$ and $\sigma_R^j\left(m_j^j\right) = \mathrm{BR}\left(m_j^j, J_k\right)$.
- $\forall m > m_k^k$ and $m \neq m_j^j$, $\mu^j(\cdot \mid m) = p_{\{1\}}$ and $\sigma_R^j(m) = \mathrm{BR}(m, \{1\})$.

Next we check that σ^j is a j-truncated equilibrium:

- $t < k+1$:

  If $m \leq m_k^k$, by the equilibrium condition of σ^k and the definition of σ^j, we have

  $$\begin{aligned} & u_S\left(t, \sigma^j\right) = u_S\left(t, \sigma^k\right) \\ \geq & u_S\left(t, m, \mathrm{BR}\left(m, \mu^k\right)\right) = u_S\left(t, m, \mathrm{BR}\left(m, \mu^j\right)\right). \end{aligned}$$

  If $m = m_j^j$, by the equilibrium condition of σ^k and the reverse single-crossing condition ($m_k^k < m_j \leq m_j^j$), we have either

  $$\begin{aligned} & u_S\left(t, \sigma^j\right) = u_S\left(t, \sigma^k\right) \\ \geq & u_S\left(t, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right) \geq u_S\left(t, m_j^j, \mathrm{BR}\left(m_j^j, J_k\right)\right) \end{aligned}$$

  when $u_S\left(k+1, m_j^j, \mathrm{BR}\left(m_j^j, J_k\right)\right) = u_S\left(k+1, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right)$ or

  $$\begin{aligned} & u_S\left(t, \sigma^j\right) = u_S\left(t, \sigma^k\right) \\ \geq & u_S\left(t, m_k^k, \mathrm{BR}\left(m_k^k, I_k\right)\right) \geq u_S\left(t, m_j^j, \mathrm{BR}\left(m_j^j, J_k\right)\right) \end{aligned}$$

  when $u_S\left(k+1, m_j^j, \mathrm{BR}\left(m_j^j, J_k\right)\right) = u_S\left(k+1, m_k^k, \mathrm{BR}\left(m_k^k, I_k\right)\right)$.

If $m > m_k^k$ and $m \neq m_j^j$, by the equilibrium condition of σ^k and monotonicity (A2), we have

$$\begin{aligned} & u_S\left(t, \sigma^j\right) = u_S\left(t, \sigma^k\right) \\ \geq & u_S\left(t, m, \mathrm{BR}\left(m, \mu^k\right)\right) \geq u_S\left(t, m, \mathrm{BR}\left(m, \{1\}\right)\right). \end{aligned}$$

– $t \in J_k$:

If $m \leq m_k^k$, by the definition of m_j^j and the single-crossing condition, we have

$$\begin{aligned} & u_S\left(t, \sigma^j\right) = u_S\left(t, m_j^j, \mathrm{BR}\left(m_j^j, J_k\right)\right) \\ \geq & u_S\left(t, m_k^k, \mathrm{BR}\left(m_k^k, I_k\right)\right) \\ \geq & u_S\left(t, m, \mathrm{BR}\left(m, \mu^k\right)\right) = u_S\left(t, m, \mathrm{BR}\left(m, \mu^j\right)\right). \end{aligned}$$

If $m > m_k^k$ and $m \neq m_j^j$, by the equilibrium condition of σ^k and monotonicity, we have

$$\begin{aligned} & u_S\left(t, \sigma^j\right) = u_S\left(t, m_j^j, \mathrm{BR}\left(m_j^j, J_k\right)\right) \\ \geq & u_S\left(t, m_k^k, \mathrm{BR}\left(m_k^k, I_k\right)\right) \\ \geq & u_S\left(t, m^l, \mathrm{BR}\left(m^l, \{1\}\right)\right) \geq u_S\left(t, m, \mathrm{BR}\left(m, \{1\}\right)\right). \end{aligned}$$

Hence, σ^j is a j-truncated equilibrium. Notice that by construction, we have

$$\begin{aligned} & u_S\left(t, \sigma^j\right) = u_S\left(t, \sigma^k\right) \geq u_S\left(t, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right) \quad \forall t < k+1 \\ & u_S\left(t, \sigma^j\right) \geq u_S\left(t, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right) \quad \forall t \in J_k. \end{aligned}$$

In particular, we have $u_S\left(t, \sigma^j\right) \geq u_S\left(t, m_j, \mathrm{BR}\left(m_j, \mu'\right)\right)$.

- Case 1.2: $u_S\left(k+1, m_k^k, \mathrm{BR}\left(m_k^k, I_k\right)\right) \geq u_S\left(k+1, m_j, \mathrm{BR}\left(m_j, J_k\right)\right)$. Now we only need to show that there exists a $k+1$-truncated equilibrium σ^{k+1} such that

$$u_S\left(t, \sigma^{k+1}\right) \geq u_S\left(t, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right) \quad \forall t \in T^{k+1}$$

and $m_{k+1}^{k+1} < m_j$. If $k+1 = j$, we prove the claim. Otherwise, we can replace $k+1$, I_k and J_k by $k+2$, $I_{k+1} = I_k \cup \{k+1\}$ and $J_{k+1} = J_k \setminus \{k+1\}$ respectively, repeat the same analysis for type $k+2$ and so on until we reach type j.

We construct a $k+1$-truncated equilibrium σ^{k+1} based on the k-truncated equilibrium

σ^k by letting type $k+1$ pool with type k. Denote $l = \min\{t \in I_k\}$. Notice that

$$u_S\left(l, m_k^k, \mathrm{BR}\left(m_k^k, I_{k+1}\right)\right) > u_S\left(l, m_k^k, \mathrm{BR}\left(m_k^k, I_k\right)\right) = u_S\left(l, \sigma^k\right)$$
$$u_S\left(l, m_j, \mathrm{BR}\left(m_j, I_{k+1}\right)\right) < u_S\left(l, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right) \leq u_S\left(l, \sigma^k\right)$$

The first inequality follows from the I_{k+1}-conditional belief first-order stochastically dominating the I_k-conditional belief. The second inequality follows from the J_{k^*}-conditional belief first-order stochastically dominating the I_{k+1}-conditional belief. The last equality follows from the property of σ^k.

Because $u_S\left(l, m, \mathrm{BR}\left(m, I_{k+1}\right)\right)$ is continuous in m, there exists a $m_{k+1}^{k+1} \in \left(m_k^k, m_j\right)$ such that

$$u_S\left(l, m_{k+1}^{k+1}, \mathrm{BR}\left(m_{k+1}^{k+1}, I_{k+1}\right)\right) = u_S\left(l, \sigma^k\right)$$

The strategy-belief profile σ^{k+1} in the $k+1$-truncated game G_{S}^{k+1} is constructed based on the k-truncated equilibrium σ^k as follows:

- $\forall t < l$, $\sigma_S^{k+1}(t) = \sigma_S^k(t)$;
- $\forall t \in I_{k+1}$, $\sigma_S^{k+1}(t) = m_{k+1}^{k+1}$;
- $\forall m < m_{k+1}^{k+1}$, $\mu^{k+1}(\cdot \mid m) = \mu^k(\cdot \mid m)$ and $\sigma_R^{k+1}(m) = \sigma_R^k(m)$.
- $\mu^{k+1}\left(\cdot \mid m_{k+1}^{k+1}\right) = p_{I_{k+1}}$ and $\sigma_R^{k+1}\left(m_{k+1}^{k+1}\right) = \mathrm{BR}\left(m_{k+1}^{k+1}, I_{k+1}\right)$.
- $\forall m > m_{k+1}^{k+1}$, $\mu^{k+1}(\cdot \mid m) = p_{\{1\}}$ and $\sigma_R^{k+1}(m) = \mathrm{BR}(m, \{1\})$.

Next we check that σ^{k+1} is a $k+1$-truncated equilibrium:

- $t < l$:

  If $m < m_{k+1}^{k+1}$, by the equilibrium condition of σ^k and the definition of σ^{k+1}, we have

  $$\begin{aligned} u_S\left(t, \sigma^{k+1}\right) &= u_S\left(t, \sigma^k\right) \\ &\geq u_S\left(t, m, \mathrm{BR}\left(m, \mu^k\right)\right) = u_S\left(t, m, \mathrm{BR}\left(m, \mu^{k+1}\right)\right). \end{aligned}$$

  If $m = m_{k+1}^{k+1}$, by the equilibrium condition of σ^k and the reverse single-crossing condition ($m_k^k < m_{k+1}^{k+1}$), we have

  $$\begin{aligned} u_S\left(t, \sigma^{k+1}\right) &\geq u_S\left(t, m_k^k, \mathrm{BR}\left(m_k^k, I_k\right)\right) \\ &> u_S\left(t, m_{k+1}^{k+1}, \mathrm{BR}\left(m_{k+1}^{k+1}, I_{k+1}\right)\right). \end{aligned}$$

If $m > m_{k+1}^{k+1}$, by the equilibrium condition of σ^k and monotonicity, we have

$$\begin{aligned} & u_S\left(t, \sigma^{k+1}\right) = u_S\left(t, \sigma^k\right) \\ \geq & u_S\left(t, m, \mathrm{BR}\left(m, \mu^k\right)\right) \geq u_S\left(t, m, \mathrm{BR}\left(m, \{1\}\right)\right). \end{aligned}$$

– $t \in I_k$:

If $m < m_{k+1}^{k+1}$, by the equilibrium condition of σ^k and the single-crossing condition, we have

$$\begin{aligned} & u_S\left(t, \sigma^{k+1}\right) = u_S\left(t, m_{k+1}^{k+1}, \mathrm{BR}\left(m_{k+1}^{k+1}, I_{k+1}\right)\right) \\ \geq & u_S\left(t, m_k^k, \mathrm{BR}\left(m_k^k, I_k\right)\right) = u_S\left(t, \sigma^k\right) \\ \geq & u_S\left(t, m, \mathrm{BR}\left(m, \mu^k\right)\right) = u_S\left(t, m, \mathrm{BR}\left(m, \mu^{k+1}\right)\right). \end{aligned}$$

If $m > m_{k+1}^{k+1}$, by the equilibrium condition of σ^k and monotonicity, we have

$$\begin{aligned} & u_S\left(t, \sigma^{k+1}\right) \geq u_S\left(t, \sigma^k\right) \\ \geq & u_S\left(t, m, \mathrm{BR}\left(m, \mu^k\right)\right) \geq u_S\left(t, m, \mathrm{BR}\left(m, \{1\}\right)\right). \end{aligned}$$

– $t = k + 1$:

If $m < m_{k+1}^{k+1}$, by the single-crossing condition and the previous results for $t \in I_k$, we have

$$\begin{aligned} & u_S\left(k+1, \sigma^{k+1}\right) = u_S\left(k+1, m_{k+1}^{k+1}, \mathrm{BR}\left(m_{k+1}^{k+1}, I_{k+1}\right)\right) \\ \geq & u_S\left(k+1, m, \mathrm{BR}\left(m, \mu^k\right)\right) = u_S\left(k+1, m, \mathrm{BR}\left(m, \mu^{k+1}\right)\right). \end{aligned}$$

If $m > m_{k+1}^{k+1}$, by the equilibrium condition of σ^k and the single-crossing condition ($m_{k+1}^{k+1} > m^l$), we have

$$\begin{aligned} & u_S\left(k+1, \sigma^{k+1}\right) = u_S\left(k+1, m_{k+1}^{k+1}, \mathrm{BR}\left(m_{k+1}^{k+1}, I_{k+1}\right)\right) \\ > & u_S\left(k+1, m^l, \mathrm{BR}\left(m^l, \{1\}\right)\right) \geq u_S\left(k+1, m, \mathrm{BR}\left(m, \{1\}\right)\right). \end{aligned}$$

Hence, σ^{k+1} is a $k+1$-truncated equilibrium. Notice that by construction, we have

$$\begin{aligned} & u_S\left(t, \sigma^{k+1}\right) = u_S\left(t, \sigma^k\right) \geq u_S\left(t, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right) \quad \forall t < l \\ & u_S\left(t, \sigma^{k+1}\right) = u_S\left(t, m_{k+1}^{k+1}, \mathrm{BR}\left(m_{k+1}^{k+1}, I_{k+1}\right)\right) \end{aligned}$$

$$\geq u_S\left(t, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right) \quad \forall t \in I_{k+1}.$$

The second inequality follows from the fact that

$$u_S\left(l, \sigma^k\right) \geq u_S\left(l, m_j, \mathrm{BR}\left(m_j, J_{k^*}\right)\right)$$
$$u_S\left(l, \sigma^k\right) = u_S\left(l, m_{k+1}^{k+1}, \mathrm{BR}\left(m_{k+1}^{k+1}, I_{k+1}\right)\right)$$

and the single-crossing condition.

In particular, we have $u_S\left(k+1, \sigma^{k+1}\right) \geq u_S\left(t, m_j, \mathrm{BR}\left(m_j, \mu'\right)\right)$. If $k+1 = j$, we prove the claim. Otherwise, if

$$\begin{aligned} & u_S\left(k+2, m_{k+1}^{k+1}, \mathrm{BR}\left(m_{k+1}^{k+1}, I_{k+1}\right)\right) \\ \geq & u_S\left(k+2, m_j, \mathrm{BR}\left(m_j, J_{k+1}\right)\right), \end{aligned}$$

we move to Case 1.2 and construct a $k+2$-truncated equilibrium based on σ^{k+1} as before. If

$$\begin{aligned} & u_S\left(k+2, m_{k+1}^{k+1}, \mathrm{BR}\left(m_{k+1}^{k+1}, I_{k+1}\right)\right) \\ < & u_S\left(k+2, m_j, \mathrm{BR}\left(m_j, J_{k+1}\right)\right), \end{aligned}$$

we move to Case 1.1 and directly construct a j-truncated equilibrium σ^j based on σ^{k+1} as before.

Case 2: Type 1 pools with type j in the equilibrium σ. Notice that we can construct a $\{1\}$-truncated equilibrium σ^1 by letting the sender send the message m^l and assigning the $\{1\}$-conditional belief to the receiver for every message.

- Case 2.1: $u_S\left(1, m^l, \mathrm{BR}\left(m^l, \{1\}\right)\right) > u_S\left(1, m_j, \mathrm{BR}\left(m_j, \mu'\right)\right)$. Then $m^l < m_j$, and we are in the same situation as Case 1 where $k=1$, $I_1 = \{1\}$, and $J_1 = \{2, \ldots, j\}$. Hence, we can construct a j-truncated equilibrium σ^j based on σ^1.
- Case 2.2: $u_S\left(1, m^l, \mathrm{BR}\left(m^l, \{1\}\right)\right) \leq u_S\left(1, m_j, \mathrm{BR}\left(m_j, \mu'\right)\right)$. Then we can directly construct a strategy-belief profile σ^j by pooling all types together as follows:
  - $\forall t \in T^j$, $\sigma_S^j(t) = m_j$;
  - $\mu^j\left(\cdot \mid m_j\right) = p_{T^j}$ and $\sigma_R^j\left(m_j\right) = \mathrm{BR}\left(m_j, T^j\right) = \mathrm{BR}\left(m_j, \mu'\right)$.
  - $\forall m \neq m_j$, $\mu^j\left(\cdot \mid m\right) = p_{\{1\}}$ and $\sigma_R^j(m) = \mathrm{BR}\left(m, \{1\}\right)$.

We check that σ^j is a j-truncated equilibrium: for all $t \in T^j$ and all $m \neq m_j$, we have

$$u_S\left(t, m_j, \mathrm{BR}\left(m_j, T^j\right)\right) \geq u_S\left(t, m^l, \mathrm{BR}\left(m^l, \{1\}\right)\right) \geq u_S\left(t, m, \mathrm{BR}\left(m, \{1\}\right)\right),$$

which follows from the single-crossing condition and A4. In particular, we have $u_S\left(j, \sigma^j\right) \geq u_S\left(j, m_j, \mathrm{BR}\left(m_j, \mu'\right)\right)$.

We have shown that if the unraveling condition (1) is violated at type j, then there exists a j-truncated equilibrium σ^j such that

$$u_S\left(j, \sigma^j\right) \geq u_S\left(j, m_j, \mathrm{BR}\left(m_j, \mu'\right)\right) > u_S\left(j, \overline{\sigma}\right).$$

However, this contradicts the fact that $\overline{\sigma}$ is the LMSE (Lemma 3). Therefore, we conclude that the interpretation of $\overline{m}$ in $\overline{\sigma}$ and σ rather than triggers an unraveling.

For the LMSE $\overline{\sigma}$ and any other equilibrium σ, we have shown the existence of a message $\overline{m}$ such that the interpretation of $\overline{m}$ in $\overline{\sigma}$ rather than σ triggers an unraveling. Hence, $\overline{\sigma}$ is most persuasive.

A.2 Proof of Theorem 2

We prove by contradiction. Suppose that there exists another most persuasive equilibrium $\hat{\sigma}$ in the game G_{S} that is not payoff-equivalent for the sender to the LMSE $\overline{\sigma}$ (Definition 4). Then, by the definition of most persuasive equilibrium (Definition 5), there exists a message $\hat{m}$ such that the interpretation of $\hat{m}$ in $\hat{\sigma}$ rather than $\overline{\sigma}$ triggers an unraveling. In particular, there exists a type $i \in T$ who sends $\hat{m}$ in the equilibrium $\hat{\sigma}$ and strictly prefers $\hat{\sigma}$ to $\tilde{\sigma}$, i.e., $u_S\left(i, \hat{\sigma}\right) > u_S\left(i, \overline{\sigma}\right)$. Now we show that the interpretation of $\hat{m}$ in $\hat{\sigma}$ rather than $\overline{\sigma}$ can never trigger an unraveling irrespective of the ranking function f, which contradicts the fact that $\hat{\sigma}$ is most persuasive.

The first observation is that type i must pool with higher types in the equilibrium $\hat{\sigma}$. Otherwise, $\hat{\sigma}$ induces a i-truncated equilibrium $\hat{\sigma}^i$ where $u_S\left(i, \hat{\sigma}^i\right) > u_S\left(i, \overline{\sigma}\right)$, which contradicts the fact that $\overline{\sigma}$ is the LMSE (Lemma 3). Secondly, if we denote by $j > i$ the highest type pooling with type i in the equilibrium $\hat{\sigma}$, then there must exist $i < k \leq j$ such that $u_S\left(k, \hat{\sigma}\right) < u_S\left(k, \overline{\sigma}\right)$. Let $\overline{m}_k$ denote the message sent by type k in the LMSE $\overline{\sigma}$. Following the notation of Definition 2, we replace σ and $\overline{\sigma}$ by $\overline{\sigma}$ and $\hat{\sigma}$ respectively. For any ranking function f, let

$$t^* = \arg \max_{t \in T_{\hat{m}}^{\hat{\sigma} < \overline{\sigma}} \cap T_{\overline{m}_k}^{\overline{\sigma}}} f\left(t\right).$$

Then, by the definition of t^*, we have $F_{\hat{m}}^{\hat{\sigma}<\overline{\sigma}}(t^*) \cap T_{\overline{m}_k}^{\overline{\sigma}} = \emptyset$. Our goal is to show that

$$u_S(t^*, \hat{\sigma}) < u_S(t^*, \overline{m}_k, \mathrm{BR}(\overline{m}_k, \overline{\mu}^*)) = u_S\left(t^*, \overline{m}_k, \mathrm{BR}\left(\overline{m}_k, U_{\overline{m}_k}^{\overline{\sigma}}\right)\right).$$

In other words, the interpretation of $\hat{m}$ in $\hat{\sigma}$ rather than $\overline{\sigma}$ cannot trigger an unraveling, and the unraveling process has an early stop when reaching type t^*.

Notice that

$$u_S(i, \hat{m}, \hat{\sigma}_R(\hat{m})) = u_S(i, \hat{\sigma}) > u_S(i, \overline{\sigma}) \geq u_S(i, \overline{m}_k, \overline{\sigma}_R(\overline{m}_k)) \tag{5}$$

$$u_S(k, \hat{m}, \hat{\sigma}_R(\hat{m})) = u_S(k, \hat{\sigma}) < u_S(k, \overline{\sigma}) = u_S(k, \overline{m}_k, \overline{\sigma}_R(\overline{m}_k)) \tag{6}$$

First, we claim that $\hat{m} < \overline{m}_k$. Otherwise, if $\hat{m} = \overline{m}_k$, then by monotonicity, (5) implies that the receiver must take a higher action in $\hat{\sigma}$ than in $\overline{\sigma}$ after observing the messages $\hat{m}$ and $\overline{m}_k$ respectively, which contradicts (6); if $\hat{m} > \overline{m}_k$, by the single-crossing condition, (5) implies that $u_S(k, \hat{\sigma}) > u_S(k, \overline{\sigma})$, which contradicts (6) again. Hence, we have $\hat{m} < \overline{m}_k$.

Next we claim that $\min\left\{t \in T_{\hat{m}}^{\hat{\sigma}}\right\} \leq \min\left\{t \in T_{\overline{m}_k}^{\overline{\sigma}}\right\}$. Otherwise, $i \in T_{\overline{m}_k}^{\overline{\sigma}}$. Given that $u_S(k, \hat{\sigma}) < u_S(k, \overline{\sigma})$ and $u_S(i, \hat{\sigma}) > u_S(i, \overline{\sigma})$, by the single-crossing condition, there exists $i \leq i' < k$ such that: (1) for any type $t \in T_{\hat{m}}^{\hat{\sigma}}$ such that $t \leq i'$, we have $u_S(t, \hat{\sigma}) \geq u_S(t, \overline{\sigma})$; (2) for any type $t \in T_{\hat{m}}^{\hat{\sigma}}$ such that $t > i'$, we have $u_S(t, \hat{\sigma}) < u_S(t, \overline{m}_k, \overline{\sigma}_R(\overline{m}_k)) \leq u_S(t, \overline{\sigma})$. Let $i'' = \max\left\{t \in T \middle| t < \min\left\{t \in T_{\hat{m}}^{\hat{\sigma}}\right\}\right\}$. We have

$$u_S(i'', \hat{\sigma}) \geq u_S(i'', \hat{m}, \hat{\sigma}_R(\hat{m})) > u_S(i'', \overline{m}_k, \overline{\sigma}_R(\overline{m}_k)) = u_S(i'', \overline{\sigma}).$$

Then, $\hat{\sigma}$ induces a i''-truncated equilibrium $\hat{\sigma}^{i''}$ where $u_S\left(i'', \hat{\sigma}^{i''}\right) > u_S(i'', \overline{\sigma})$, which contradicts the fact that $\overline{\sigma}$ is the LMSE (Lemma 3).

Given that $u_S(k, \hat{\sigma}) < u_S(k, \overline{\sigma})$ and $\min\left\{t \in T_{\hat{m}}^{\hat{\sigma}}\right\} \leq \min\left\{t \in T_{\overline{m}_k}^{\overline{\sigma}}\right\}$, by the single-crossing condition, there exists a cutoff type $k' < k$ such that for any type $t \in T_{\hat{m}}^{\hat{\sigma}} \cap T_{\overline{m}_k}^{\overline{\sigma}}$ and $t \leq k'$, we have $u_S(t, \hat{\sigma}) \geq u_S(t, \overline{\sigma})$.[33] Then,

$$\begin{aligned}
T_{\hat{m}}^{\hat{\sigma}\geq\overline{\sigma}} &= T_{\hat{m}}^{\hat{\sigma}\geq\overline{\sigma}} \cap \{t \in T | t \leq k'\} \\
U_{\overline{m}_k}^{\overline{\sigma}} &= T_{\overline{m}_k}^{\overline{\sigma}} \setminus \left(F_{\hat{m}}^{\hat{\sigma}<\overline{\sigma}}(t^*) \cup T_{\hat{m}}^{\hat{\sigma}\geq\overline{\sigma}}\right) = T_{\overline{m}_k}^{\overline{\sigma}} \cap \{t \in T | t > k'\} \\
u_S(t^*, \hat{\sigma}) &< u_S(t^*, \overline{\sigma}) \leq u_S\left(t^*, \overline{m}_k, \mathrm{BR}\left(\overline{m}_k, U_{\overline{m}_k}^{\overline{\sigma}}\right)\right),
\end{aligned}$$

which implies that the unraveling process stops early when reaching type t^*.

Therefore, we have shown that the interpretation of $\hat{m}$ in $\hat{\sigma}$ rather than $\overline{\sigma}$ can never trigger an unraveling irrespective of the ranking function, which contradicts the fact that $\hat{\sigma}$ is

[33] $k' < \min\left\{t \in T_{\overline{m}_k}^{\overline{\sigma}}\right\}$ implies that $u_S(t, \hat{\sigma}) < u_S(t, \overline{\sigma})$ for any type $t \in T_{\hat{m}}^{\hat{\sigma}} \cap T_{\overline{m}_k}^{\overline{\sigma}}$.

most persuasive. Hence, the most persuasive equilibrium is unique up to payoff equivalence for the sender, which is determined by the LMSE.

A.3 Proof of Theorem 3

Suppose that there are two most persuasive equilibria $\hat{\sigma}$ and $\overline{\sigma}$ in the game G_{S}. By Theorem 2, we know that they are payoff-equivalent for the sender, i.e., $u_S(t, \hat{\sigma}) = u_S(t, \overline{\sigma})$ for all $t \in T$, and they are both LMSE. To show the uniqueness of the most persuasive equilibrium outcome, we only need to show that $\hat{\sigma}_S(t) = \overline{\sigma}_S(t)$ for all $t \in T$. When the sender's equilibrium strategy is pinned down, so does the equilibrium outcome.[34]

Notice that $\hat{\sigma}$ and $\overline{\sigma}$ must induce the same partition of the type space. If not, perturbing the prior will affect the posterior after each message, and thus the equality in payoffs will vanish, which implies that the most persuasive equilibrium outcome is generically unique. Now we show that the sender of the same type must send the same message in both equilibria. Suppose not. If $\hat{\sigma}_S(j) = \hat{m}_j \neq \overline{m}_j = \overline{\sigma}_S(j)$ for some $j \in T$, then

$$u_S(j, \hat{m}_j, \hat{\sigma}_R(\hat{m}_j)) = u_S(j, \overline{m}_j, \overline{\sigma}_R(\overline{m}_j))$$

implies that

$$u_S(i, \hat{m}_j, \hat{\sigma}_R(\hat{m}_j)) \neq u_S(i, \overline{m}_j, \overline{\sigma}_R(\overline{m}_j))$$

if type j pools with any type $i \neq j$ by the single-crossing condition, which contradicts the assumption that both equilibria generate the same equilibrium payoff for type i. Hence, type j does not pool with any other type in both equilibria.

Consider the j-truncated game G_{S}^j. Since type j does not pool with any other type, $\overline{\sigma}$ and $\hat{\sigma}$ induce two j-truncated equilibria $\overline{\sigma}^j$ and $\hat{\sigma}^j$. Strict quasi-concavity implies that $u_S(j, m, \mathrm{BR}(m, \{j\})) > u_S(j, \overline{\sigma}^j) = u_S(j, \hat{\sigma}^j)$ for all m between $\overline{m}_j$ and $\hat{m}_j$.

If there is no type lower than j, i.e., $j = 1$, then we can construct another j-truncated equilibrium $\tilde{\sigma}^j$ by letting type j sending some message between $\overline{m}_j$ and $\hat{m}_j$. Then, type j achieves a higher payoff in $\tilde{\sigma}^j$ than in $\overline{\sigma}^j$, contradicting the assumption that $\overline{\sigma}$ is the LMSE.

If there exist types lower than j, then

$$u_S\left(j-1, \overline{m}_j, \overline{\sigma}_R^j(\overline{m}_j)\right) \neq u_S\left(j-1, \hat{m}_j, \hat{\sigma}_R^j(\hat{m}_j)\right).$$

Otherwise, the single-cross condition implies that $u_S(j, \overline{\sigma}^j) \neq u_S(j, \hat{\sigma}^j)$, which contradicts

[34] $\hat{\sigma}$ and $\overline{\sigma}$ can differ in terms of the off-path beliefs and what happens following an off-path message.

our assumption. Since $u_S\left(j-1,\overline{\sigma}^j\right)=u_S\left(j-1,\hat{\sigma}^j\right)$, we have either

$$u_S\left(j-1,\overline{\sigma}^j\right)>u_S\left(j-1,\overline{m}_j,\overline{\sigma}_R^j\left(\overline{m}_j\right)\right)$$

or

$$u_S\left(j-1,\hat{\sigma}^j\right)>u_S\left(j-1,\hat{m}_j,\hat{\sigma}_R^j\left(\hat{m}_j\right)\right).$$

Without loss of generality, we can take $\overline{\sigma}^j$ for example. Given that

$$u_S\left(j,m,\mathrm{BR}\left(m,\{j\}\right)\right)>u_S\left(j,\overline{\sigma}^j\right)$$

for all m between $\overline{m}_j$ and $\hat{m}_j$, continuity implies that there exists a message $\tilde{m}_j$ close to $\overline{m}_j$ such that

$$\begin{aligned}u_S\left(j,\tilde{m}_j,\mathrm{BR}\left(\tilde{m}_j,\{j\}\right)\right)&>u_S\left(j,\overline{\sigma}^j\right)\\ u_S\left(j-1,\tilde{m}_j,\mathrm{BR}\left(\tilde{m}_j,\{j\}\right)\right)&<u_S\left(j-1,\overline{\sigma}^j\right).\end{aligned}$$

By the single-crossing condition, we have $u_S\left(t,\overline{\sigma}^j\right)>u_S\left(t,\tilde{m}_j,\mathrm{BR}\left(\tilde{m}_j,\{j\}\right)\right)$ for all $t<j$. Then, we can construct a j-truncated equilibrium $\tilde{\sigma}^j$ based on $\overline{\sigma}^j$ by letting $\tilde{\sigma}^j\left(t\right)=\overline{\sigma}^j\left(t\right)$ for $t<j$, and $\tilde{\sigma}^j\left(j\right)=\tilde{m}_j$. Then, type j achieves a higher payoff in $\tilde{\sigma}^j$ than in $\overline{\sigma}^j$, contradicting the assumption that $\overline{\sigma}$ is the LMSE.

Hence, the sender of the same type must send the same message in the two most persuasive equilibria $\overline{\sigma}$ and $\hat{\sigma}$, and they produce a unique lex max outcome, which is generically the unique most persuasive equilibrium outcome.

B Intuitive Explanation of the Criteria in Table 4

We use the intuitive criterion as the benchmark and compare it with other refinements. We follow the two-step approach in Section 3. The explanations are not meant to provide exact characterizations but to convey the essential intuition underlying each criterion. For ease of comparison, we begin by restating the intuitive criterion.

Intuitive Criterion (Cho and Kreps, 1987)

- Step 1: Which types of the sender *could benefit* by sending an off-path message m?

We denote the set of such types as D. Formally,

$$D = \left\{ t \in T | \, u_S(t, \sigma) \leq \max_{a \in \text{BR}(m, \Delta(T))} u_S(t, m, a) \right\},$$

where $\text{BR}(m, \Delta(T)) = \cup_{\mu \in \Delta(T)} \text{BR}(m, \mu)$.

- Step 2: If deviations only come from the set of types of the sender identified in Step 1, is the *lowest* payoff from deviating higher than their equilibrium payoff for some type of the sender?

  Formally, if there exists $t \in D$ such that

  $$\min_{a \in \text{BR}(m, \Delta(D))} u_S(t, m, a) > u_S(t, \sigma),$$

  then this equilibrium σ fails the intuitive criterion.

D1 Criterion (Banks and Sobel, 1987)

- Step 1: Which types of the sender *are most likely to benefit* by sending an off-path message m?
- Step 2: The same as Step 2 of the intuitive criterion.

Type t_1 is more likely to benefit than type t_2 by sending an off-path message m, if whenever type t_2 can weakly benefit by sending m under some action of the receiver, type t_1 can strictly benefit by sending m under the same action. In the Spencian game, when the high-type worker pools with the medium-type worker in an equilibrium, the high-type worker is more likely to benefit by sending an off-path message than the medium-type worker. Then, in Step 2, we only consider the higher-type worker when checking whether the equilibrium fails the D1 criterion, even though the medium-type worker could also benefit from deviating. Hence, the D1 criterion is stronger than the intuitive criterion, and it uniquely selects the Riley outcome in monotone signaling games.[35]

G-P Criterion (Grossman and Perry, 1986)

- Step 1: The same as Step 1 of the intuitive criterion.

[35] This explanation follows from Munoz-Garcia and Espinola-Arredondo (2011).

- Step 2: If deviations only come from a *subset* $\tilde{D} \subseteq D$ of the types of the sender identified in Step 1, is the *expected* payoff from deviating higher than their equilibrium payoff *only for* $t \in \tilde{D}$? Formally, if there exists $\tilde{D} \subseteq D$ such that

$$\begin{aligned} u_S\left(t, m, \mathrm{BR}\left(m, p_{\tilde{D}}\right)\right) &\geq u_S\left(t, \sigma\right) \quad \forall t \in \tilde{D} \\ u_S\left(\tilde{t}, m, \mathrm{BR}\left(m, p_{\tilde{D}}\right)\right) &> u_S\left(\tilde{t}, \sigma\right) \quad \exists \tilde{t} \in \tilde{D} \\ u_S\left(t, m, \mathrm{BR}\left(m, p_{\tilde{D}}\right)\right) &\leq u_S\left(t, \sigma\right) \quad \forall t \in T \setminus \tilde{D} \end{aligned}$$

  where $p_{\tilde{D}}\left(t\right) = p\left(t \middle| t \in \tilde{D}\right)$ is the $\tilde{D}$-conditional belief, then the equilibrium fails the G-P Criterion.

The G-P criterion is stronger than the intuitive criterion, because the belief applied in Step 2 is less pessimistic than that in the intuitive criterion. As a result, the G-P criterion uniquely select σ^{m_2} in Example 5. However, the G-P criterion suffers from the non-existence problem. In Example 1, the intuitive criterion selects the Riley outcome irrespective of the prior probability p. Then, the only equilibrium outcome that could pass the G-P criterion is the Riley outcome. However, when p is close to zero, the Riley outcome fails the G-P criterion, because both the high-type and low-type workers could benefit from deviating to an education level close to zero, thereby obtaining an expected wage of $2 - p$.

Undefeated Equilibrium (Mailath, Okuno-Fujiwara and Postlewaite, 1993)

σ' is defeated by σ if (1) there exists an *off-path* message m in σ' that is *on-path* in σ such that for *every type* t that sends m in σ, the *equilibrium* payoff in σ is weakly higher than in σ'; and (2) there exists one type t' sending m in σ whose equilibrium payoff is strictly higher than in σ'.

An equilibrium is undefeated if it is not defeated by any other equilibrium. The undefeated equilibrium is typically not unique. In Example 2, both the pooling equilibrium σ^{Pooling} and the LMSE $\overline{\sigma}$ are undefeated. Neither the pooling equilibrium nor the LMSE defeats the other, because the medium-type worker prefers the pooling equilibrium, while the high-type worker prefers the LMSE. Similarly, in Example 4, both pooling equilibria σ^{Beer} and σ^{Quiche} are undefeated, because the surly type prefers σ^{Beer}, while the wimp type prefers σ^{Quiche}.