Robust Neologism Proofness*

Daniel Clark†

November 2, 2021

Abstract

We introduce robust neologism proofness, an equilibrium refinement that applies to both cheap-talk and costly signaling games. Robust neologism proofness eliminates equilibria that can be undone by a certain kind of credible communication from the sender to the receiver, formalized as a “credible robust neologism.” We show that robust neologism proof equilibria exist both in a class of “monotonic” signaling games and any signaling game where the sender can give a transfer to the receiver. Additionally, we apply robust neologism proofness to various examples and compare it with other equilibrium refinements. Finally, we show that in the special class of “monotone-concave-supermodular” signaling games with transfers, robust neologism proofness selects the “sender-optimal” separating equilibria.

Keywords: signaling games, equilibrium refinements, credible communication, robust neologism proofness, signaling games with transfers, monotone-concave-supermodular signaling games with transfers, sender-optimal separating equilibria.

*The author wishes to thank Drew Fudenberg for his invaluable guidance and suggestions, as well as In-Koo Cho, Glenn Ellison, Joseph Farrell, Françoise Forges, Giacomo Lanzani, Elliot Lipnowski, Zhuoran Lu, Stephen Morris, Harry Pei, Debraj Ray, Bruno Strulovici, Alex Wolitzky, and Muhamet Yildiz for helpful comments and conversations.

†Department of Economics, MIT. Email: dgclark@mit.edu.
1 Introduction

Many economic settings have been modeled as signaling games between two players, a sender and receiver. In equilibrium, the sender may be able to transmit information to the receiver through their choice of costly action. However, communication avenues outside of signaling through costly actions may be available. When these avenues are sufficiently flexible, credible communication between the sender and receiver can unravel many equilibria that satisfy traditional refinements like the Intuitive Criterion and D1. This paper develops the equilibrium refinement of “robust neologism proofness,” which refines away equilibria that can be undone by a particular kind of credible communication.

For concreteness, consider the following example.

Example 1. The sender’s type space is $\Theta = \{\theta_1, \theta_2\}$, and the receiver’s prior is that the two types are equally likely. The sender’s action space is $X = \{x_1, x_2\}$, and the receiver’s action space is $Y = \{y_1, y_2, y_3\}$. The payoffs to the sender and receiver are given in Table 1 below. Note that the sender’s action does not affect the payoffs of either the sender or receiver.

<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$\theta_2$</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>2,2</td>
<td>-2, -2</td>
<td>1,1</td>
<td>$x_1$</td>
<td>-2, -2</td>
<td>2,2</td>
<td>1,1</td>
</tr>
<tr>
<td>$x_2$</td>
<td>2,2</td>
<td>-2, -2</td>
<td>1,1</td>
<td>$x_2$</td>
<td>-2, -2</td>
<td>2,2</td>
<td>1,1</td>
</tr>
</tbody>
</table>

Table 1: The payoffs for Example 1. In this table and in all other tables in the paper, the payoff to the sender is the first number in each pair, while the payoff to the receiver is the second.

In this game, the sender and receiver have completely aligned preferences over action profiles given the sender’s type. When the sender’s type is $\theta_1$, the ideal receiver action for both sender and receiver is $y_1$. When the sender’s type is $\theta_2$, the ideal action is $y_2$. 
The equilibrium outcomes of this game are the pooling outcome in which the receiver always plays $y_3$ regardless of the sender’s type, and the separating outcome in which the receiver plays $y_1$ when the sender’s type is $\theta_1$, and the receiver plays $y_2$ when the sender’s type is $\theta_2$. □

Equilibria where the receiver plays $y_3$ in response to both $x_1$ and $x_2$ are allowed by many equilibrium refinements such as the Intuitive Criterion and D1. It seems likely, however, that the sender should be able to credibly signal their type, given the strong alignment of preferences between the sender and receiver and the strong misalignment of preferences between the two sender types. If for instance sender type $\theta_1$ were to declare to the receiver that they were $\theta_1$, the receiver would have strong reason to believe them and consequently take action $y_1$, since only $\theta_1$ would wish to induce this belief and subsequent receiver action.\footnote{A similar argument applies to the credibility of $\theta_2$ declaring their type.} Robust neologism proofness, which refines away the pooling equilibria and selects the separating equilibria in this game, attempts to capture some of the effects of this kind of credible communication in signaling games.

The above example, though illustrative, is quite stark. However, robust neologism proofness is natural in many settings of interest with flexible communication between a sender and a receiver. For example, a firm that offers a performance-based incentive contract to their employee may be more informed than the employee about how the employee’s effort will translate into profit. If the compensation specified by the contract depends on the firm’s profit, then the potential employee’s perception of what the firm knows will be important for their decision of how hard to work, and the informed firm’s choice of incentive scheme signals their knowledge to the worker. It is natural to think that informal communication is prevalent in many organizational settings such as that of a firm and worker where there is close interaction between parties. Robust neologism proofness has significant cutting power in this setting of an informed firm and worker, as we illustrate with a brief example (Example 3) later in the paper.

Robust neologism proofness is not the first communication-based refinement for
signaling games. Neologism-proofness (Farrell [1993]), has a similar motivation of eliminating equilibria that can be undone by credible communication between sender and receiver, and oftentimes the equilibrium selections of these refinements coincide.\(^2\) Robust neologism proofness is much more restrictive in what constitutes credible communication that unravels equilibria. In this paper, we argue that it is often a more sensible refinement, both conceptually and in terms of its properties. One of the issues with neologism-proofness is that it frequently rules out –all– equilibria, including seemingly sensible equilibria that other refinements allow. One example of this is the Spence signaling model (Spence [1973]). Here neologism-proofness can rule out all equilibria, including those with the least-cost separating outcome, which is also known as the Riley outcome. In contrast, most refinements like the Intuitive Criterion preserve the least-cost separating outcome and some, such as D1 and robust neologism proofness, select it as the unique outcome.

We show that robust neologism proof equilibria always exist in two broad classes of signaling games: (1) Monotonic games, where all sender types share the same preferences over receiver actions, and (2) Signaling games with transfers, where the sender’s action includes a transfer to the receiver. In contrast, neologism-proof equilibria need not exist in these games. The proof of the existence result for signaling games with transfers constructs an equilibrium that is a limit of equilibria in a sequence of modified games, and shows that this equilibrium is robust neologism proof. The pre-limit games are modified so that the sender types can use messages that prevent the receiver from playing certain actions, but the payoffs of the sender types from doing this depend on the incentives that other sender types would have to do the same. These modifications are done in such a way that any profile that emerges in the limit both satisfies robust neologism proofness and is an equilibrium of the original game.\(^3\)

We then characterize the robust neologism proof equilibria in the class of “monotone-

\(^2\)Neologism-proofness also selects the separating equilibria in Example 1.

\(^3\)As noted in Subsection 4.2, the existence result for signaling games with transfers extends to other settings, such as those where the sender can exert effort or “burn money” that does not impact the receiver’s utility.
concave-supermodular (MCS)” signaling games with transfers, a more restrictive class
that nests many settings of interest, such as the informed-firm example discussed above.
We prove that robust neologism proofness selects the sender-optimal separating equi-
libria in this class of games, and illustrate this in the context of the informed-firm example.

Furthermore, we compare robust neologism proofness with refinements like the
Intuitive Criterion and D1. Frequently, credible communication gives robust neologism
proofness cutting power that the Intuitive Criterion or D1 lack. Likewise, the Intuitive
Criterion and D1 sometimes eliminate equilibria that are deemed robust neologism proof. As a dramatic example of how different robust neologism proofness can behave
from other refinements, we present a game in which there are robust neologism proof
equilibria, but there is no equilibrium that is both robust neologism proof and D1.
Despite the lack of a general relationship between robust neologism proofness and
the other refinements, in the class of monotonic games, there is a strong relationship
between robust neologism proofness and D1, in that every D1 equilibrium is robust
neologism proof.

2 Related Literature

Of the most commonly used equilibrium refinements for signaling games, the most
permissive is the Intuitive Criterion (Cho and Kreps [1987]) and the most restrictive
is strategic stability (Kohlberg and Mertens [1986]). Kohlberg and Mertens [1986]
proved that there are individual equilibrium outcomes that are strategically stable
in generic extensive-form games; however, strategic stability does not imply robust
neologism proofness. Consequently, even though robust neologism proofness is upper
hemicontinuous in payoffs, arguments such as those in Cho [1987] cannot be used to
prove the existence of robust neologism proof equilibria.

The most commonly used signaling game refinement motivated by the possibility of
credible communication is neologism-proofness. While originally formulated for cheap-
talk games, it can be adapted for costly signaling games as well. When this is done, neologism-proofness is similar to robust neologism proofness, and we will frequently compare the two. Several papers, such as Farrell and Gibbons [1989], Maskin and Tirole [1992], Mylovanov and Tröger [2012, 2014], and Hidir and Vellodi [2018], have applied variants of neologism-proofness in various settings. Lai and Lim [2018] finds some experimental evidence supporting neologism-proofness in a lab setting. Matthews et al. [1991] develops a related refinement, “announcement-proofness,” which eliminates equilibria with credible “announcements” as opposed to credible “neologisms” and is neither stronger nor weaker than neologism-proofness in general. “Invulnerability to credible deviations” (Eső and Schummer [2009]) and the Strong Intuitive Criterion (Riley [2012]) are also related; we discuss them in detail in Section 6.

“Credible message equilibrium” (Rabin [1990]) uses ex-ante behavioral assumptions to eliminate equilibria, in contrast to neologism-proofness and robust neologism proofness, which eliminate equilibria on the basis of credible deviations. Blume and Sobel [1995], Zapater [1997], and Olszewski [2006] also develop refinements for cheap-talk games that are not based on credible deviations from equilibria. Clark and Fudenberg [2021] gives a learning foundation for the communication-based refinement of “justified communication equilibrium (JCE).” JCE has substantial refinement power in many costly signaling games, but does not have any cutting power in cheap-talk games.

Cho and Sobel [1990] studies a broad class of signaling games where only the sender-optimal separating perfect Bayesian equilibria survive D1. Ramey [1996] obtains a similar result in a setting with a continuum of types. We establish an analogous uniqueness result for robust neologism proofness in the setting of “monotone-concave-supermodular” signaling games with transfers.

---

4Another recent experimental paper, de Groot Ruiz et al. [2014], finds some support for “Average Credible Deviation Criterion,” a cheap-talk refinement related to neologism-proofness.

5Eső and Schummer [2009] shows that, in this class of games, the equilibria that are not vulnerable to credible deviations are also precisely those that survive D1.
3 Robust Neologism Proofness

This section presents the definition of robust neologism proofness, shows that it satisfies an upper hemicontinuity property, and applies it to the example game discussed earlier. We first present the framework used in the paper.

3.1 The Model

The sender’s type is \( \theta \in \Theta \), where \( \Theta = \{\theta_1, ..., \theta_N\} \) is a finite type space, and the receiver has a full-support prior belief about \( \theta \), denoted by \( \mu \in \Delta(\Theta) \). The sender takes action \( x \in X \) and the receiver takes action \( y \in Y \), where the action spaces \( X \) and \( Y \) are compact metric spaces. The sender also sends a message \( m \in M \), where the message space \( M \) is also a compact metric space and is large in the sense that \( \Delta(Y) \) can be injectively mapped into \( M \).\(^6\) To lighten the notation, we assume that \( \Delta(Y) \subseteq M \). This reflects a setting with communication avenues sufficiently flexible that the sender at least has the ability to recommend an arbitrary mixed action to the receiver, as is possible with verbal communication, for instance. The sender’s utility function is \( u : \Theta \times X \times Y \to \mathbb{R} \) and the receiver’s utility function is \( v : \Theta \times X \times Y \to \mathbb{R} \), where \( u(\theta, x, y) \) and \( v(\theta, x, y) \) give the payoffs to the sender and receiver respectively when the sender’s type is \( \theta \) and the action profile is \( (x, y) \). Both utility functions are continuous. Note that neither the payoffs of the sender or the receiver depend directly on the sender’s message \( m \).

The timing is as follows. Upon observing their type, the sender takes an action \( x \) and sends a message \( m \). The receiver observes the sender’s choice of both \( x \) and \( m \) and responds with an action \( y \). After this the payoffs are realized.

The strategy space of type \( \theta \) is the space of probability distributions over \( X \times M \), denoted by \( \Delta(X \times M) \).\(^7\) If the strategy of type \( \theta \) is \( \pi_{\theta} \in \Delta(X \times M) \), then \( \pi_{\theta}(\alpha) \) denotes

\(^6\)Manelli [1996] uses this assumption to establish the existence of perfect Bayesian equilibria in signaling games with infinite action spaces; here we use it to prove the existence of robust neologism proof perfect Bayesian equilibria in monotonic signaling games and signaling games with transfers.

\(^7\)Throughout the paper, we denote the set of probability distributions over a set \( \Omega \) by \( \Delta(\Omega) \), and
the probability with which the \((x, m)\) chosen by type \(\theta\) belongs to the measurable set \(\alpha \subseteq X \times M\). The receiver’s strategy can be any measurable function mapping \(X \times M\) into \(\Delta(Y)\). When the receiver’s strategy is \(\pi_R\), \(\pi_R(\beta|x, m)\) denotes the probability with which \(y\) belongs to the measurable set \(\beta \subseteq Y\) when the receiver observes \((x, m)\).

We use Bayes Nash equilibrium as the basic equilibrium concept throughout the paper and refer to Bayes Nash equilibria simply as equilibria. A strategy profile \(\pi = (\pi_{\theta_1}, \ldots, \pi_{\theta_N}, \pi_R)\) constitutes an equilibrium if every sender type maximizes their expected payoff given the receiver’s strategy, and the receiver’s strategy maximizes their expected payoff given the strategies of the sender types. We also use the equilibrium concept of perfect Bayesian equilibrium (PBE, see Fudenberg and Tirole [1991]).

The following notion of PBE adapts the definition of sequential equilibrium in signaling games with infinite action spaces from Manelli [1996]. The strategy profile \(\pi\) constitutes a PBE if it meets three criteria: (1) For every sender type \(\theta\), \(\pi_\theta\) maximizes the expected payoff of type \(\theta\) given the receiver’s strategy \(\pi_R\), (2) For every \((x, m)\), \(\pi_R(\cdot|x, m)\) maximizes the receiver’s expected payoff when the sender takes action \(x\), given the receiver’s posterior belief over the sender’s type after observing \((x, m)\), and (3) There is a regular conditional probability distribution over \(\Theta\) given \(X \times M\) derived from the prior \(\mu\) and the strategy profile of the sender types \((\pi_{\theta_1}, \ldots, \pi_{\theta_N})\) that gives the receiver’s posterior belief over the sender’s type following the observation of any \((x, m)\). Throughout the paper, we abuse notation and denote the expected payoff of type \(\theta\) from strategy profile \(\pi\) by \(u(\theta, \pi)\) and the receiver’s expected payoff from \(\pi\) by \(v(\pi)\).

Robust neologism proofness eliminates equilibria that can be undone by credible communication from sender types to the receiver. An important aspect of whether communication is credible is whether the action it seeks to induce in the receiver is incentive compatible. Naturally, the receiver’s best response correspondence plays an important role in determining this. We denote the set of receiver best responses when the sender takes action \(x\) and the receiver’s belief over the sender’s type is \(\tilde{\mu} \in \Delta(\Theta)\) we imbue this with the topology of weak convergence.
by $BR(\tilde{\mu}, x)$: That is, $BR(\tilde{\mu}, x) = \arg \max_{y \in Y} E_{\tilde{\mu}}[v(\theta, x, y)]$. When the sender takes action $x$ and their type is known to be $\theta$, we denote the set of receiver best responses by $BR(\theta, x)$. Also, for any $\tilde{\Theta} \subseteq \Theta$, we let $BR(\tilde{\Theta}, x) = \cup_{\tilde{\mu} \in \Delta(\tilde{\Theta})} BR(\tilde{\mu}, x)$ be the set of receiver best responses to $x$ for some belief with support contained in $\tilde{\Theta}$, and likewise we let $MBR(\tilde{\Theta}, x) = \{ \beta \in \Delta(Y) | \exists \tilde{\mu} \in \Delta(\tilde{\Theta}) \text{ s.t. } E_{\tilde{\mu}}[v(\theta, x, \beta)] \geq E_{\tilde{\mu}}[v(\theta, x, y)] \forall y \in Y \}$ be the set of mixed receiver actions that are best responses to $x$ for some belief whose support is contained in $\tilde{\Theta}$.

### 3.2 Preliminaries

This subsection presents the definitions of credible robust neologisms and robust neologism proofness, and also shows that the correspondence mapping games to robust neologism proof outcome distributions is upper hemicontinuous.

**Definition 1.** The strategy profile $\pi$ has a **credible robust neologism** if there exists some sender action $x \in X$ and non-empty subset of sender types $\tilde{\Theta} \subseteq \Theta$ such that

1. $\min_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y) > u(\theta, \pi)$ for all $\theta \in \tilde{\Theta}$, and
2. $\max_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y) < u(\theta', \pi)$ for all $\theta' \notin \tilde{\Theta}$.

**Definition 2.** A strategy profile is **robust neologism proof** if it does not have a credible robust neologism.

A strategy profile $\pi$ has a credible robust neologism when there is some sender action $x$ and non-empty subset of sender types $\tilde{\Theta}$ such that, for any receiver best response to $x$ and a belief over sender types whose support is contained in $\tilde{\Theta}$, every sender type in $\Theta$ ($\theta \in \tilde{\Theta}$) obtains a strictly higher payoff than under $\pi$ while every sender type outside $\tilde{\Theta}$ ($\theta' \notin \tilde{\Theta}$) obtains a strictly lower payoff than under $\pi$. Robust neologism proofness eliminates strategy profiles with credible robust neologisms.

The motivation for robust neologism proofness is that when an equilibrium $\pi$ has a credible robust neologism corresponding to $x$ and $\tilde{\Theta}$, a type $\theta \in \tilde{\Theta}$ sender can, instead of
following the equilibrium, play \( x \) and convince the receiver that their type is contained in \( \tilde{\Theta} \). This is because type \( \theta \) could play action \( x \) and tell the receiver that their type is in \( \tilde{\Theta} \) and so the receiver should play only actions in \( BR(\tilde{\Theta}, x) \). Type \( \theta \) could argue that this is credible because the sender types in \( \tilde{\Theta} \) are precisely the types that would prefer such an outcome to the prevailing equilibrium, since for any \( y \in BR(\tilde{\Theta}, x) \), every sender type in \( \tilde{\Theta} \) would obtain a strictly higher payoff from \( (x, y) \) than the equilibrium while all other types \( \theta' \notin \tilde{\Theta} \) would obtain a strictly lower payoff from \( (x, y) \). This is similar to the “informal speech” motivation of the Intuitive Criterion\(^8\), but with the difference that here the sender convinces the receiver of a specific subset of the sender types to which they belong through their action and communication, rather than just convincing the receiver that they are a type for whom a given off-path action is not equilibrium dominated.

The requirements for a strategy profile to have a credible robust neologism are fairly strong, and so robust neologism proofness may not eliminate all the equilibria that are unlikely to prevail in situations with a high degree of communication. Many stronger communication-based equilibrium refinements can be formulated that may seem reasonable for different settings. One possible refinement could eliminate any equilibrium \( \pi \) in which there was some action \( x \), subset of the sender types \( \tilde{\Theta} \), and type \( \theta \in \tilde{\Theta} \) such that

1'. \( \min_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y) > u(\theta, \pi) \), and

2. \( \max_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y) < u(\theta', \pi) \) for all \( \theta' \notin \tilde{\Theta} \).

Unlike robust neologism proofness, where \( \min_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y) > u(\theta, \pi) \) must hold for all \( \theta \) in the subset of sender types \( \tilde{\Theta} \) corresponding to the robust neologism, for this refinement it suffices that this holds for just one \( \theta \in \tilde{\Theta} \). The idea here is that type \( \theta \) could play \( x \) and credibly reveal their type to be in \( \tilde{\Theta} \), because every type outside of \( \tilde{\Theta} \) strictly prefers their outcome in \( \pi \) to \( (x, y) \) for every \( y \in BR(\tilde{\Theta}, x) \). Type \( \theta \) will be better off under \( (x, y) \) for every \( y \in BR(\tilde{\Theta}, x) \) than they would be under \( \pi \), although

\(^8\)We give a formal definition of the Intuitive Criterion in Subsection 6.3.
this is not necessarily the case for other types in $\tilde{\Theta}$. Note that this refinement, which is stronger than robust neologism proofness, is also stronger than the Intuitive Criterion, whereas robust neologism proofness is neither stronger nor weaker than the Intuitive Criterion in general. We focus on robust neologism proofness because the requirement that all types in $\tilde{\Theta}$ strictly prefer $(x, y)$ to their outcome in $\pi$ for all $y \in BR(\tilde{\Theta}, x)$ may make a sender type’s argument that their type belongs to $\tilde{\Theta}$ simpler to understand for the receiver, and the relative permissiveness of robust neologism proofness makes its strong refinement predictions all the more striking.

We conclude this subsection by showing that the correspondence mapping games to outcome distributions induced by robust neologism proof strategy profiles is upper hemicontinuous. Fix some game with sender type space $\Theta$, receiver prior $\mu$, sender action space $X$, sender message space $M$, receiver action space $Y$, continuous sender payoff function $u : \Theta \times X \times Y \to \mathbb{R}$, and continuous receiver payoff function $v : \Theta \times X \times Y \to \mathbb{R}$. Consider sequences of games indexed by $k \in \mathbb{N}$, where the $k$-th game has the same sender type space $\Theta$ and message space $M$, but has a possibly different receiver prior $\mu_k$, sender action space $X_k$, receiver action space $Y_k$, sender payoff function $u_k : \Theta \times X_k \times Y_k \to \mathbb{R}$, and receiver payoff function $v_k : \Theta \times X_k \times Y_k \to \mathbb{R}$.

We assume that there is some larger metric space that contains the $X_k$ and $X$ as subspaces, and similarly that there is there is some larger metric space containing the $Y_k$ and $Y$. Suppose that the original game is the “limit game” of this sequence of games in that (1) The $\mu_k$ converge to $\mu$, (2) The $X_k$ and $Y_k$ converge to $X$ and $Y$ respectively according to the Hausdorff metric, and (3) The $u_k$ and $v_k$ converge to $u$ and $v$ respectively in the sense that, for all $\varepsilon > 0$, there exists some $K$ and $\delta > 0$ such that $|u_k(\theta, x', y') - u(\theta, x, y)| < \varepsilon$ and $|v_k(\theta, x', y') - v(\theta, x, y)| < \varepsilon$ whenever $|x' - x| < \delta$, $|y' - y| < \delta$, and $k > K$.

**Theorem 1.** Suppose that $p_k \in \Delta(\Theta \times X \times Y)$ is an outcome distribution induced by a robust neologism proof strategy profile in the $k$-th game, and suppose that $\lim_{k \to \infty} p_k = p$ for some $p \in \Delta(\Theta \times X \times Y)$. Then $p$ is an outcome distribution induced by a robust
neologism proof strategy profile in the limit game.

Intuitively, the proof, which is in A.1, shows that if a sequence of outcome distributions converges to an outcome distribution with a credible robust neologism, then there would be credible robust neologisms for some of the pre-limit outcome distributions.

3.3 Application to Example 1

We illustrate robust neologism proofness by applying it to Example 1 from Section 1. In a separating equilibrium, both sender types achieve their highest possible payoff, so robust neologism proofness has no bite. However, in a pooling equilibrium both sender types obtain a payoff of 1. Because the payoff of type $\theta_1$ from $(x, y_1)$ is 2, the payoff of type $\theta_2$ from $(x, y_1)$ is $-2$, and the receiver’s strict best response is $y_1$ when the sender’s type is $\theta_1$, it follows that this equilibrium has a credible robust neologism and so is eliminated by robust neologism proofness.9

4 Existence

Although robust neologism proof equilibria do not always exist, we establish their existence in two broad settings. First, in “monotonic” signaling games, there is always a robust neologism proof equilibrium. Second, a robust neologism proof equilibrium exists when the sender’s action space includes the ability to give transfers to the receiver.

4.1 Monotonic Games

As in Cho and Sobel [1990], we say that a signaling game is monotonic if, for every $x$, all the sender types share the same induced preference over the mixed actions of the receiver that are best responses to $x$ for some belief.

9Note that the same argument, but with the sender types switched, shows that there is a credible robust neologism corresponding to type $\theta_2$. 

11
**Definition 3.** A signaling game is **monotonic** if for all $\theta, \theta' \in \Theta$, $x \in X$, and $\beta, \beta' \in MBR(\Theta, x)$, $u(\theta, x, \beta) \geq u(\theta, x, \beta') \Leftrightarrow u(\theta', x, \beta) \geq u(\theta', x, \beta')$.

The following conditions ensure a signaling game is monotonic: (1) For all $x$, either all sender types prefer a higher level of $y$ or a lower level of $y$, and (2) $v(\theta, x, y)$ is strictly quasiconcave in the receiver’s action $y$ for all $(\theta, x)$.

**Theorem 2.** In monotonic signaling games, a robust neologism proof PBE exists.

Theorem 2 follows from combining the following lemma, which establishes that every D1 equilibrium\(^{10}\) is robust neologism proof in monotonic signaling games, with the fact that a D1 equilibrium exists in every signaling game.\(^{11}\)

**Lemma 1.** In monotonic signaling games, every D1 equilibrium is robust neologism proof.

We discuss the relationship between robust neologism proofness and D1 in greater detail in Subsection 6.3. We note that Lemma 1 says that, in the class of monotonic signaling games, D1 is a refinement of robust neologism proofness. The intuition behind Lemma 1 is that if an equilibrium in a monotonic signaling game has a credible robust neologism corresponding to sender action $x$ and the non-empty subset of sender types $\tilde{\Theta}$, then D1 requires that the receiver’s beliefs after $x$ be supported in $\tilde{\Theta}$.\(^{12}\) This means that the receiver responds to $x$ with some $\beta \in MBR(\tilde{\Theta}, x)$, so playing $x$ is a strictly profitable deviation for types $\theta \in \tilde{\Theta}$, contradicting equilibrium. The proof is given in A.2.1.

---

\(^{10}\)We give a formal definition of D1 equilibrium, which is a refinement of PBE, in Subsection 6.3.

\(^{11}\)D1 equilibria exist in all signaling games with finite action spaces, but not necessarily in signaling games with infinite action spaces. However, an argument similar to the proof of Theorem 1 establishes an analogous upper hemicontinuity result for D1, which along with Theorem 2 of Manelli [1996] implies the existence of D1 equilibria in all signaling games with large message spaces.

\(^{12}\)This follows from combining the fact that the preference ordering over $MBR(\Theta, x)$ is the same for all sender types with the fact that, in such a credible robust neologism, types $\theta \in \Theta$ strictly prefer $(x, y)$ to their equilibrium outcome while types $\theta' \notin \Theta$ strictly prefer their equilibrium outcome, for all $y \in BR(\tilde{\Theta}, x)$. 

12
4.2 Transfers

We show that in any signaling game where, in addition to the action $x$, the sender can give a transfer to the receiver, a robust neologism proof PBE exists. Formally, we assume that the sender’s action space is $X \times T$, where $T = \mathbb{R}_+$ and $t \in T$ represents a quasilinear transfer from the sender to the receiver. The payoffs to the sender and receiver when the sender’s type is $\theta$ and the action profile is $(x,t,y)$ are $U(\theta,x,t,y) \equiv u(\theta,x,y) - t$ and $V(\theta,x,t,y) \equiv v(\theta,x,y) + t$, respectively.\(^{13}\)

**Theorem 3.** Every signaling game with transfers has a robust neologism proof PBE.

We first establish this result for the special case where the action spaces $X$ and $Y$ are both finite. The result for general compact $X$ and $Y$ then follows because any such game has a sequence of “approximation” games with the same utility functions but action spaces that are finite subsets of the true $X$ and $Y$ which converge to the original game in the sense discussed in Subsection 3.2. Theorem 2 of Manelli [1996] ensures that for any such sequence of games and PBE outcomes in said games, there is a corresponding subsequence of PBE outcomes which converges to a PBE outcome in the original game. Theorem 1 in Subsection 3.2 guarantees that this limit outcome is robust neologism proof when the pre-limit PBE outcomes are robust neologism proof in their respective games.

**Lemma 2.** If $X$ and $Y$ are finite, there is a PBE $\pi^*$ that is robust neologism proof.

The proof constructs such a $\pi^*$. It partially mimics the usual proof of the existence of trembling-hand perfect equilibria (Selten [1975]) in that it modifies the game so that players are forced to make various trembles, and then computes a limit of the resulting equilibria as the trembles become vanishingly small. However, to ensure that the resulting limit profile is not just a PBE but also satisfies robust neologism proofness, we modify the sequence of games so that, for each sender action and non-empty subset $\theta$, we ensure $t > 2 \max_{(\theta,x,y)} |u(\theta,x,y)|$, so we could equivalently think of $T$ as being a sufficiently large, but compact interval subset of $\mathbb{R}_+$.\(^{13}\)
of sender types $\Theta$, there is a message that types $\theta \in \Theta$ belonging to the subset can use that prevents the receiver from playing any actions outside of $BR(\Theta, x)$ when $x$ is played.

In equilibrium in the actual game, the different sender types can always mimic each other. Thus, we ensure that a sender type $\theta \in \Theta$ uses such a message in the artificial games only when no type $\theta' \notin \Theta$ outside of $\Theta$ would prefer the corresponding outcome to the prevailing profile. We do so by modifying the payoff functions of the sender types so that there are costs from using such messages that depend on the incentives of the other sender types.

The construction of $\pi^*$ is in A.2.2. While we modify the sender type payoff functions so that they depend on the incentives of the other sender types, we are careful to do so in a way that does not lead to the various objects being recursively defined in terms of each other. The following two lemmas complete the proof of Lemma 2.

**Lemma 3.** There is no sender action $x$, transfer $t$, and non-empty subset of sender types $\tilde{\Theta}$, such that

1. $\min_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y) - t > U(\theta, \pi^*)$ for some $\theta \in \tilde{\Theta}$, and

2. $\max_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y) - t < U(\theta', \pi^*)$ for all $\theta' \notin \tilde{\Theta}$.

**Lemma 4.** $\pi^*$ is a PBE.

The proof of Lemma 3 is in A.2.3 while the proof of Lemma 4 is in A.2.4. The greatest difficulty is ensuring that $\pi^*$ is a PBE, and in particular ensuring that the various sender types are playing optimally. The reason for this is that modifying the sender types’ payoff functions in the pre-limit games to depend on the payoffs the other sender types obtain may lead to some of the sender types’ limit strategies not being optimal given their true payoff function and the receiver’s strategy $\pi^*_R$. This problem emerges for some type $\theta$ when they have a strictly profitable deviation to some action and receiver response that is induced by a message corresponding to some
subset of sender types $\tilde{\Theta}$ that contains $\theta$, but there are other types outside of $\tilde{\Theta}$ that in the pre-limit games are near indifferent between their equilibrium payoffs and this outcome. However, when quasilinear transfers are present, such indifferences do not cause a problem, as a type $\theta$ can break any indifferences of other sender types while negligibly affecting their own payoff.

Before proceeding, we note that quasilinear transfers are sufficient for the existence of robust neologism proof equilibria, but they are more than necessary even for the proof technique in this paper. In fact, the proof extends if the sender’s utility function $U(\theta, x, t, y)$ is strictly decreasing in the transfer $t$ and satisfies $\lim_{t \to \infty} \max_{(\theta,x,y)} U(\theta, x, t, y) = -\infty$ and the transfer does not affect the incentives of the receiver. This would be the case, for instance, if the receiver’s utility function were of the form $V(\theta, x, t, y) = f_1(t)v(\theta, x, y) + f_2(\theta, x, t)$ for some $f_1 : \mathbb{R}_{+} \to \mathbb{R}_{++}$ and $f_2 : \Theta \times X \times \mathbb{R}_{+} \to \mathbb{R}$. This nests the formulation with quasilinear transfers as the special case with $f_1(t) = 1$ and $f_2(\theta, x, t) = t$ for all $(\theta, x, t) \in \Theta \times X \times \mathbb{R}_{+}$.

### 4.3 Non-Existence

Here is a signaling game, based on Example 3 of Farrell [1993], in which there is no robust neologism proof equilibrium. However, when the game is altered so that the sender can give transfers to the receiver, there are robust neologism proof equilibria, as must be the case by Theorem 3.

**Example 2.** The sender type space is $\Theta = \{\theta_1, \theta_2\}$ and the receiver’s prior is $\mu(\theta_1) = \mu(\theta_2) = 1/2$. The sender’s action space is singleton, that is $X = \{x\}$, and the receiver’s action space is $Y = \{y_1, y_2, y_3\}$. The payoffs to the sender types and receiver are given in Table 2.

\[^{14}\text{Less restrictive conditions, like } \lim_{t \to \infty} \max_{(\theta,x,y)} U(\theta, x, t, y) < \min_{(\theta,x,y)} U(\theta, x, 0, y), \text{ also suffice.}\]
It is shown in OA.1 that the only equilibrium outcome is the one in which the receiver always plays \( y_3 \) with probability 1 regardless of the sender type. Intuitively, the receiver takes action \( y_2 \), which is the worst receiver action for both sender types, when the receiver ascribes a large probability to the sender type being \( \theta_2 \). This drives sender type \( \theta_2 \) to mimic sender type \( \theta_1 \), so only pooling is possible in equilibrium.

However, no such equilibrium is robust neologism proof. The reason for this is that type \( \theta_2 \) obtains their highest payoff from \( y_3 \), but type \( \theta_1 \) would obtain a strictly higher payoff from \( y_1 \), which is precisely the receiver’s optimal action when the sender type is \( \theta_1 \). Hence, there is a credible robust neologism for type \( \theta_1 \).

Now, consider the altered game that allows transfers. There is a separating equilibrium that is robust neologism proof\(^{15} \) where type \( \theta_1 \) plays \( t = 1 \) and the receiver responds with \( y_1 \), and type \( \theta_2 \) plays \( t = 0 \) and the receiver responds with \( y_2 \). To see that this equilibrium has no credible robust neologisms, first note that no sender type would obtain a strictly higher payoff, regardless of the transfer level \( t \), were the receiver to respond with \( y_2 \). Thus, there is no credible robust neologism corresponding to a \( \tilde{\Theta} \) that contains \( \theta_2 \). Second, note that type \( \theta_1 \) obtains a strictly higher payoff when the receiver plays \( y_1 \) only if they played some transfer \( t' < 1 \); however, type \( \theta_2 \) would also prefer such an outcome to their equilibrium outcome, so there is no credible robust neologism corresponding to \( \tilde{\Theta} = \{ \theta_1 \} \). \( \square \)

\(^{15}\)In contrast, there are no neologism-proof equilibria in this game.

\[\begin{array}{ccc}
\theta_1 & y_1 & y_2 & y_3 \\
x & 2, 2 & 0, -2 & 1, 1 \\
\theta_2 & y_1 & y_2 & y_3 \\
x & 1, -2 & 0, 2 & 2, 1 \\
\end{array}\]

Table 2: The payoffs for Example 2.
5 Monotone-Concave-Supermodular Signaling Games with Transfers

We analyze robust neologism proofness in the setting of “monotone-concave-supermodular” signaling games with transfers. These games are a special class of both the monotonic games of Subsection 4.1 and the signaling games with transfers of Subsection 4.2. Here we assume that the sender types are ordered according to $\theta_1 < \ldots < \theta_N$ and that the receiver’s action space $Y$ is fully ordered. The formal definition of monotone-concave-supermodular signaling games with transfers is as follows.

Definition 4. A signaling game with transfers is monotone-concave-supermodular (MCS) if

1. $u(\theta, x, y)$ is strictly increasing in $y$ for all $(\theta, x)$,

2. $u(\theta, x, y)$ has strictly increasing differences in $(\theta, y)$ for all $x$, and

3. $y^*(\tilde{\mu}, x) \equiv BR(\tilde{\mu}, x)$ is singleton and strictly increasing in $\tilde{\mu}$ according to the FOSD partial ordering of $\Delta(\Theta)$ for all $x$.$^{16}$

The first criterion states that all sender types strictly prefer a higher level of $y$, regardless of $x$. The second criterion imposes supermodularity on the dependence of the sender’s payoffs on their type and the receiver’s action: The difference in sender utility from a higher receiver action to a lower receiver action, holding fixed the sender’s action, must be strictly increasing in the sender’s type. Finally, criterion 3 requires the receiver’s best response to be strictly increasing in their posterior belief about the sender’s type for any $x$. A sufficient condition for criterion 3 is that $v(\theta, x, y)$ be strictly quasiconcave in $y$ and, modulo minor differentiability assumptions, have strictly increasing differences in $(\theta, y)$ for all $x$.

Our main result in this section is that the robust neologism proof equilibria coincide with the “sender-optimal” separating equilibria in MCS games with transfers. In

$^{16}$For notational ease, we adapt $y^*(\tilde{\mu}, x) \equiv BR(\tilde{\mu}, x)$ and $y^*(\theta, x) \equiv BR(\theta, x)$ in this section.
this setting, an equilibrium \( \pi = (\pi_{\theta_1}, ..., \pi_{\theta_N}, \pi_R) \) is separating if there is a family of measurable subsets \( A_\theta \subseteq X \times T \times M \), one for each \( \theta \in \Theta \), such that (1) \( \pi_\theta(A_\theta) = 1 \) for all \( \theta \in \Theta \), and (2) \( A_\theta \cap A_{\theta'} = \emptyset \) if \( \theta \neq \theta' \). Condition (1) says that when sender type \( \theta \) plays according to \( \pi_\theta \), the resulting \((x, t, m)\) comes from \( A_\theta \) with probability 1, while Condition (2) says that the \( A_\theta \) subsets are disjoint for distinct sender types. We now define sender-optimal separating equilibria and show that they exist in all MCS games with transfers. Denote the set of separating equilibria by \( \Pi_s \). A sender-optimal separating equilibrium is any \( \pi_s \in \Pi_s \) that gives every sender type a higher payoff than they obtain in any other separating equilibrium.

**Definition 5.** Separating equilibrium \( \pi_s \in \Pi_s \) is sender-optimal if \( U(\theta, \pi_s) \geq U(\theta, \pi_s') \) for all \( \pi_s, \pi_s' \in \Pi_s \), \( \theta \in \Theta \).

In general signaling games, separating equilibria are not guaranteed to exist; however, in the setting of MCS signaling games with transfers, sender-optimal separating equilibria always exist and have a simple characterization. Consider the profile of sender type payoffs \((U_{\theta_1}, ..., U_{\theta_N})\) defined inductively by

\[
U_{\theta_n} = \max_{(x,t)} u(\theta_n, x, y^*(\theta_n, x)) - t \tag{1}
\]

subject to \( u(\theta_m, x, y^*(\theta_n, x)) - t \leq U_{\theta_m} \) for all \( m < n \).

For each \( \theta \in \Theta \), let \( \overline{A}_\theta \subseteq X \times T \times M \) consist of the \((x, t, m)\) where \((x, t)\) solves (1) for \( \theta \).\(^{17}\)

**Lemma 5.** Sender-optimal separating equilibria exist and are precisely the equilibria \( \pi \) in which \( \pi_\theta(\overline{A}_\theta) = 1 \) for all \( \theta \in \Theta \). Moreover, every sender-optimal separating equilibrium is outcome equivalent to a PBE.

Consider a separating equilibrium \( \pi_s \in \Pi_s \) with corresponding \( A_{\theta_1}, ..., A_{\theta_n} \subseteq X \times T \times M \) and sender type payoffs \((U_{\theta_1}, ..., U_{\theta_N})\). For each \( \theta \in \Theta \), the receiver responds

\(^{17}\)Formally, \( \overline{A}_\theta = \{(x, t, m) \in X \times T \times M\mid u(\theta_m, x, y^*(\theta_n, x)) - t \leq U_{\theta_m} \text{ for all } m \leq n \text{ with equality for } m = n\} \).
with \( y^*(\theta, x) \) for \( \pi_\theta \) almost all \((x, t, m) \in A_\theta\). Hence, \( U_{\theta_1} \leq \overline{U}_{\theta_1} \). Moreover, since 
\[
 u(\theta_1, x, y^*(\theta_2, x)) - t \leq U_{\theta_1} \leq \overline{U}_{\theta_1}
\]
for all \((x, t, m) \in A_{\theta_2}\), it follows that \( U_{\theta_2} \leq \overline{U}_{\theta_2} \). Continuing this argument shows that \( u_\theta \leq \overline{U}_\theta \) for all \( \theta \in \Theta \). Additionally, because of the monotonicty of the sender preferences over the receiver’s action and the fact that the receiver’s best response for a given \( x \) is different for different types, the incentive constraints in (1) imply that \( A_\theta \cap \overline{A}_{\theta'} = \emptyset \) whenever \( \theta \neq \theta' \). Thus, any equilibrium in which each type \( \theta \) plays \((x, t, m) \in \overline{A}_\theta \) with probability 1 is necessarily a sender-optimal separating equilibrium, so it suffices to show that an arbitrary strategy profile \( \pi \) in which \( \pi_\theta(A_\theta) = 1 \) and \( \pi_R(y^*(\theta, x)|(x, t, m)) = 1 \) for \( \pi_\theta \) almost all \((x, t, m) \in \overline{A}_\theta \) is outcome equivalent to a PBE.

We construct such a PBE by having the receiver’s belief concentrate on the lowest sender type \( \theta_1 \) after any \((x, t, m) \not\in \bigcup_{\theta \in \Theta} \overline{A}_\theta\).\footnote{The receiver’s belief after any \((x, t, m) \in \overline{A}_\theta \) concentrates on \( \theta \).} To get a sense of why this constitutes a PBE, note that because of the incentive constraints in (1), no type \( \theta \) strictly prefers to play any \((x, t, m) \in \overline{A}_{\theta'} \) for \( \theta' > \theta \). If instead some type \( \theta \) strictly preferred to play some \((x, t, m) \in \overline{A}_{\theta'} \) for some \( \theta' < \theta \), then they would also strictly prefer the outcome corresponding to \((x, t, m) \) when the receiver plays a best response to type \( \theta \) rather than type \( \theta' \). Because of supermodularity, the transfer type \( \theta \) gives can be increased so that \( \theta \) still obtains a strictly higher payoff than \( \overline{U}_\theta \), but every smaller type \( \tilde{\theta} < \theta \) would obtain a smaller payoff than \( \overline{U}_{\tilde{\theta}} \). However, this contradicts \( \overline{U}_\theta \) being the maximum payoff of type \( \theta \) in (1). Likewise, no type \( \theta \) can strictly prefer to take an off-path action for similar reasons. The proof of Lemma 5 is given in A.3.1.

We now establish that the robust neologism proof equilibria are the sender-optimal separating equilibria.

**Theorem 4.** In any MCS game with transfers, the robust neologism proof equilibria are the sender-optimal separating equilibria.

That there are no robust neologism proof equilibria other than the sender-optimal separating equilibria is a consequence of the following two lemmas.
Lemma 6. In every robust neologism proof equilibrium $\pi$, $U(\theta, \pi) \geq U_\theta$ for all $\theta \in \Theta$.

Lemma 7. Every robust neologism proof equilibrium must be separating.

To see the intuition for Lemma 6, consider an equilibrium where $U(\theta, \pi) < U_\theta$ for some $\theta \in \Theta$, and let $\theta$ be the lowest sender type that obtains a strictly lower payoff than $U_\theta$. Any $(x, t) \in A_\theta$ in the sender-optimal separating equilibrium, with a slightly increased transfer to the receiver, is either a credible robust neologism for type $\theta$, or the same $x$ with a higher transfer is a credible robust neologism for a higher type.

Intuitively, Lemma 7 follows from the fact that an equilibrium that is not separating involves pooling for some types. While all pooling types would prefer the receiver to respond as if they believed the type to be the highest among them, the highest pooling type would gain the most from this by supermodularity. With the ability to give the receiver a higher transfer, either the highest types involved in pooling can distinguish themselves with a credible robust neologism or there is an even higher type that was not originally pooling for whom there is a credible robust neologism. The proof of Lemma 6 is given in A.3.2, and the proof of Lemma 7 is given in A.3.3.

The existence results of Section 4 show that there must be robust neologism proof equilibria in MCS games with transfers. Because each sender-optimal separating equilibrium gives the sender types the same payoff as every other sender-optimal separating equilibrium, and there are no robust neologism proof equilibria other than the sender-optimal separating equilibria, it follows that each sender-optimal equilibrium is robust neologism proof. This establishes Theorem 4.

We now present an example of an MCS game with transfers in which the incentive scheme an informed firm offers to a worker may signal the firm’s knowledge about the potential profitability of their project.\textsuperscript{19}

Example 3. Consider the interaction of an informed firm with an uninformed worker they employ. Suppose that the firm knows $\theta \in \{1, 2, 3\}$, a parameter governing the

\textsuperscript{19}Beaudry [1994] analyzes a similar setting when the firm has two possible types, and notes there that the Intuitive Criterion selects the firm-optimal separating equilibrium. However, unlike robust neologism proofness, the Intuitive Criterion does not in general make this selection when the firm has more than two types.
profitability of the effort $e$ that the worker exerts. The worker does not know $\theta$, but has a prior belief that $\theta = 1$ with probability $1/6$, $\theta = 2$ with probability $1/2$, and $\theta = 3$ with probability $1/3$. The firm can offer linear incentive pay, consisting of a base-rate transfer $t \in \mathbb{R}_+$ and a bonus $b \in [0, 1]$ per unit of profitability to the worker. The firm’s payoff when their type is $\theta$, the worker exerts effort level $e$, and the contract between them is $(b, t)$ is $U(\theta, b, t, e) = (1 - b)\theta e - t$, while the worker’s payoff is $V(\theta, b, t, e) = b\theta e - e^2/2 + t$. To see that this game is (essentially) MCS,\(^{20}\) note that, when $b < 1$, $(1 - b)\theta e$ is strictly increasing in $e$, regardless of $b$ and $\theta$, and has strictly increasing differences in $(\theta, e)$, so the first and second criteria hold. Likewise, the worker’s best response level of effort to any belief $\tilde{\mu}$ about the firm type $\theta$ and bonus level $b$ is $e^*(\tilde{\mu}, b) = bE_{\tilde{\mu}}[\theta]$, which is singleton and strictly increasing whenever $\tilde{\mu}$ strictly increases in the sense of FOSD when $b > 0$.

It can be shown that the firm-optimal separating outcome has all firm types giving a bonus of $b^* = 1/2$, while firm type 1 transfers $t^*(1) = 0$ (to which the worker responds with effort $1/2$), firm type 2 transfers $t^*(2) = 1/4$ (to which the worker responds with effort $1$), and firm type 3 transfers $t^*(3) = 3/4$ (to which the worker responds with effort $3/2$); here firm type 1 obtains $1/4$, firm type 2 obtains $3/4$, and firm type 3 obtains $3/2$. By Theorem 4, this is the unique robust neologism proof equilibrium outcome. \(\square\)

Theorem 4 establishes the equivalence of the robust neologism proof equilibria and the sender-optimal separating equilibria in a setting where transfers impact the utility of both sender and receiver quasilinearly. However, this result can be generalized to other settings with transfers. In particular, it can be shown that sender-optimal separating equilibria exist and coincide with the robust neologism proof equilibria when the following four conditions are met: (1) The sender’s utility function $U(\theta, x, t, y)$ is strictly increasing in $y$ for all $(\theta, x, t)$, (2) The sender’s utility function $U(\theta, x, t, y)$ is \(^{20}\)Technically, the possibility of $b = 0$ or $b = 1$ prevents this game from being MCS in the sense of Definition 4. However, no firm type would ever play $b = 0$ or $b = 1$ in equilibrium, so the general results of this section continue to apply.
single-crossing in \((\theta, t, y)\) in that, for all \(x, \theta > \theta', t > t',\) and \(y > y',\) \(U(\theta', x, t, y) \geq U(\theta', x, t', y')\) implies \(U(\theta, x, t, y) > U(\theta, x, t', y'),\) (3) \(\lim_{t \to \infty} \max_{(\theta, x, y)} U(\theta, x, t, y) = -\infty,\) and (4) The receiver’s best response \(y^*(\tilde{\mu}, x, t) \equiv BR(\tilde{\mu}, x, t)\) is singleton and strictly increasing in \(\tilde{\mu}\) according to the FOSD partial ordering of \(\Delta(\Theta)\) for all \((x, t)\).

Since D1 is a refinement of robust neologism proofness in such a setting, it also follows that D1 selects the sender-optimal separating equilibria. Propositions 4.2 and 4.5 of Cho and Sobel [1990] establish a similar result in a related class of signaling games that satisfy the “monotonicity” conditions 1 and 3 in Definition 4, but also satisfy a single-crossing condition regarding how the sender’s type, receiver’s action \(y,\) and the components of the sender’s (possibly multidimensional) action \(x\) interact in the payoff functions of the sender types. Ramey [1996] notes that D1 continues to select the sender-optimal separating equilibrium under a more relaxed single-crossing condition that applies to just one component of the sender’s action.

6 Relationship with Other Refinements

We compare robust neologism proofness with various other equilibrium refinements. We first discuss differences with neologism-proofness, including the fact that the existence results for robust neologism proof equilibria presented in Section 4 do not extend to this more restrictive refinement.

6.1 Neologism-Proofness

As originally formulated in Farrell [1993], neologism-proofness is a refinement specific to cheap-talk games. However, as noted in Maskin and Tirole [1992], neologism-proofness can be adapted for costly signaling games as well. For ease of comparison, we define a version of neologism-proofness for general signaling games.

Fix an equilibrium \(\pi.\) For any \((x, y),\) let \(WH(x, y, \pi)\) be the set of types that would obtain a weakly higher payoff from \((x, y)\) than in \(\pi,\) and let \(H(x, y, \pi)\) be the
set of types that would obtain a strictly higher payoff from \((x, y)\) than in \(\pi\). Let 
\[ M(x, y, \pi, \mu) \subseteq \Delta(\Theta) \]
be the set of beliefs given by

\[
M(x, y, \pi, \mu) = \left\{ \tilde{\mu} \in \Delta(WH(x, y, \pi)) \left| \frac{\tilde{\mu}(\theta')}{\tilde{\mu}(\theta)} \leq \frac{\mu(\theta')}{\mu(\theta)} \text{ for all } \theta' \in \Theta, \theta \in H(x, y, \pi) \right. \right\}.
\]

That is, \(M(x, y, \pi, \mu)\) is precisely the set of beliefs that both put support only on
the sender types who weakly gain by a deviation to \((x, y)\) from the equilibrium \(\pi\), and
weakly increases the relative weight on each type that strictly gains by such a deviation.

We now present the definition of credible neologisms and the definition of neologism-
proofness.

**Definition 6.** If there exists some \((x, y)\) such that \(H(x, y, \pi) \neq \emptyset\) and some \(\tilde{\mu} \in M(x, y, \pi, \mu)\) such that \(y \in BR(\tilde{\mu}, x)\), then \(\pi\) has a **credible neologism**.

**Definition 7.** An equilibrium is **neologism-proof** if it has no credible neologisms.

Neologism-proofness is a stronger refinement than robust neologism proofness since
all credible robust neologisms correspond to credible neologisms. A stark feature of
neologism-proofness is that in a credible neologism, the receiver action \(y\) need only be
a best response to a particular posterior belief supported on the type set corresponding
to the neologism, which is essentially the receiver’s prior updated to the sender types
that gain by deviating to the neologism from the equilibrium, with some flexibility to
allow weakening the relative weight on the types that do not strictly gain. In contrast,
in a robust neologism corresponding to some sender action \(x\) and non-empty subset of
sender types \(\tilde{\Theta}\), all possible receiver best responses to every possible receiver belief with
support contained in \(\tilde{\Theta}\) are considered in determining whether the robust neologism is
credible.

Having stronger requirements concerning receiver beliefs than neologism-proofness
seems desirable since, even when the receiver puts full support only on those types
that gain by a proposed deviation from equilibrium, it is not clear why the receiver’s
belief should update in the way neologism-proofness demands. Indeed, this requires
the receiver to believe that every type who strictly gains from the deviation to be equally likely to have proposed it, but this seems unlikely to be the case in general when some of these types may prefer an alternative credible deviation. Additionally, the relatively lax requirement that neologism-proofness has concerning receiver beliefs also causes some of the problems with the existence of neologism-proof equilibria that do not plague robust neologism proofness.

We have seen an example in which robust neologism proof equilibria do not exist, so a fortiori no neologism-proof equilibria exist either. Section 4 presents two settings (monotonic signaling games and signaling games with transfers) in which robust neologism proof equilibria are guaranteed to exist. These existence results do not carry over to neologism-proofness. To illustrate this, we show that the sender-optimal separating equilibrium outcome from Example 3 in Section 5 is not neologism-proof. The reason is that playing \((b, t) = (1/2, 0)\) and advocating the receiver take action \(e = 13/12\) is a credible neologism for all types. This is because the payoff of type \(\theta\) from the neologism would be \(13\theta/24\), so the payoffs of all firm types strictly exceed their equilibrium payoff. Additionally, the worker’s best response to \(b = 1/2\) under their prior over the firm’s type is \(e = 13/12\).

Before we proceed, we note that in some previous results in the literature where neologism-proofness has been used as a refinement, robust neologism proofness could instead be used to obtain the same, and sometimes nicer, results. An example of this is Maskin and Tirole [1992], which analyzes the interaction of an informed principal with an agent when the principal’s type affects the agent’s payoff from the allocation the principal implements.\(^{21}\) The paper analyzes this interaction using the framework of an extended signaling game. It mostly focuses on the set of PBE, but also analyzes the resulting set of equilibria under the Intuitive Criterion and neologism-proofness. One of the paper’s results, Proposition 7, shows that under some natural assumptions,\(^{21}\)

\(^{21}\)As observed in Myerson [1983], many informed-principal settings are likely to have a high degree of flexible communication between the principal (sender) and agent (receiver). This makes them a natural topic of interest for robust neologism proofness and other communication-based equilibrium refinements.
the unique equilibrium outcome satisfying the Intuitive Criterion is what is termed the “Rothschild-Stiglitz-Wilson (RSW) allocation,” which is similar to the “sender-optimal” separating outcomes analyzed in Section 5 of this paper.

Additionally, Proposition 7 shows that the RSW allocation is the unique equilibrium outcome under neologism-proofness precisely when it is Pareto-efficient for the sender types among all allocations that respect certain individual rationality conditions. However, when this is not the case, neologism-proofness rules out all equilibria. In contrast, robust neologism proofness selects the RSW allocation as the unique equilibrium outcome under the conditions that Proposition 7 identifies for the Intuitive Criterion to make this selection. Additionally, in many common classes of signaling games analyzed in the literature, neologism-proofness may rule out all equilibria. For some parameters in the Spence signaling model, neologism-proofness rules out all equilibria, including those with the least-cost separating outcome, which is the equilibrium outcome selected by robust neologism proofness.

6.2 Vulnerability to Credible Deviations and the Strong Intuitive Criterion

Eső and Schummer [2009] and Riley [2012] provide refinements that only apply to off-path sender actions. Eső and Schummer [2009] says that an equilibrium is “vulnerable to a credible deviation” if there is some off-path action $x$ and a unique subset of sender types $\tilde{\Theta}$ such that $\tilde{\Theta} = \{\theta \in \Theta | \min_{y \in BR(\tilde{\Theta},x)} u(\theta,x,y) > u(\theta,\pi)\}$. The implicit requirement that $\min_{y \in BR(\tilde{\Theta},x)} u(\theta',x,y) \leq u(\theta',\pi)$ for all $\theta' \notin \tilde{\Theta}$ is substantially weaker than the corresponding condition for a credible robust neologism, $\max_{y \in BR(\tilde{\Theta},x)} u(\theta',x,y) < u(\theta',\pi)$ for all $\theta' \notin \tilde{\Theta}$, and it is not clear in general why types outside of $\tilde{\Theta}$ should be deterred from attempting to convince the receiver their type belongs to $\tilde{\Theta}$ when only a worst-case receiver best response to $\tilde{\Theta}$ and $x$ would (weakly) deter them. Additionally, this weaker requirement for credibility leads to failure of upper hemicontinuity in the correspondence mapping games into equilibrium.
outcomes that are not vulnerable to a credible deviation. It also precludes the existence of equilibria that are not vulnerable to credible deviations in settings of interest, such as signaling games with transfers, where robust neologism proof equilibria exist.

Riley [2012]'s Strong Intuitive Criterion\textsuperscript{22} eliminates any equilibrium where there is a credible robust neologism corresponding to some off-path action and singleton subset of sender types, i.e. $\tilde{\Theta} = \{\theta\}$ for some $\theta \in \Theta$. Thus, it eliminates any equilibrium $\pi$ where there is some off-path action $x$ and sender type $\theta$ such that (a) $\min_{y \in BR(\theta, x)} u(\theta, x, y) > u(\theta, \pi)$, and (b) $\max_{y \in BR(\theta, x)} u(\theta', x, y) < u(\theta', \pi)$ for all $\theta' \neq \theta$. This refinement is weaker than robust neologism proofness,\textsuperscript{23} so equilibria satisfying the Strong Intuitive Criterion necessarily exist whenever there is a robust neologism proof equilibrium, although equilibria satisfying the Strong Intuitive Criterion still do not exist in all games.

Unlike robust neologism-proofness, the correspondence mapping games to equilibrium outcomes satisfying the Strong Intuitive Criterion is not upper hemicontinuous, because types that impact the payoffs of the sender and receiver identically in some game may generate different receiver best responses when the game is slightly perturbed.\textsuperscript{24} Relatedly, there are examples (one of which is given in OA.3) where the Strong Intuitive Criterion preserves unreasonable equilibria in which there is a credible robust neologism corresponding to some non-empty subset of sender types $\tilde{\Theta}$, but where some of the types in $\tilde{\Theta}$ generate different best responses for the receiver.\textsuperscript{25}

Because the Eső and Schummer [2009] refinement and the Strong Intuitive Criterion

\textsuperscript{22}The author wishes to thank Zhuoran Lu for informing him of the Strong Intuitive Criterion. Riley [2012] only defines the Strong Intuitive Criterion for signaling games in which the receiver has a unique best response for every sender action and type; the definition given here extends the Strong Intuitive Criterion to general signaling games.

\textsuperscript{23}Thus, despite the name, the Strong Intuitive Criterion is not a stronger refinement than the Intuitive Criterion in general games.

\textsuperscript{24}This implicitly assumes that types that impact the payoffs of the sender and receiver identically are identified as the same type for the purpose of checking whether an equilibrium satisfies the Strong Intuitive Criterion. Riley [2012] does not discuss how to handle this situation.

\textsuperscript{25}It is worth noting however that the Strong Intuitive Criterion makes the same selection of the sender-optimal separating equilibria in the class of “monotone-concave-supermodular” signaling games with transfers.
only apply to off-path sender actions, neither refinement has any bite in cheap-talk games, since any equilibrium outcome in a cheap-talk game can be obtained with an equilibrium in which all messages are used on-path. Robust neologism proofness does not impose this restriction since whether an action is on-path or off-path does not seem important for the credibility of communication.

6.3 The Intuitive Criterion and D1

Here, we further compare robust neologism proofness with the Intuitive Criterion and D1. For ease of reference, we formulate these two refinements in the language of this paper in this subsection.

**Definition 8.** \( \pi \) fails the **Intuitive Criterion** if there exists some sender action \( x \in X \), non-empty subset of sender types \( \tilde{\Theta} \subseteq \Theta \), and sender type \( \theta \in \tilde{\Theta} \) such that

1'. \( \min_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y) > u(\theta, \pi) \), and

2'. \( \max_{y \in BR(\Theta, x)} u(\theta', x, y) < u(\theta', \pi) \) for all \( \theta' \notin \tilde{\Theta} \).

Condition 2' says that sender action \( x \) is equilibrium dominated for types outside of \( \tilde{\Theta} \). Condition 1' (which is formally the same as the first condition in the definition of the alternative refinement discussed in Subsection 3.2) says that type \( \theta \) obtains a strictly higher payoff from playing action \( x \) for any receiver best response to \( x \) and a belief over sender types whose support is contained in \( \tilde{\Theta} \) than under \( \pi \). The motivation for the Intuitive Criterion is that, when conditions 1' and 2' hold, the type \( \theta \) sender can, instead of following the equilibrium, play \( x \) and convince the receiver that their type is contained in \( \tilde{\Theta} \) since no type for whom \( x \) is equilibrium dominated prefers \( x \) under any receiver best-response to the prevailing equilibrium. As discussed in Subsection 3.2, robust neologism proofness accounts for the possibility of the sender convincing the receiver that they belong to other subsets of the sender types than just the set of all types for whom a given action is not equilibrium dominated. This corresponds to the differences in Condition 2 of Definition 6 and Condition 2' above.
To formulate the D1 refinement, we first establish some relevant notation. For a given sender type \( \theta \), sender action \( x \), and strategy profile \( \pi \), let \( D_\theta(x, \pi) = \{ \beta \in MBR(\Theta, x)|u(\theta, x, \beta) > u(\theta, \pi)\} \) be the set of receiver mixed best responses to \( x \) and any belief over sender types that makes type \( \theta \) strictly better off than under \( \pi \). Similarly, let \( D^0_\theta(x, \pi) = \{ \beta \in MBR(\Theta, x)|u(\theta, x, \beta) = u(\theta, \pi)\} \) be the set of receiver mixed best responses to \( x \) and any belief over sender types that makes type \( \theta \) equally well off as under \( \pi \). Let \( \Theta_{D1}(x, \pi) = \Theta \setminus \{ \theta \in \Theta|\exists \theta' \text{ s.t. } D_\theta(x, \pi) \cup D^0_\theta(x, \pi) \subseteq D_{\theta'}(x, \pi)\} \).

**Definition 9.** \( \pi \) is a **D1 equilibrium** if it is a PBE and, for every sender action \( x \), there exists some \( \beta \in MBR(\Theta_{D1}(x, \pi), x) \) such that \( u(\theta, x, \beta) \leq u(\theta, \pi) \) for all \( \theta \in \Theta \).

By Lemma 1, D1 is a refinement of robust neologism proofness in the class of monotonic signaling games. Because of this, for some games, robust neologism proofness and D1 offer different foundations for the same equilibrium selections, despite their different motivations. This is the case for the class of “monotone-concave-supermodular” signaling games with transfers analyzed in Section 5.

However, outside the class of monotonic signaling games, robust neologism proofness and D1 can behave quite differently. The following example is a game in which both robust neologism proof and D1 equilibria exist, but there is no equilibrium that is both robust neologism proof and D1. This illustrates how different the communication-based refinement of robust neologism proofness can be from traditional refinements in terms of equilibrium selection.

**Example 4.** The sender’s type space is \( \Theta = \{ \theta_1, \theta_2 \} \) and the receiver’s prior is \( \mu(\theta_1) = \mu(\theta_2) = 1/2 \). The sender’s action space is \( X = \{ x_1, x_2 \} \) and \( Y = \{ y_1, y_2, y_3 \} \). The payoffs to the sender types and receiver are given in Table 3.
We show in OA.2 that the unique D1 equilibrium outcome is the pooling equilibrium on $x_1$, while the unique robust neologism proof equilibrium outcome is the pooling equilibrium on $x_2$. (Both of these equilibrium outcomes satisfy the Intuitive Criterion.)

The reason why D1 rules out the pooling equilibrium on $x_2$ is that the receiver’s beliefs following $x_1$ must concentrate on $\theta_1$ since the payoff type $\theta_1$ obtains from any $(x_1, y)$ outcome is strictly higher than the payoff type $\theta_2$ obtains, and the payoff type $\theta_1$ obtains from any $(x_2, y)$ outcome is strictly lower than the payoff type $\theta_2$ obtains. The reason why robust neologism proofness rules out the pooling equilibrium on $x_1$ is that there is a credible robust neologism corresponding to sender action $x_1$ and type $\theta_1$, since the receiver’s strict best response to type $\theta_1$ taking action $x_1$ is $y_1$, and type $\theta_1$ obtains a strictly higher payoff from $(x_1, y_1)$ than $(x_1, y_3)$, while type $\theta_2$ obtains a strictly higher payoff from $(x_1, y_3)$.

\[\begin{array}{|c|c|c|c|} \hline \theta_1 & y_1 & y_2 & y_3 \\ \hline x_1 & 2, 2 & 1/8, -2 & 1, 1 \\ x_2 & -1/4, 2 & -5/4, -2 & 1/4, 1 \\ \hline \end{array}\]

\[\begin{array}{|c|c|c|c|} \hline \theta_2 & y_1 & y_2 & y_3 \\ \hline x_1 & 2/3, -2 & 0, 2 & 3/4, 1 \\ x_2 & 0, -2 & -1, 2 & 1/2, 1 \\ \hline \end{array}\]

Table 3: The payoffs for Example 4.

7 Discussion

Robust neologism proofness is a natural refinement for informed-principal settings and other organizational settings with close interaction between parties. Clark [2021] applies a version of robust neologism proofness to study principal-agent settings with informed principals and agents whose actions are subject to moral hazard.\footnote{The version of robust neologism proofness in Clark [2021] accounts for principals proposing mechanisms that are more complicated than the actions played by senders in signaling games.}

It would be useful to develop foundations for robust neologism proofness and similar communication-based equilibrium refinements. One possible approach would be to understand what communication-based refinements emerge from settings in which there
are costs of lying or some agents are not strategic.\footnote{Chen et al. [2008] takes this approach and provides a foundation for a NITS (no incentive to separate) condition which they use to select equilibria in the Crawford and Sobel [1982] cheap-talk setting.} A different approach would be to study learning foundations for communication-based refinements. Clark and Fudenberg [2021] takes this approach and obtains the solution concept of “justified communication equilibrium.” However, obtaining learning foundations for refinements with cutting power in cheap-talk games, such as robust neologism proofness, remains an open issue.

A Appendix

A.1 Proof of Theorem 1

**Theorem 1.** Suppose that $p_k \in \Delta(\Theta \times X \times Y)$ is an outcome distribution induced by a robust neologism proof strategy profile in the $k$-th game, and suppose that $\lim_{k \to \infty} p_k = p$ for some $p \in \Delta(\Theta \times X \times Y)$. Then $p$ is an outcome distribution induced by a robust neologism proof strategy profile in the limit game.

**Proof.** By Theorem 2 of Manelli [1996], there is a strategy profile that induces the outcome distribution $p$. We show that this strategy profile is robust neologism proof. We abuse notation and denote the expected payoff of sender type $\theta$ from a strategy profile inducing outcome distribution $p_k$ in the $k$-th game by $u_k(\theta, p_k)$ and from a strategy profile inducing the outcome distribution $p$ in the limit game by $u(\theta, p)$.\footnote{Formally, $u_k(\theta, p_k) = E_{p_k}[u_k(\hat{\theta}, x, y) | \hat{\theta} = \theta]$ and $u(\theta, p) = E_p[u(\hat{\theta}, x, y) | \hat{\theta} = \theta]$.}

Consider the non-empty subset of sender types $\tilde{\Theta}$, sender action $x \in X$, and some sequence $x_k \in X_k$ such that $\lim_{k \to \infty} x_k = x$. Because each $p_k$ corresponds to a robust neologism proof strategy profile, it follows that, for all $k$, either (1) there is some $\theta \in \tilde{\Theta}$ such that $u_k(\theta, x_k, y_k) \leq u_k(\theta, p_k)$ for some $y_k \in BR_k(\tilde{\Theta}, x_k)$, or (2) $u_k(\theta', x_k, y_k) \geq u_k(\theta', p_k)$ for some $y_k \in BR_k(\tilde{\Theta}, x_k)$, $\theta' \not\in \tilde{\Theta}$. Thus, at least one of these conditions holds for infinitely many $k$. Suppose that the first condition holds for infinitely many $k$; a similar argument handles the case when the second condition holds for infinitely
many $k$. It follows that there is some $\theta \in \tilde{\Theta}$ and some subsequence $\{k_l\}_{l \geq 1}$ such that $u_{k_l}(\theta, x_{k_l}, y_{k_l}) \leq u_{k_l}(\theta, p_{k_l})$ for some $y_{k_l} \in BR_{k_l}(\tilde{\Theta}, x_{k_l})$, and $\lim_{l \to \infty} y_{k_l} = y$ for some $y \in BR(\tilde{\Theta}, x)$. Since $\lim_{l \to \infty} u_{k_l}(\theta, x_{k_l}, y_{k_l}) = u(\theta, x, y)$ and $\lim_{l \to \infty} u_{k_l}(\theta, p_{k_l}) = u(\theta, p)$, it follows that $u(\theta, x, y) \leq u(\theta, p)$. Hence, there is no credible robust neologism corresponding to the sender action $x$ and non-empty subset of sender types $\tilde{\Theta}$. ■

A.2 Proofs of Results in Section 4

A.2.1 Proof of Lemma 1

Lemma 1. In monotonic signaling games, every $D_1$ equilibrium is robust neologism proof.

Proof. We show that every PBE with a credible robust neologism fails the $D_1$ criterion. Suppose that $\pi$ is a PBE with a credible robust neologism corresponding to sender action $x$ and the non-empty subset of sender types $\tilde{\Theta}$. Then, $D_{\theta'}(x, \pi) \cup D_0(\pi) \subseteq D_\theta(x, \pi)$ for all $\theta' \notin \tilde{\Theta}, \theta \in \tilde{\Theta}$, so $\Theta_{D_1}(x, \pi) \subseteq \tilde{\Theta}$. This implies that $u(\theta, x, \beta) > u(\theta, \pi)$ for all $\beta \in MBR(\Theta_{D_1}(\pi), x), \theta \in \tilde{\Theta}$, so $\pi$ cannot be $D_1$. ■

A.2.2 Construction of Strategy Profile $\pi^*$. 

We construct the strategy profile $\pi^*$ described in Subsection 4.2.

Construction of Strategy Profile. Let $P_+(\Theta) = P(\Theta) \setminus \{\emptyset\}$ be the set of non-empty subsets of $\Theta$. For this construction, we take the message space to be $M = P_+(\Theta) \cup \{\emptyset\}$. We are free to do this since every large message space has a strictly greater cardinality than $P_+(\Theta)$, so we can identify some subset of the message space with $P_+(\Theta)$ and we can treat all other messages as redundant “null” messages.

Let $\{t_k\}_{k \geq 1}$ be an enumeration of the non-negative rational numbers and let $T_k = \{t_k\}_{1 \leq k \leq k}$. We set the strategy space of sender type $\theta$ in the $k$-th game to be $\Pi_{k, \theta} = \{\pi_{k, \theta} \in \Delta(X \times T_k \times M)\mid (1) \pi_{k, \theta}(x, 0, \emptyset) \geq \frac{1}{k} \forall x \in X, (2) \forall x \in X, \pi_{k, \theta}(x, t, \emptyset) = 0 \forall t > 0, (3) \forall \tilde{\Theta} \text{ s.t. } \theta \notin \tilde{\Theta}, \pi_{k, \theta}(x, t, \tilde{\Theta}) = 0 \forall (x, t) \in X \times T_k\}$. Throughout the construction,
we implicitly assume that \( k \) is sufficiently large so that this strategy space is well-defined. The first condition is a trembling condition which ensures that all sender types play every combination of action \( x \) with \( t = 0 \) and the null message \( \emptyset \) with positive probability, while the second condition guarantees that no sender type ever plays the combination of an action \( x \) with some \( t > 0 \) and the null message \( \emptyset \). The third condition prevents any sender type from ever sending a message that corresponds to some subset of sender types to which they do not belong.

We turn now to the strategy space of the receiver in the \( k \)-th game. For every \( x \in X \) and \( \tilde{\Theta} \in P_+(\Theta) \), let \( \Pi_{k,R}(x,\tilde{\Theta}) = \Delta(BR(\tilde{\Theta},x)) \). We set the strategy space of the receiver in the \( k \)-th game to be \( \Pi_{k,R} = \Delta(Y)^X \times \prod_{(x,\tilde{\Theta}) \in X \times P_+(\Theta)} \Pi_{k,R}(x,\tilde{\Theta}) \). Thus, a receiver strategy in the \( k \)-th game \( \pi_{k,R} \in \Pi_{k,R} \) consists of some \( \pi_{k,R}(\cdot|x,\emptyset) \in \Delta(Y) \) for every \( x \in X \), which gives the receiver’s response when the sender action \( x \) is paired with the null-message, as well as some \( \pi_{k,R}(\cdot|x,\tilde{\Theta}) \in BR(\tilde{\Theta},x) \) for every \( x \in X \) and \( \tilde{\Theta} \in P_+(\Theta) \), which gives the receiver’s response when the sender action \( x \) is paired with the message \( \tilde{\Theta} \). Note that we are constraining the responses of the receiver in two ways. First, the receiver cannot condition on the transfer \( t \), but rather can only condition on \( x \) and the message \( m \). Second, whenever some non-null message \( m = \tilde{\Theta} \in P_+(\Theta) \) is paired with some \( x \), the receiver can only play actions that are a best response to \( x \) and some belief distribution over sender types with support contained in \( \tilde{\Theta} \).

We set the payoff of the receiver from the strategy profile \( \pi_k = (\pi_{k,R},\pi_{k,-R}) \) in the \( k \)-th game to be

\[
V_k(\pi_{k,R},\pi_{k,-R}) = \sum_{\theta \in \Theta} \sum_{x \in X} \sum_{y \in Y} v(\theta,x,y)\mu(\theta)\pi_{k,\theta}(x,0,\emptyset)\pi_{k,R}(y|x,\emptyset) + \sum_{x \in X} \sum_{t \in T_k} \sum_{\tilde{\Theta} \in P_+(\Theta)} \sum_{\theta \in \tilde{\Theta}} \sum_{y \in BR(\tilde{\Theta},x)} (v(\theta,x,y) + t)\mu(\theta)\pi_{k,\theta}(x,t,\tilde{\Theta})\pi_{k,R}(y|x,\tilde{\Theta}).
\]

This is precisely the expected payoff the receiver would obtain from using the strategy \( \pi_{k,R} \) if the sender types play according to the strategy profile \( \pi_{k,-R} = (\pi_{k,\theta_1},\ldots,\pi_{k,\theta_N}) \).

Additionally, we set the fictitious payoff of type \( \theta \) from the strategy profile \( \pi_k =
\((\pi_{k,\theta}, \pi_{k,-\theta})\) in the \(k\)-th game to be

\[
\bar{U}_k(\theta, \pi_{k,\theta}, \pi_{k,R}) = \sum_{x \in X} \sum_{y \in Y} u(\theta, x, y) \pi_{k,\theta}(x, 0, \emptyset) \pi_{k,R}(y|x, \emptyset) + \sum_{x \in X} \sum_{t \in T_k} \sum_{\tilde{\Theta} \in P_+(\Theta), y \in BR(\tilde{\Theta}, x)} (u(\theta, x, y) - t) \pi_{k,\theta}(x, t, \tilde{\Theta}) \pi_{k,R}(y|x, \tilde{\Theta}).
\]

This is exactly the expected payoff type \(\theta\) would obtain from using the strategy \(\pi_{k,\theta}\) if the receiver responded according to \(\pi_{k,R}\).

Let \(f_k : \mathbb{R} \to \mathbb{R}_+\) be the family of continuous functions given by

\[
f_k(a) = \begin{cases} 
0 & \text{if } a \leq 0 \\
\frac{1}{g(k)} a & \text{if } 0 \leq a \leq g(k) \\
1 & \text{if } a \geq g(k)
\end{cases},
\]

where \(g(k) = \min_{1 \leq k_1 < k_2 \leq k} |t_{k_1} - t_{k_2}|\). Note that \(f_k(a) = 0\) for all \(a \leq 0\) and \(k\), and \(\lim_{k \to \infty} f_k(a) = 1\) for all \(a > 0\). Let \(A \in \mathbb{R}_+\) be such that \(A > 2 \max_{(\theta, x, y)} |u(\theta, x, y)|\) and let \(c_k : \Theta \times X \times T_k \times P_+(\Theta) \times \Pi_{k,S}^N \times \Pi_{k,R} \to \mathbb{R}_+\) be the “cost” function given by

\[
c_k(\theta, x, t, \tilde{\Theta}, \pi_{k,-\theta}) = A \sum_{\theta' \notin \tilde{\Theta}} f_k \left( \sum_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y) \pi_{k,R}(y|x, \tilde{\Theta}) - t - \bar{U}_k(\theta', \pi_{k,\theta'}, \pi_{k,R}) \right).
\]

Note that \(c_k(\theta, x, t, \tilde{\Theta}, \pi_{k,-\theta}) \geq A\) if \(\sum_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y) \pi_{k,R}(y|x, \tilde{\Theta}) - t \geq \bar{U}_k(\theta', \pi_{k,\theta'}, \pi_{k,R}) + g(k)\) for some \(\theta' \notin \tilde{\Theta}\), while \(c_k(\theta, x, t, \tilde{\Theta}, \pi_{k,-\theta}) = 0\) if \(\sum_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y) \pi_{k,R}(y|x, \tilde{\Theta}) - t \leq \bar{U}_k(\theta', \pi_{k,\theta'}, \pi_{k,R})\) for all \(\theta' \notin \tilde{\Theta}\). We set the payoff of type \(\theta\) from the strategy profile \((\pi_{k,\theta}, \pi_{k,-\theta})\) in the \(k\)-th game to be

\[
U_k(\theta, \pi_{k,\theta}, \pi_{k,R}) = \bar{U}(\theta, \pi_{k,\theta}, \pi_{k,R}) - \sum_{x \in X} \sum_{t \in T_k} \sum_{\tilde{\Theta} \in P_+(\Theta), \theta \in \Theta} c_k(\theta, x, t, \tilde{\Theta}, \pi_{k,-\theta}) \pi_{k,\theta}(x, t, \tilde{\Theta}).
\]

This equals the fictitious payoff of type \(\theta\) from the strategy profile \((\pi_{k,\theta}, \pi_{k,-\theta})\) minus
a cost term that penalizes type $\theta$ for combining $(x,t)$ with a non-null message $\tilde{\Theta}$ containing $\theta$ whenever there is some type $\theta' \notin \tilde{\Theta}$ sender such that the expected payoff of type $\theta'$ from playing $(x,t)$ when the receiver responds according to $\pi_{k,R}(y|x,\tilde{\Theta})$ strictly exceeds their fictitious payoff. The cost term is such that no sender type $\theta$ would ever play $(x,t,m)$ where $m = \tilde{\Theta}$ is a non-null message if $\sum_{y \in BR(x)} u(\theta',x,y)\pi_{k,R}(y|x,\tilde{\Theta}) - t \geq \hat{U}_k(\theta',\pi_{k,\theta},\pi_{k,R}) + g(k)$ for some $\theta' \notin \tilde{\Theta}$.

Note that $\Pi_{k,S}$ and $\Pi_{k,R}$ can be thought of as non-empty, convex, and compact subsets of some Euclidean space. Additionally, $U_k(\theta,\pi_{k,\theta},\pi_{k,-\theta})$ and $V_k(\pi_{k,R},\pi_{k,-R})$ are continuous in $\pi_k$, $U_k(\theta,\pi_{k,\theta},\pi_{k,-\theta})$ is linear in $\pi_{k,\theta}$, and $V_k(\pi_{k,R},\pi_{k,-R})$ is linear in $\pi_{k,R}$. Thus, Kakutani’s fixed point theorem ensures the existence of a strategy profile $\pi^*_k$ such that $\pi^*_k(\theta) \in \arg\max_{\pi_{k,\theta} \in \Pi_{k,\theta}} U_k(\theta,\pi_{k,\theta},\pi^*_k(\theta))$ for all $\theta \in \Theta$ and $\pi^*_k \in \arg\max_{\pi_{k,R} \in \Pi_{k,R}} V_k(\pi_{k,R},\pi^*_k(\theta))$.

Since no sender type every uses any transfer $t > A$, $\pi^*_k(\theta) \in \Delta(X \times [0,A] \times M)$ for all $\theta \in \Theta$ and $k$. Thus, there is a $\tilde{\pi} = (\tilde{\pi}_{\theta_1},...,\tilde{\pi}_{\theta_N},\tilde{\pi}_R)$, where $\tilde{\pi}_\theta \in \Delta(X \times T \times M)$ for each $\theta \in \Theta$ and $\tilde{\pi}_R \in \Delta(Y)^X \times \Pi(\Theta) \times \Pi_{k,R}(x,\tilde{\Theta})$, that is a limit point of $\{\pi^*_k\}_k$. Throughout the Appendix, we will assume that $\{\pi^*_k\}_{k \geq 1}$ is a subsequence of $\{\pi^*_k\}_k$ such that $\lim_{l \to \infty} \pi^*_{k_l,\theta} = \tilde{\pi}_\theta$ for all $\theta \in \Theta$ and $\lim_{l \to \infty} \pi^*_{k_l,R} = \tilde{\pi}_R$.

We now use $\tilde{\pi}$ to construct a strategy profile $\pi^*$ in the original game. $\pi^*$ is such that the strategies of all sender types are given by $\tilde{\pi}$, i.e. $\pi^*_\theta \equiv \tilde{\pi}_\theta$ for all $\theta \in \Theta$. The strategy of the receiver is such that the receiver’s response is the same as in $\tilde{\pi}$ for all $(x,t,m)$ where $(t,m) = (0,0)$, i.e. $\pi^*_R(y|x,0,0) = \tilde{\pi}_R(y|x,0)$. Also, for all $(x,t,m)$ with $m = \tilde{\Theta} \in P_+(\Theta)$ that are in the support of some type $\theta$ sender’s strategy $\pi^*_\theta$, the receiver’s response is the same as in $\tilde{\pi}$, i.e. $\pi^*_R(y|x,t,\tilde{\Theta}) = \tilde{\pi}_R(y|x,\tilde{\Theta})$ when there is some $\theta$ such that $(x,t,\tilde{\Theta}) \in \text{supp}(\tilde{\pi}_\theta)$. The receiver’s response to any action when coupled with $m = \emptyset$ is $\pi^*_R(y|x,t,\emptyset) = \tilde{\pi}_R(y|x,\emptyset)$ for all $t \in T$. The receiver’s response to any action when coupled with $m = \tilde{\Theta} \in P_+(\Theta)$ that is not in the support of some type $\theta$ sender’s strategy $\pi^*_\theta$ is $\pi^*_R(y|x,t,\tilde{\Theta}) = \tilde{\pi}_R(y|x,\emptyset)$.

To conclude that $\pi^*$ is a valid strategy profile in the original game, we must show that $\pi^*_R$ is a measurable function mapping $X \times T \times M$ into $\Delta(Y)$. We do so by
showing that, for all $x \in X$ and $m \in M$, there is at most one $t \in T$ such that $(x, t, m) \in \text{supp}(\tilde{\pi}_\theta)$ for some $\theta \in \Theta$. This follows because, for all $\theta \in \Theta$, $x \in X$, and $\tilde{\Theta} \in P_+(\Theta)$, $\pi_{k,\rho}(x, t, \tilde{\Theta}) > 0$ can hold for at most two $t \in T_k$, which must be adjacent, for all $k$ since $g(k) = \min_{1 \leq k_1 < k_2 \leq k} |t_{k_1} - t_{k_2}|$, and the gap in any two such transfers converges to 0 as $k \to \infty$. Similarly, for any two types $\theta, \theta'$, sender action $x$, and subset of sender types $\tilde{\Theta} \in P_+(\Theta)$, if there is $t$ where $(x, t, \tilde{\Theta}) \in \text{supp}(\tilde{\pi}_\theta)$ and $t'$ where $(x, t', \tilde{\Theta}) \in \text{supp}(\tilde{\pi}_{\theta'})$, then $t = t'$. Intuitively, because the receiver’s response in the pre-limit games does not condition on transfers, when choosing $x$ and sending message $\tilde{\Theta}$, neither type should give a much higher transfer than the other (relative to $g(k)$), which translates into them coupling the same transfer with $x$ and $\tilde{\Theta}$ in the limit.

We conclude this construction by noting that the sequences of fictitious payoffs of the sender types in the $k_l$-th game from $\pi^*_{k_l}$ converge to the payoffs of the sender types in the original game from the strategy profile $\pi^*$. Formally, $U(\theta, \pi^*) = \lim_{l \to \infty} \tilde{U}_{k_l}(\theta, \pi^*_{k_l, \theta}, \tau^*_{k_l, R})$ for all $\theta \in \Theta$.

A.2.3 Proof of Lemma 3

Lemma 3. There is no sender action $x$, transfer $t$, and non-empty subset of sender types $\tilde{\Theta}$, such that

1. $\min_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y) - t > U(\theta, \pi^*)$ for some $\theta \in \tilde{\Theta}$, and

2. $\max_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y) - t < U(\theta', \pi^*)$ for all $\theta' \notin \tilde{\Theta}$.

Proof. Suppose there were such a sender action $x$, transfer $t$, subset of sender types $\tilde{\Theta}$, and sender type $\theta \in \tilde{\Theta}$. Without loss of generality, we assume that $t$ is rational. Then, because $\lim_{l \to \infty} \tilde{U}_{k_l}(\theta', \pi^*_{k_l, \theta'}, \tau^*_{k_l, R}) = U(\theta', \pi^*)$ for all $\theta' \notin \tilde{\Theta}$, there is some $L$ such that $\max_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y) - t < \tilde{U}_{k_l}(\theta', \pi^*_{k_l})$ for all $\theta' \notin \tilde{\Theta}$, $l > L$, and consequently $\sum_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y)\pi^*_{k_l, R}(y|x, \tilde{\Theta}) - t < \tilde{U}_{k_l}(\theta', \pi^*_{k_l})$ for all $\theta' \notin \tilde{\Theta}$, $l > L$. Thus, $\liminf_{l \to \infty} U_{k_l}(\theta, \pi^*_{k_l}) \geq \min_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y) - t$, since type $\theta$ would secure this as the limit of their payoffs in the sequence of the $k_l$-th games if they were to play
(x, t, \tilde{\Theta}) against \pi_{k_i,\theta} with the highest probability possible. Since \tilde{U}_{k_1}(\theta, \pi_{k_1,\theta}, \pi_{k_1,R}) \geq U_{k_1}(\theta, \pi_*)$, it follows that \( \lim_{l \to \infty} \tilde{U}_{k_1}(\theta, \pi_{k_1,\theta}, \pi_{k_1,R}) \geq \min_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y) - t \). Because \( U(\theta, \pi_*) = \lim_{l \to \infty} \tilde{U}_{k_1}(\theta, \pi_{k_1,\theta}, \pi_{k_1,R}) \), we conclude that \( U(\theta, \pi_*) \geq \min_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y) - t \), which is a contradiction. \hfill \blacksquare

A.2.4 Proof of Lemma 4

Lemma 4. \pi_* is a PBE.

Proof. By the construction in A.2.2, for any \((x, t, \emptyset)\), the receiver’s play, given by \( \pi_{R}(y|x, t, \emptyset) = \tilde{\pi}_{R}(y|x, \emptyset) \), is a best response to a belief over \( \Theta \) consistent with \( \pi^{-R} \). The same also holds for all \( \tilde{\Theta} \in P_+(\Theta) \) and \((x, t, \tilde{\Theta})\) for which there is no \( \theta \) such that \((x, t, \tilde{\Theta}) \in \text{supp}(\pi_*)\). Moreover, for all \( \tilde{\Theta} \in P_+(\Theta) \) and \((x, t, \tilde{\Theta})\) for which there is some sender type \( \theta \) with \((x, t, \tilde{\Theta}) \in \text{supp}(\pi_*)\), the receiver’s play, given by \( \pi_{R}(y|x, t, \tilde{\Theta}) = \tilde{\pi}_{R}(y|x, \tilde{\Theta}) \), is a best response to a belief over \( \Theta \) consistent with \( \pi^{-R} \). This is because, in the pre-limit games, the receiver’s play, given by \( \pi_{R}(y|x, t, \tilde{\Theta}) \), is a best response to \( x \) and the posterior distribution of sender types, conditional on \( x \) and \( \tilde{\Theta} \) being played, induced by \( \pi^{-R} \). Furthermore, by the construction in A.2.2, \( t \) must be the unique transfer level with which \( x \) and \( \tilde{\Theta} \) are paired according to \( \pi_* \), and any types that play \((x, t, \tilde{\Theta})\) must belong to \( \tilde{\Theta} \). Thus, it follows that playing according to \( \tilde{\pi}_{R}(y|x, \tilde{\Theta}) = \lim_{k \to \infty} \pi_{k,R}(y|x, \tilde{\Theta}) \) following such \((x, t, \tilde{\Theta})\) is optimal for the receiver.

All that remains is to show that all sender types are playing optimally. Since \( \pi_{R}(y|x, t, \emptyset) = \tilde{\pi}_{R}(y|x, \emptyset) \), there can be no profitable deviation for any sender type to any \((x, t, \emptyset)\), by the construction in A.2.2. Likewise, since \( \pi_{R}(y|x, t, \tilde{\Theta}) = \tilde{\pi}_{R}(y|x, \emptyset) \) whenever there is no \( \theta \) with \((x, t, \tilde{\Theta}) \in \text{supp}(\pi_*)\), there can be no profitable deviation for any sender type to such \((x, t, \tilde{\Theta})\). Finally, there are no profitable deviations for any sender type to any \((x, t, \tilde{\Theta})\) for which there is some \( \theta \) such that \((x, t, \tilde{\Theta}) \in \text{supp}(\pi_*)\). To see this, note that for \((x, t, \tilde{\Theta}) \in \text{supp}(\pi_*)\) to hold, it must be that \( \sum_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y)\pi_{R}(y|x, t, \tilde{\Theta}) - t \leq U(\theta', \pi_*) \) for all \( \theta' \notin \tilde{\Theta} \). Otherwise, there would be some \( \theta' \notin \tilde{\Theta} \) and some sufficiently high \( L \) such that \( \sum_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y)\pi_{R}(y|x, \tilde{\Theta}) - t \leq U(\theta', \pi_*) \) for all \( \theta' \notin \tilde{\Theta} \).
t > \tilde{U}_{k_l}(\theta', \pi_{k_l, \theta'}, \pi_{k_l, R}^*) + g(k_l) \text{ for all } l > L, \text{ in which case no sender type would ever want to pair } x \text{ and transfers near } t \text{ with the message } \tilde{\Theta} \text{ in the pre-limit games corresponding to } l > L. \text{ Thus, no type } \theta' \notin \tilde{\Theta} \text{ has a profitable deviation to } (x, t, \tilde{\Theta}). \text{ If type } \theta \in \tilde{\Theta} \text{ had a profitable deviation to } (x, t, \tilde{\Theta}) \text{ in that } \sum_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y)\pi_R^*(y|x, t, \tilde{\Theta}) - t > U(\theta, \pi^*), \text{ then for some rational } t' \text{ slightly larger than } t, \sum_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y)\pi_R^*(y|x, t, \tilde{\Theta}) - t' > U(\theta, \pi^*) \text{ and } \sum_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y)\pi_R^*(y|x, t, \tilde{\Theta}) - t' < U(\theta', \pi^*) \text{ for all } \theta' \notin \tilde{\Theta}. \text{ Then, because } \lim_{l \to \infty} \tilde{U}_{k_l}(\theta', \pi_{k_l, \theta'}, \pi_{k_l, R}^*) = U(\theta', \pi^*) \text{ for all } \theta' \notin \tilde{\Theta}, \text{ there is some } L \text{ such that } \sum_{y \in BR(\tilde{\Theta}, x)} u(\theta', x, y)\pi_{k_l, R}^*(y|x, \tilde{\Theta}) - t' < \tilde{U}_{k_l}(\theta', \pi_{k_l}^*) \text{ for all } \theta' \notin \tilde{\Theta}, l > L. \text{ Thus, } \lim \inf_{l \to \infty} U_{k_l}(\theta, \pi_{k_l}^*) \geq \sum_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y)\pi_R^*(y|x, t, \tilde{\Theta}) - t', \text{ since type } \theta \text{ would secure this as the limit of their payoffs in the sequence of the } k_l\text{-th games if they were to play } (x, t', \tilde{\Theta}) \text{ against } \pi_{k_l, -\theta}^* \text{ with the highest probability possible. Since } \tilde{U}_{k_l}(\theta, \pi_{k_l, \theta}, \pi_{k_l, R}^*) \geq U_{k_l}(\theta, \pi_{k_l}^*), \text{ it follows that } \lim_{l \to \infty} \tilde{U}_{k_l}(\theta, \pi_{k_l, \theta}, \pi_{k_l, R}^*) \geq \sum_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y)\pi_R^*(y|x, t, \tilde{\Theta}) - t'. \text{ Because } U(\theta, \pi^*) = \lim_{l \to \infty} \tilde{U}_{k_l}(\theta, \pi_{k_l, \theta}, \pi_{k_l, R}^*), \text{ we conclude that } U(\theta, \pi^*) \geq \sum_{y \in BR(\tilde{\Theta}, x)} u(\theta, x, y)\pi_R^*(y|x, t, \tilde{\Theta}) - t', \text{ which is a contradiction.} \quad \blacksquare

A.3 Proofs of Results in Section 5

A.3.1 Proof of Lemma 5

Lemma 5. Sender-optimal separating equilibria exist and are precisely the equilibria \( \pi \) in which \( \pi_{\theta}(A_{\theta}) = 1 \) for all \( \theta \in \Theta \). Moreover, every sender-optimal separating equilibrium is outcome equivalent to a PBE.

Proof. Consider a strategy profile \( \pi \) in which, for all \( \theta \in \Theta \), \( \pi_{\theta}(A_{\theta}) = 1 \) and \( \pi_{R}(y^*(\theta, x)|(x, t, m)) = 1 \) for all \( (x, t, m) \in A_{\theta} \), and \( \pi_{R}(y^*(\theta_1, x)|(x, t, m)) = 1 \) for all \( (x, t, m) \notin \bigcup_{\theta \in \Theta} A_{\theta} \). The receiver’s behavior is optimal given posterior beliefs that concentrate on type \( \theta \) following any \( (x, t, m) \in A_{\theta} \) and on type \( \theta_1 \) following any \( (x, t, m) \notin \bigcup_{\theta \in \Theta} A_{\theta} \). Moreover, these beliefs are consistent with the strategies of the sender types. Thus, to show this is a PBE, we need only show that no sender type has a strict incentive to deviate.

First, note that the payoff each sender type \( \theta \) obtains from this strategy profile
is $\overline{U}_\theta$. Therefore, the incentive constraints in (1) imply that no sender type $\theta$ can have a strict incentive to mimic a higher type $\theta'$. We now show that no sender type has a strict incentive to mimic a lower type. Suppose that some type higher than $\theta'$ has a strict incentive to mimic $\theta'$. Let $\tilde{\theta}$ be the lowest type higher than $\theta'$ for which this is true. Then there is some $(x,t,m) \in \overline{A}_{\theta'}$ such that $u(\tilde{\theta}, x, y^*(\theta', x)) - t > \overline{U}_{\tilde{\theta}}$ and $u(\theta, x, y^*(\theta', x)) - t \leq \overline{U}_\theta$ for all $\theta \leq \tilde{\theta}$. Consider $\tilde{t} = t + u(\tilde{\theta}, x, y^*(\theta', x)) - u(\tilde{\theta}, x, y^*(\theta', x))$. Note that $u(\tilde{\theta}, x, y^*(\tilde{\theta}, x)) - \tilde{t} = u(\tilde{\theta}, x, y^*(\theta', x)) - t > \overline{U}_{\tilde{\theta}}$, while by supermodularity $u(\tilde{\theta}, x, y^*(\tilde{\theta}, x)) - \tilde{t} < u(\tilde{\theta}, x, y^*(\theta', x)) - t \leq \overline{U}_\theta$ for all $\theta < \tilde{\theta}$. However, this implies type $\tilde{\theta}$ could obtain a strictly higher payoff than $\overline{U}_{\tilde{\theta}}$ while satisfying all the constraints in (1), a contradiction. A similar argument shows that no sender type has a strict incentive to deviate to any $(x,t,m) \not\in \cup_{\theta \in \Theta} \overline{A}_\theta$. ■

A.3.2 Proof of Lemma 6

Lemma 6. In every robust neologism proof equilibrium $\pi$, $U(\theta, \pi) \geq \overline{U}_\theta$ for all $\theta \in \Theta$.

Proof. Suppose towards a contradiction that $U(\theta, \pi) < \overline{U}_\theta$ for some $\theta \in \Theta$. Let $\underline{\theta}$ be the lowest such sender type. Fix some $(x,t,m) \in \overline{A}_{\underline{\theta}}$, and consider the set $\Theta_{(\underline{\theta},x,\pi)} = \arg\max_{\theta \in \Theta}\{u(\theta, x, y^*(\theta, x)) - U(\theta, \pi)\}$. By definition, $u(\underline{\theta}, x, y^*(\underline{\theta}, x)) - \underline{t} > U(\underline{\theta}, \pi)$ and $u(\theta, x, y^*(\theta, x)) - t \leq U(\theta, \pi)$ for all $\theta < \underline{\theta}$, so no type $\theta < \underline{\theta}$ can belong to $\Theta_{(\underline{\theta},x,\pi)}$. We thus have two cases: Case (1), where $\{\underline{\theta}\} = \Theta_{(\underline{\theta},x,\pi)}$, and Case (2), where there is some $\theta > \underline{\theta}$ such that $\theta \in \Theta_{(\underline{\theta},x,\pi)}$. For Case (1), consider $\tilde{t} = u(\underline{\theta}, x, y^*(\underline{\theta}, x)) - U(\underline{\theta}, \pi) > 0$. Then we have $u(\underline{\theta}, x, y^*(\underline{\theta}, x)) - \tilde{t} = U(\underline{\theta}, \pi)$ and $u(\theta, x, y^*(\theta, x)) - \tilde{t} < U(\theta, \pi)$ for all $\theta \neq \theta'$. It thus follows that for some slightly decreased transfer $\tilde{t} - \varepsilon$, there is a credible robust neologism corresponding to sender action $x$, transfer $\tilde{t} - \varepsilon$, and type $\underline{\theta}$.

For Case (2), consider $\overline{\theta} = \max\{\Theta_{(\underline{\theta},x,\pi)}\}$ and $\Theta_{(\overline{\theta},x,\pi)} = \arg\max_{\theta \in \Theta}\{u(\theta, x, y^*(\overline{\theta}, x)) - U(\theta, \pi)\}$. By supermodularity and the fact that $\overline{\theta} \in \Theta_{(\overline{\theta},x,\pi)}$, it follows that $u(\overline{\theta}, x, y^*(\overline{\theta}, x)) - U(\overline{\theta}, \pi) > u(\theta, x, y^*(\overline{\theta}, x)) - U(\overline{\theta}, \pi)$ for all $\theta < \overline{\theta}$. Thus, again we have two cases: Case (1'), where $\{\overline{\theta}\} = \Theta_{(\overline{\theta},x,\pi)}$, and Case (2'), where there is some $\theta > \overline{\theta}$ such that $\theta \in \Theta_{(\overline{\theta},x,\pi)}$. A similar argument to Case (1) above shows that in Case (1') there is a
credible robust neologism corresponding to sender action \(x\), some transfer, and type \(\tilde{\theta}\).

For Case (2'), we can perform the same procedure and can continue to do so until the analog of Cases (1) or (1') holds. Since the set of sender types is finite, this process must eventually result in a credible robust neologism.

\[\square\]

A.3.3 Proof of Lemma 7

Lemma 7. Every robust neologism proof equilibrium must be separating.

Proof. Consider an equilibrium \(\pi\) that is not separating. Then there is some sender type \(\tilde{\theta}\) such that the set of \((x, t, m)\) at which \(\pi_R\) dictates the receiver play some pure action \(y\) such that \(y < y^*(\tilde{\theta}, x)\) has positive probability under \(\pi_{\tilde{\theta}}\).\[^{29}\] Thus, there exists some \((x, t, m)\) and \(y\) such that \(U(\tilde{\theta}, \pi) = u(\tilde{\theta}, x, y) - t < u(\tilde{\theta}, x, y^*(\tilde{\theta}, x)) - t\) and \(U(\theta, \pi) \geq u(\theta, x, y) - t\) for all \(\theta \in \Theta\).

Consider the set \(\Theta_{(\tilde{\theta},x,\pi)} = \arg \max_{\theta \in \Theta} \{u(\theta, x, y^*(\tilde{\theta}, x)) - U(\theta, \pi)\}\). By supermodularity, it follows that \(u(\tilde{\theta}, x, y^*(\tilde{\theta}, x)) - U(\tilde{\theta}, \pi) > u(\theta, x, y^*(\tilde{\theta}, x)) - U(\theta, \pi)\) for all \(\theta < \tilde{\theta}\). Thus, as in the proof of Lemma 6, we have two cases: Case (1), where \(\{\tilde{\theta}\} = \Theta_{(\tilde{\theta},x,\pi)}\), and Case (2), where there is some \(\theta > \tilde{\theta}\) such that \(\theta \in \Theta_{(\tilde{\theta},x,\pi)}\). A similar argument to the proof of Lemma 6 shows that in Case (1) there is a credible robust neologism corresponding to sender action \(x\), some transfer, and type \(\tilde{\theta}\), and in Case (2) we can iteratively perform the same procedure until we obtain a credible robust neologism. \[\square\]

References


\[^{29}\] Formally, \(\pi_{\tilde{\theta}}(\{(x, t, m) \in X \times M \times T | \exists y \in Y \text{ s.t. } \pi_R(y|x, t, m) = 1 \text{ and } y < y^*(\tilde{\theta}, x)\}) > 0\).


