Learning and Equilibrium Refinements in Signalling Games

Drew Fudenberg† Kevin He‡

First version: August 31, 2017
This version: September 21, 2017

Abstract

We propose two new signalling-game refinements that are microfounded in a model of patient Bayesian learning. Agents are born into player roles and play the signalling game against a random opponent each period. Inexperienced agents know their opponents’ payoff functions but not the prevailing distribution of opponents’ play. One refinement corresponds to an upper bound on the set of possible learning outcomes while the other provides a lower bound. Both refinements are closely related to divine equilibrium (Banks and Sobel, 1987).

*This material was previously part of a larger paper titled “Type-Compatible Equilibria in Signalling Games”. We thank Laura Doval, Glenn Ellison, Lorens Imhof, Yuichiro Kamada, Robert Kleinberg, David K. Levine, Kevin K. Li, Eric Maskin, Dilip Mookherjee, Harry Pei, Matthew Rabin, Bill Sandholm, Lones Smith, Joel Sobel, Philipp Strack, Bruno Strulovici, Tomasz Strzalecki, Jean Tirole, and Juuso Toikka for helpful comments and conversations, and National Science Foundation grant SES 1643517 for financial support.

†Department of Economics, MIT. Email: drew.fudenberg@gmail.com
‡Department of Economics, Harvard University. Email: hesichao@gmail.com
1 Introduction

Signalling games typically have many perfect Bayesian equilibria, because Bayes rule does not pin down the receivers’ off-path beliefs about the senders’ types. Different off-path beliefs for the receiver can justify different off-path receiver behavior, which in turn sustain equilibria with a variety of on-path outcomes. For this reason, applied work using signalling games typically invokes some equilibrium refinement to obtain a smaller and (hopefully) more accurate subset of predictions.

However, most refinements impose restrictions on the equilibrium beliefs without any reference to the process that might lead to equilibrium. Our earlier paper Fudenberg and He (2017) provided a learning-theoretic foundation for a weak form of equilibrium refinement in signalling games, based on the idea that “out of equilibrium” signals are not zero-probability events during learning, but instead arise as relatively rare experiments by inexperienced patient senders trying to learn how the receivers respond to different signals. Unlike the classic refinement literature, we did not assume that agents know their opponents’ payoff functions. This paper explores the additional equilibrium restrictions that follow from supposing that agents do have this prior knowledge. In addition, we complement our past work by providing a sufficient condition for an outcome to emerge as the result of our learning process.

In our learning model, agents are Bayesians who believe they face a fixed but unknown distribution of the opposing players’ strategies. Importantly, the senders hold independent beliefs about how receivers respond to different signals, so they cannot use response to one signal to infer anything about the response to a different signal. This introduces an exploration-exploitation trade-off, as each sender only observes the response to the signals she sends. Long-lived and patient senders will therefore experiment with every signal that they think might yield a substantially higher payoff than the myopically best play. The key to our results is that different types of senders have different incentives for experimenting with various signals, so that some of the sender types will send certain signals more often than others do. Consequently, even though long-lived senders only experiment for a vanishingly small fraction of their lifetimes, the play of the long-lived receivers will be a best response to beliefs about the senders types that reflects this difference in experimentation probabilities.

Of course, the senders’ possible payoffs to various signals depend on which receiver responses the senders deem plausible after each signal. In Fudenberg and He (2017), we assumed that senders’ beliefs have full support over the entire set of receiver strategies, and so assign positive probability to the receivers playing actions that would not be a best response to any belief about sender’s type. In this paper, we instead assume that players know their opponents’ payoff functions and assign zero probability to the event that their opponents play strictly dominated strategies. This means that a smaller set of sender types will experiment with each signal, leading to a stronger restriction on the receiver beliefs. Conversely, receivers know that
no sender type would ever want to experiment with a signal that does not best respond to any receiver strategy, because no possible response by the receiver would make playing that signal worthwhile and the experiment would provide no information about the payoff values of any other signal due to independent prior beliefs. For this reason, the receivers’ beliefs after each signal assign probability zero to the types for whom that signal is dominated.

With these restrictions on the priors, Theorem 1 shows that every patient learning outcome is consistent with “rational compatibility,” while Theorem 2 shows that every equilibrium satisfying a uniform version of rational compatibility and some strictness assumptions can arise as a patient learning outcome. As we show in Section 3 these belief restrictions resemble those imposed by divine equilibrium (Banks and Sobel, 1987): Every divine equilibrium is also consistent with rational compatibility and that every equilibrium satisfying the uniform version of rational compatibility is universally divine.

1.1 Related Literature

This paper is most closely related to the work of Fudenberg and Levine (1993), Fudenberg and Levine (2006), and Fudenberg and He (2017) on patient learning by Bayesian agents who believe they face a steady state distribution of play. Except for the support of the agents’ priors, our learning model is exactly the same as that of Fudenberg and He (2017), and the proof of Theorem 1 closely parallels our results there. Theorem 2 is the main innovation, which establishes that certain equilibria are patient learning outcomes by constructing a suitable prior and investigating the set of patient learning outcomes that arise from it. The only other constructive sufficient condition for strategy profiles to be patient learning outcomes is the subgame-confirmed equilibrium of Fudenberg and Levine (2006), which only applies to a subclass of perfect-information games. The paper is also related to other models of Bayesian non-equilibrium learning, such as Kalai and Lehrer (1993) and Esponda and Pouzo (2016), and to the equilibrium concepts of the Intuitive Criterion Cho and Kreps (1987) and divine equilibrium (Banks and Sobel, 1987). One other contribution of this work relative to Fudenberg and He (2017) is that we compare our learning-based equilibrium refinements with these equilibrium refinements, both of which implicitly assume that players are certain of the payoff functions of their opponents.

2 Two Equilibrium Refinements for Signalling Games

2.1 Signalling Game Notation

A signalling game has two players, a sender (“she”, player 1) and a receiver (“he”, player 2). At the start of the game, the sender learns her type \( \theta \in \Theta \), but the receiver only knows the
sender’s type distribution $\lambda \in \Delta(\Theta)$. Next, the sender chooses a signal $s$ from the finite set $S$. The receiver observes the signal and chooses an action $a$ from the finite set $A$ in response. We assume that $\Theta$ is finite and that $\lambda(\theta) > 0$ for all $\theta$.

The players’ payoffs depend on the triple $(\theta, s, a)$. Let $u_1 : \Theta \times S \times A \to \mathbb{R}$ and $u_2 : \Theta \times S \times A \to \mathbb{R}$ denote the utility functions of the sender and the receiver, respectively.

For $P \subseteq \Delta(\Theta)$, we let $BR(P, s) := \bigcup_{p \in P} \left( \arg \max_{a' \in A} u_2(p, s, a') \right)$. This is the set of best responses to $s$ supported by some belief in $P$. The action $a$ is conditionally dominated after signal $s$ if it is not a best response to any belief about the sender’s type, that is if

$$a \notin BR(\Delta(\Theta), s).$$

Abbreviate $A_s := BR(\Delta(\Theta), s)$. Thus $\Pi_2 := \times_{s \in S} \Delta(A_s)$ is the set of behavior strategies that never play a conditionally dominated action after any signal. $^3$ We call the elements $\pi_2$ of $\Pi_2^*$ the rational receiver strategies.

A behavior strategy for the sender is a collection $\pi_1 = (\pi_1(\cdot | \theta))_{\theta \in \Theta}$, where each $\pi_1(\cdot | \theta)$ is a distribution on $S$. For a given $\pi_1$, signal $s$ is off the path of play if it has probability 0, i.e. $\pi_1(s | \theta) = 0$ for all $\theta$.

Analogous to the definition of $\Pi_2^*$, call a signal $s$ dominated for type $\theta$ if it is not a best response to any $\pi_2 \in \Pi_2$, that is if

$$s \notin \bigcup_{\pi_2 \in \Pi_2} \left( \arg \max_{s' \in S} u_1(\theta, s', \pi_2(\cdot | s')) \right).$$

We denote the set of undominated signals for type $\theta$ by $S_\theta$, so $\Pi_1^* := \times_{\theta \in \Theta} \Delta(S_\theta)$ is the subset of the sender’s behavior strategies where no type ever sends a dominated signal. We also write $\Theta_s$ for the types $\theta$ for whom $s$ is not dominated.

### 2.2 Rationality-Compatible Equilibria

We now introduce rationality-compatible equilibrium (RCE) and uniform rationality-compatible equilibrium (uRCE), two refinements of Nash equilibrium in signalling games.

In Section 4, we develop a steady-state learning model where populations of senders and receivers, initially uncertain as to the aggregate play of the opponent population, undergo random anonymous matching each period to play the signalling game. We study the steady states when agents are patient and long lived, which we term “patiently stable.” Under some strictness assumptions, we show that only RCE can be patiently stable (Theorem 1) and that

---

$^1$The notation $\Delta(X)$ means the set of all probability distribution on $X$.

$^2$Here we extend the domain of $u_2$ to $\Delta(\Theta) \times S \times A$ by expected utility, that is $u_2(p, s, a) := \sum_{\theta \in \Theta} u_2(\theta, s, a) \cdot p(\theta)$.

$^3$Recall that the set of mixed best responses need not be convex.
every uRCE is path-equivalent to a patiently stable profile (Theorem 2). Thus we provide a learning foundation for these solution concepts.

**Definition 1.** Signal $s$ is *more rationally-compatible* with $\theta'$ than $\theta''$, written as $\theta' \succ_s \theta''$, if for every $\pi_2 \in \Pi^2$ such that

$$u_1(\theta'', s, \pi_2(\cdot|s)) \geq \max_{s' \neq s} u_1(\theta'', s', \pi_2(\cdot|s')),$$

we have

$$u_1(\theta', s, \pi_2(\cdot|s)) > \max_{s' \neq s} u_1(\theta', s', \pi_2(\cdot|s')).$$

In words, $\theta' \succ_s \theta''$ means whenever $s$ is a weak best response for $\theta''$ against some rational receiver behavior strategy $\pi_2$, it is also a strict best response for $\theta'$ against $\pi_2$.\(^4\)

We require two auxiliary definitions before defining RCE.

**Definition 2.** For any two types $\theta', \theta''$, let $P_{\theta' \Delta \theta''}$ be the set of beliefs where the odds ratio of $\theta'$ to $\theta''$ exceeds their prior odds ratio, that is

$$P_{\theta' \Delta \theta''} := \left\{ p \in \Delta(\Theta) : \frac{p(\theta'')}{p(\theta')} \leq \frac{\lambda(\theta'')}{\lambda(\theta')} \right\}. \quad (1)$$

Note that if $\pi_1(s'|\theta') \geq \pi_1(s'|\theta''), \pi_1(s'|\theta') > 0$, and the receiver updates beliefs using $\pi_1$, then the receiver’s posterior belief about the sender’s type after observing $s'$ falls in the set $P_{\theta' \Delta \theta''}$. In particular, in any Bayesian Nash equilibrium, the receiver’s on-path belief falls in $P_{\theta' \Delta \theta''}$ after any on-path signal $s$ with $\theta' \succ_s \theta''$.

We now introduce some additional definitions to let us investigate the implications of the agents’ knowledge of their opponent’s payoff function.

**Definition 3.** For any strategy profile $\pi^*$, let

$$\tilde{J}(s, \pi^*) := \left\{ \theta \in \Theta : \max_{a \in A_s} u_1(\theta, s, a) \geq u_1(\theta; \pi^*) \right\}.$$

This is the set of types for which *some* best response to signal $s$ is at least as good as their payoff under $\pi^*$.

**Definition 4.** The set of *rationality-compatible beliefs* for the receiver at strategy profile $\pi^*$, $\{ \tilde{P}(s, \pi^*) \}_{s}$ is defined as follows:

- If $\tilde{J}(s, \pi^*)$ is non-empty,

$$\tilde{P}(s, \pi^*) := \Delta(\tilde{J}(s, \pi^*)) \cap \left\{ P_{\theta' \Delta \theta''} : \theta' \succ_s \theta'' \right\}.$$\(^5\)

\(^4\)This order is more complete than the compatibility order from Fudenberg and He (2017), though it shares the same notation.

\(^5\)With the convention $0 \div 0 := 0$. 

4
• Otherwise, define \( \tilde{P}(s, \pi^*) := \Delta(\Theta) \).

Note that \( \tilde{P} \) assigns probability 0 to equilibrium dominated types; this is similar to the belief restriction of the Intuitive Criterion.

**Definition 5.** Strategy profile \( \pi^* \) is a *rationality-compatible equilibrium (RCE)* if it is a Nash equilibrium and \( \pi^*_2(\cdot|s) \in \Delta(\text{BR}(\tilde{P}(s, \pi^*), s)) \) for every \( s \).

As we show in Section 3, RCE rules out the implausible equilibria in a number of games, but is weaker than some past signalling game refinements in the literature. However, RCE is only a necessary condition for patient stability, which leaves open the question of whether patient learning has additional implication. For this reason, we now define uRCE, a subset of RCE (up to path-equivalence). As we show below, uRCE is a sufficient condition for patient stability, thus providing a bound on its implications.

**Definition 6.** The set of *uniformly rationality-compatible beliefs* for the receiver is \( \{\hat{P}(s)\}_{s \in S} \) where

\[
\hat{P}(s) := \Delta(\Theta_s) \cap \left\{ P_{\theta',\theta''} : \theta' \succ^s \theta'' \right\}.
\]

Note that \( \{\hat{P}(s)\}_{s \in S} \) makes no reference to a particular strategy profile, unlike \( \{\tilde{P}(s, \pi^*)\}_{s \in S} \). Since \( \Delta(\Theta_s) \) contains types for whom \( s \) is undominated and \( \tilde{J}(s, \pi^*) \) contains types for whom \( s \) is equilibrium-undominated (relative to the profile \( \pi^* \)), we have \( \tilde{P}(s, \pi^*) \subseteq \hat{P}(s) \) for every \( \pi^* \) and \( s \).

**Definition 7.** A Nash equilibrium strategy profile \( \pi^* \) is called a *uniform rationality-compatible equilibrium (uRCE)* if for all \( \theta \), all off-path signals \( s \) and all \( a \in \text{BR}(\hat{P}(s), s) \), we have \( u_1(\theta; \pi^*) \geq u_1(\theta, s, a) \).

The “uniformity” in uniform RCE comes from the requirement that *every* best response to *every* belief in \( \hat{P}(s) \) deters *every* type from deviating to the off-path \( s \). By contrast, a RCE is a Nash equilibrium where *some* best response to \( \tilde{P}(s, \pi^*) \) deters every type from deviating to \( s \). Starting with a uRCE, we can modify receiver’s off-path behavior so that they best respond to \( \tilde{P}(s, \pi^*) \) after each off-path signal \( s \). The resulting strategy profile is still a Nash equilibrium due to \( \tilde{P}(s, \pi^*) \subseteq \hat{P}(s) \). This shows that every uRCE is path-equivalent to an RCE.

The following example illustrates that uRCE is a strict subset of RCE in some games.

**Example 1.** Suppose a worker has either high ability (\( \theta_H \)) or low ability (\( \theta_L \)). She chooses between three levels of higher education: None (\( N \)), College (\( C \)), or PhD (\( D \)). An employer observes the worker’s education level and pays a wage. The game has \( u_1(\theta, s, a) = z(a) + v(\theta, s) \) where \( z(\text{low wage}) = 0 \), \( z(\text{medium wage}) = 6 \), \( z(\text{high wage}) = 9 \) and \( v(\theta_H, N) = 0 \), \( v(\theta_L, N) = 0 \), \( v(\theta_H, C) = 2 \), \( v(\theta_L, C) = 1 \), \( v(\theta_H, D) = -2 \), \( v(\theta_L, D) = -4 \). (With this payoff function, going to college has a consumption value while getting a PhD is costly.)
reflect a desire to pay a wage corresponding to the worker’s ability and increased productivity with education.

\[
\begin{array}{|c|c|c|c|} 
\hline
N & low & med & high \\
\hline
\theta_H & 0,2 & 6,0 & 9,1 \\
\theta_L & 0,1 & 6,0 & 9,2 \\
\hline
\end{array}
\quad
\begin{array}{|c|c|c|c|} 
\hline
C & low & med & high \\
\hline
\theta_H & 2,1 & 8,1 & 11,2 \\
\theta_L & 1,2 & 7,1 & 10,1 \\
\hline
\end{array}
\quad
\begin{array}{|c|c|c|c|} 
\hline
D & low & med & high \\
\hline
\theta_H & -2,0 & 4,2 & 7,3 \\
\theta_L & -4,3 & 2,2 & 5,0 \\
\hline
\end{array}
\]

No education level is dominated for any type and no wage is dominated after any signal. Since \( v(\theta_H, \cdot) - v(\theta_L, \cdot) \) is maximized at \( D \), it is simple to verify that \( \theta_H \) is more rationally-compatible with \( D \) than \( \theta_L \) is. Similarly, \( \theta_L \) is more rationally-compatible with \( N \) than \( \theta_H \) is. There is no compatibility relation at signal \( C \).

When the prior is \( \lambda(\theta_H) = 0.5 \), the strategy profile where the employer always pays a medium wage and both types of worker choose \( C \) is a uRCE. This is because \( \hat{P}(N) \) contains only those beliefs with \( p(\theta_H) \leq 0.5 \) and both best responses supported on \( \hat{P}(N) \), low wage and medium wage, deter every type from deviating. At the same time, no type wants to deviate to \( D \), even if she gets paid the best wage. On the other hand, the equilibrium \( \pi^* \) where the employer pays a low wage for \( N \) and \( C \), a medium wage for \( D \), and both types choose \( D \) is an RCE but not a uRCE. The belief that puts probability 1 on the worker being \( \theta_L \) belongs to \( \tilde{P}(N, \pi^*) \) and \( \tilde{P}(C, \pi^*) \) and induces the employer to choose low wage. However, medium salary is a best response to \( \lambda \in \hat{P}(N) \) and medium wage would tempt type \( \theta_L \) to deviate to \( N \). ♦

3 Comparison to Other Equilibrium Refinements

This section compares RCE to other equilibrium refinement ideas in the literature.

3.1 The Compatibility Criterion

First, we relate RCE to the compatibility criterion from Fudenberg and He (2017). Because the rational compatibility order is based on a (weakly) smaller set of receiver strategies than the type compatibility order from Fudenberg and He (2017), it is (weakly) more complete: any pair of types that are ordered by type compatibility are ordered by rational compatibility. Hence, RCE is always at least as restrictive as the equilibrium refinement based on the compatibility criterion. Moreover, RCE can eliminate some equilibria that pass the compatibility criterion.

Example 2. Consider a modification of the beer-quiche game where the receiver can play the new action “Charity” \( (C) \) after observing B. Charity has the same effect as “Not Fight” against a strong type, but gives the receiver slightly less payoff. When played against the weak type, Charity amounts to a large utility transfer from the receiver to the sender.
The quiche-pooling equilibrium in this game satisfies the compatibility criterion, since there is no type compatibility relation at signal B after adding the action C. However, C is never a best response for the receiver after B, so there is no \( \pi_2 \in \Pi^* \) with \( \pi_2(C|B) > 0 \). In this modified game, we still have that B is rationally more compatible with \( \theta_{\text{strong}} \) than with \( \theta_{\text{weak}} \), so that RCE continues to rule out the quiche-pooling equilibrium.

### 3.2 Iterated dominance

We now relate rational compatibility to a form of iterated dominance in the ex-ante strategic form of the game, where the sender chooses a signal as function of her type. We show that every sender strategy that specifies playing signal \( s' \) as a less compatible type \( \theta'' \) but not as a more compatible type \( \theta' \) will be removed by iterated deletion. The idea is that such a strategy is never a weak best response to any receiver strategy in \( \Pi^*_2 \): if the less compatible \( \theta'' \) does not have a profitable deviation, then the more compatible type strictly prefers deviating to \( s' \).

**Proposition 1.** Suppose \( \theta' \succ_{\pi'} \theta'' \). Then any ex-ante strategy of the sender \( \pi_1 \) with \( \pi_1(s'|\theta'') > 0 \) but \( \pi_1(s'|\theta') < 1 \) is removed by strict dominance once the receiver is restricted to using strategies in \( \Pi^*_2 \).

**Proof.** Fix a \( \pi_1 \) with \( \pi_1(s'|\theta'') > 0 \) but \( \pi_1(s'|\theta') < 1 \). Because the space of rational receiver strategies \( \Pi^*_2 \) is convex, it suffices to show there is no receiver strategy \( \pi_2 \in \Pi^*_2 \) such that \( \pi_1 \) is a best response to \( \pi_2 \) in the ex-ante strategic form. If \( \pi_1 \) is an ex-ante best response, then it needs to be at least weakly optimal for type \( \theta'' \) to play \( s' \) against \( \pi_2 \). By \( \theta' \succ_{\pi'} \theta'' \), this implies
\(s'\) is strictly optimal for type \(\theta'\). This shows \(\pi_1\) is not a best response to \(\pi_2\), as the sender can increase her ex-ante expected payoffs by playing \(s'\) with probability 1 when her type is \(\theta'\).

3.3 The Intuitive Criterion and Divine Equilibrium

We next relate RCE to the Intuitive Criterion.

**Proposition 2.** Every RCE satisfies the Intuitive Criterion.

**Proof.** Suppose \(\pi^*\) does not pass the Intuitive Criterion. Then there exists a type \(\theta\) and a signal \(s'\) such that

\[
u_1(\theta; \pi^*) < \min_{a \in \text{BR}(\Delta(J(s', \pi^*)), s)} u_1(\theta, s', a).
\]

If \(\pi^*\) were an RCE, then we would have \(\pi^* \in \Delta(\text{BR}(P(s, \pi^*), s))\). Since \(\tilde{P}(s, \pi^*) \subseteq \Delta(J(s', \pi^*))\), we have

\[
u_1(\theta; \pi^*) < u_1(\theta, s', \pi^* \cdot | s').
\]

This means \(\pi^*\) is not a Nash equilibrium, contradiction.

The next example shows that the set of RCE is strictly smaller than the set of equilibria that pass the Intuitive Criterion.

**Example 3.** Consider a signalling game where the prior probabilities of the two types are 
\(\lambda(\theta_1) = 3/4\) and \(\lambda(\theta_2) = 1/4\), and the payoffs are:

<table>
<thead>
<tr>
<th></th>
<th>(s')</th>
<th>(a')</th>
<th>(a'')</th>
<th>(s'')</th>
<th>(a')</th>
<th>(a'')</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\theta')</td>
<td>4, 1</td>
<td>0, 0</td>
<td></td>
<td>(\theta')</td>
<td>7, 1</td>
<td>3, 0</td>
</tr>
<tr>
<td>(\theta'')</td>
<td>6, 0</td>
<td>2, 1</td>
<td></td>
<td>(\theta'')</td>
<td>7, 0</td>
<td>3, 1</td>
</tr>
</tbody>
</table>

Against any receiver strategy, the two types \(\theta'\) and \(\theta''\) get the same payoffs from \(s''\), but \(\theta''\) gets strictly higher payoffs than \(\theta'\) from \(s'\). So, \(\theta' \succ_{s''} \theta''\).

Consider now the Nash equilibrium in which the types pool on \(s'\), i.e. \(\pi^*_1(s' | \theta') = \pi^*_1(s' | \theta'') = 1\), \(\pi^*_2(a' | s') = 1\), and \(\pi^*_2(a'' | s'') = 1\). It passes the Intuitive Criterion since the off-path signal \(s''\) is not equilibrium dominated for either type. On the other hand, RCE requires that every action played with positive probability in \(\pi^*_2(\cdot | s'')\) best responds to some belief \(p\) about sender’s type satisfying 
\[p(\theta_2) \leq \frac{\lambda(\theta_2)}{\lambda(\theta_1)} = \frac{1}{3}\]. But action \(a''\) does not best respond to any such belief, so \(\pi^*\) is not an RCE.

Finally, we compare divine equilibrium with RCE and uRCE. For a strategy profile \(\pi^*\), let

\[D(\theta, s; \pi^*) := \{\alpha \in \text{MBR}(s) \text{ s.t. } u_1(\theta; \pi^*) < u_1(\theta, s, \alpha)\}\]
be the subset of mixed best responses\textsuperscript{6} to \(s\) that would make type \(\theta\) strictly prefer deviating from the strategy \(\pi_1^*(|\theta|)\). Similarly let

\[ D^\circ(\theta, s; \pi^*) := \{\alpha \in \text{MBR}(s) \text{ s.t. } u_1(\theta; \pi^*) = u_1(\theta, s, \alpha)\} \]

be the set of mixed best responses that would make \(\theta\) indifferent to deviating.

**Proposition 3.** 1. If \(\pi^*\) is a Nash equilibrium where \(s'\) is off-path, and \(\theta' \succ_{s'} \theta''\), then

\[ D(\theta'', s'; \pi^*) \cup D^\circ(\theta'', s'; \pi^*) \subseteq D(\theta', s'; \pi^*). \]

2. Every divine equilibrium is a RCE.

**Proof.** To show (a), note first that if \(D(\theta'', s'; \pi^*) \cup D^\circ(\theta'', s'; \pi^*) = \emptyset\) the conclusion holds vacuously. If \(D(\theta'', s'; \pi^*) \cup D^\circ(\theta'', s'; \pi^*)\) is not empty, take any \(\alpha' \in D(\theta'', s'; \pi^*) \cup D^\circ(\theta'', s'; \pi^*)\) and define \(\pi_2^\prime \in \Pi_2^\ast\) by \(\pi_2^\prime(|s') = \alpha'\), \(\pi_2^\prime(|s) = \pi_2^*(|s)\) for \(s \neq s'\). Then

\[ u_1(\theta''; \pi^*) = \max_{s \neq s'} u_1(\theta'', s, \pi_2^\prime(|s')) \leq u_1(\theta'', s', \pi_2^\prime(|s')) = u_1(\theta'', s', \alpha'), \]

and when \(\theta' \succ_{s'} \theta''\), this implies that

\[ u_1(\theta'; \pi^*) = \max_{s \neq s'} u_1(\theta', s, \pi_2^\prime(|s')) < u_1(\theta', s', \pi_2^\prime(|s')) = u_1(\theta', s, \alpha'). \]

Hence \(\alpha' \in D(\theta', s'; \pi^*)\).

To show (b), suppose \(\pi^*\) is a divine equilibrium. Then it is a Nash equilibrium, and furthermore for any off-path signal \(s'\) where \(\theta' \succ_{s'} \theta''\), Proposition 3(a) implies that

\[ D(\theta'', s'; \pi^*) \cup D^\circ(\theta'', s'; \pi^*) \subseteq D(\theta', s'; \pi^*). \]

Since \(\pi^*\) is a divine equilibrium, \(\pi_2^*(|s')\) must then best respond to some belief \(p \in \Delta(\Theta)\) with

\[ \frac{p(\theta'')}{p(\theta''')} \leq \frac{\lambda(\theta'')}{\lambda(\theta')} \cdot \]

Considering all \((\theta', \theta'')\) pairs, we see that in a divine equilibrium \(\pi_2^*(|s')\) best responds to some belief in \(\{P_{\theta'\diamond \theta'': \theta' \succ_{s'} \theta''}\}\). At the same time, in every divine equilibrium belief after off-path \(s'\) puts zero probability on equilibrium-dominated types, meaning \(\pi_2^*(|s')\) best responds \(\Delta(\tilde{J}(s', \pi^*))\). This shows \(\pi^*\) is an RCE. \(\square\)

However, the converse is not true, as the following example illustrates.

**Example 4.** Consider the following signalling game with two types and three signals, with prior \(\lambda(\theta_1) = 2/3\).

\textsuperscript{6}To be precise, \(\text{MBR}(p, s) := \arg \max_{\alpha \in \Delta(A)} u_2(p, s, \alpha)\) and \(\text{MBR}(s) := \cup_{p \in \Delta(\Theta)} \text{MBR}(p, s)\).
We claim that the following is a pure-strategy RCE: $\pi_1(s'|θ') = π_1(s'|θ'') = 1, π_2(a'|s') = 1, π_2(a''|s'') = 1$. Evidently $π$ is a Nash equilibrium and no type is equilibrium-dominated at any off-path signal. We now check that we do not have $θ' \succ s'' θ''$ or $θ'' \succ s'' θ'$. Observe that against the receiver strategy $\hat{π}_2(a'|s) = \frac{1}{3}$ for every $s$, $s''$ is strictly optimal for $θ''$ but $s''$ is strictly optimal for $θ'$, so $θ' \not\succ s'' θ''$. And for the receiver strategy $\hat{π}_2(a'|s) = 1$ for every $s$, $s''$ is strictly optimal for $θ'$ but $s''$ is strictly optimal for $θ''$, so $θ'' \not\succ s'' θ'$. This shows the strategy profile is an RCE.

However, $D(θ'', s''; π) \cup D^c(θ'', s''; π)$ is the set of distributions on $\{a', a''\}$ that put at least weight 0.5 on $a'$. Any such distribution is in $D(θ', s''; π)$. So in every divine equilibrium, the receiver plays a best response to a belief that puts weight no less than 2/3 on $θ'$ after signal $s''$, which can only be $a'$.

This example illustrates one difference between divine equilibrium and RCE: under divine equilibrium, the beliefs after signal $s''$ only depend on the comparison between the payoffs to $s''$ with those of the equilibrium signal $s'$, while the compatibility criterion also considers the payoffs to a third signal $s'''$. In the learning model, this corresponds to the possibility that $θ'$ chooses to experiment with $s'''$ at beliefs that induce $θ''$ to experiment with $s''$.

Our RCE differs from divine equilibrium in another way: divine equilibrium involves an iterative application of a belief restriction. The next example illustrates this difference.

**Example 5.** There are three types, $θ', θ'', θ'''$, all equally likely. The signal space is $S = \{s', s''\}$, and the set of receiver actions is $A = \{a^1, a^2, a^3, a^4\}$. When any sender type chooses the signal $s'$, all parties get a payoff of 0 regardless of the receiver’s action. When the sender chooses $s''$, the payoffs are determined by the following matrix.

<table>
<thead>
<tr>
<th>$s'$</th>
<th>$a^1$</th>
<th>$a^2$</th>
<th>$a^3$</th>
<th>$a^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$θ'$</td>
<td>1, 0.9</td>
<td>-1, 0</td>
<td>-2, 0</td>
<td>-7, 0</td>
</tr>
<tr>
<td>$θ''$</td>
<td>5, 0</td>
<td>3, 1</td>
<td>-1, 0</td>
<td>-5, 0.8</td>
</tr>
<tr>
<td>$θ'''$</td>
<td>-3, 0</td>
<td>5, 0</td>
<td>1, 1.7</td>
<td>-3, 0.8</td>
</tr>
</tbody>
</table>

Consider the pure strategy profile $\pi_1(s'|θ) = 1$ for all $θ ∈ Θ$ and $\pi_2(a^4|s) = 1$ for all $s ∈ S$. Since $θ'''$ gains more from deviating to $s''$ than $θ'$ does, applying the divine belief restriction

---

7As noted by Van Damme (1987), it may seem more natural to replace the set $α ∈ MBR(m)$ in the definitions of $D$ and $D^c$ with the larger set $α ∈ co(BR(s))$, which leads to the weaker equilibrium refinement that Sobel, Stole, and Zapater (1990) call “co-divinity”. This example also shows that RCE need not be co-divine.

8We thank Joel Sobel for this example.
for the off-path signal $s''$ eliminates the action $a^1$, since it is not a best response to any belief $p \in \Delta(\Theta)$ with $p(\theta'') \geq p(\theta')$. But after action $a^1$ is deleted for the receiver after signal $s''$, type $\theta''$ now gains more from deviating to $s''$ than $\theta''$ does. So, applying the divine belief restriction again eliminates actions $a^2$ and $a^4$, since it is not a best response against any $p \in \Delta(\Theta)$ with $p(\theta') = 0$ (for now $s''$ is equilibrium dominated for $\theta'$) and $p(\theta''') \geq p(\theta'')$. So $\pi^*$ is not a divine equilibrium.

On the other hand, no type is equilibrium dominated at $s''$ and the only rational compatibility order is $\theta'' \succ_{s''} \theta'$. But $a^4$ is a best response against the belief $p(\theta') = 0$, $p(\theta'') = 0.6$, $p(\theta''') = 0.4$, which belongs to the set $\Delta(\Theta_{s''}) \cap P_{\theta''} \cup \theta'$. So $\pi^*$ is an RCE.

Finally, we show that every uRCE is path-equivalent to an equilibrium that is not ruled out by the “NWBR in signalling games” test (Banks and Sobel, 1987; Cho and Kreps, 1987), which comes from iterative applications of the following pruning procedure: after signal $s$ the receiver is required to put 0 probability on those types $\theta$ such that

$$D^\omega(\theta, s, \pi^*) \subseteq \bigcup_{\theta' \neq \theta} D(\theta', s; \pi^*).$$

If this would delete every type, then the procedure instead puts no restriction on receiver’s beliefs and no type is deleted.

By “path-equivalent” we mean that by modifying some of the receiver’s off-path responses, but without altering sender’s strategy or receiver’s on-path responses, we can change the uRCE into another uRCE that passes the NWBR test. Since every equilibrium passing the NWBR test is universally divine (Cho and Kreps, 1987), this implies that every uRCE is path-equivalent to a universally divine equilibrium.

**Proposition 4.** Every uRCE is path-equivalent to a uRCE that passes the NWBR test.

**Proof.** Consider a uRCE $\pi^*$. For every off-path $s$, perform the following modifications on $\pi^*_2(\cdot|s)$: if the first-round application of the NWBR procedure would have deleted every type, then do not modify $\pi^*_2(\cdot|s)$. Otherwise, find some $\theta_s$ not deleted by the iterated NWBR procedure, then change $\pi^*_2(\cdot|s)$ to some action in $\text{BR}(\theta_s, m)$, i.e. a best response to the belief putting probability 1 on $\theta_s$.

This modified strategy profile passes the NWBR test. We now establish that it remains a uRCE by checking that for those off-path $s$ where $\pi^*_2(\cdot|s)$ was modified, the modified version is still a best response to $\hat{P}(s)$. (By uniformity, this would ensure that the modified receiver play continues to deter every type from deviating to $s$.)

Type $\theta_s$ satisfies $\theta_s \in \Theta_s$. Otherwise, $D^\omega(\theta_s, s; \pi^*) = \emptyset$ and $\theta_s$ would have been deleted by NWBR in the first round. Now it suffices to argue there is no $\theta'$ such that $\theta' \succ_{s'} \theta_s$, which implies the belief putting probability 1 on $\theta_s$ is in $\hat{P}(s)$. If there were such $\theta'$, by Proposition

---

9This is closely related to, but not the same as, the NWBR property of Kohlberg and Mertens (1986).
3(a) we would have $D^o(\theta_s, s; \pi^*) \subseteq D(\theta', s; \pi^*)$, so $\theta_s$ should have been deleted by NWBR in the first round, contradicting the fact that $\theta_s$ survives all iterations of the NWBR procedure. □

**Corollary 1.** Every uRCE is path-equivalent to a universally divine equilibrium.

*Proof.* This is follows from Proposition 4 because every NWBR equilibrium is a universally divine equilibrium. □

To summarize this subsection, we note that for strategy profiles that are on-path strict for the receiver, we have the following strict inclusions (where the first $\subseteq$ should be understood as inclusion up to path-equivalence).

\[
uRCE \subsetneq \text{universally divine equilibria} \subsetneq \text{RCE} \subsetneq \text{Intuitive Criterion} \subsetneq \text{Nash equilibria}.
\]

## 4 Steady-State Learning in Signalling Games

### 4.1 Random Matching and Aggregate Play

We study the same discrete-time steady-state learning model as Fudenberg and He (2017) except for an extra restriction on the players’ prior beliefs over other players’ strategies. For that reason we describe it relatively tersely; please see the earlier paper for a fuller motivation of its assumptions.

There is a continuum of agents in the society, with a unit mass of receivers and $\lambda(\theta)$ mass of type $\theta$ senders. Each population is further stratified by age, where $(1 - \gamma) \cdot \gamma^t$ fraction of each population is aged $t$ for $t = 0, 1, 2, \ldots$ At the end of each period, each agent has probability $0 \leq \gamma < 1$ of surviving into the next period, increasing their age by 1. With complementary probability, the agent dies. At the start of the next period, $(1 - \gamma)$ new receivers and $\lambda(\theta)(1 - \gamma)$ new type $\theta$ are born into the society, thus preserving population sizes and the age distribution.

Agents play the signalling game every period against a randomly matched opponent. Each sender has probability $(1 - \gamma)\gamma^t$ of matching with a receiver of age $t$, while each receiver has probability $\lambda(\theta)(1 - \gamma)\gamma^t$ of matching with a type $\theta$ of age $t$.

### 4.2 Learning by Individual Agents

Each agent is born into a player role in the signalling game: either a receiver or a type $\theta$ sender. Agents know their role, which is fixed for life. The agents’ payoffs each period are determined by the outcome of the signalling game they played, which consists of the sender’s type, the signal sent, and the action played in response. The agents observe this outcome, but senders does not observe how her matched receiver would have played had she sent a different signal.
Agents survive between periods with probability $0 \leq \gamma < 1$. They further discount future utility flows by $0 \leq \delta < 1$ and seek to maximize expected discounted utility. Letting $u_t$ represent payoff in $t$ periods, each agent’s objective function is $E[\sum_{t=0}^{\infty} (\gamma \delta)^t \cdot u_t]$.

Agents believe they face a fixed but unknown distribution of opponents’ aggregate play, updating their beliefs at the end of every period based on their personal experience. Formally, each sender is born with a prior density function over receiver’s behavior strategies, $g_1 : \Pi_2 \to \mathbb{R}_+$. Similarly, each receiver is born with a prior density over the sender’s behavior strategies, $g_2 : \Pi_1 \to \mathbb{R}_+$. We denote the marginal distribution of $g_1$ on signal $s$ as $g_1^{(s)}$, so that $g_1^{(s)}(\pi_2(\cdot|s))$ is the density of the new senders’ prior over how receivers respond to signal $s$. Similarly, we denote the $\theta$ marginal of $g_2$ as $g_2^{(\theta)}$, so that $g_2^{(\theta)}(\pi_1(\cdot|\theta))$ is the new receivers’ prior density over $\pi_1(\cdot|\theta) \in \Delta(S)$.

We now state a regularity assumption on agents’ priors that will be maintained throughout.

**Definition 8.** A prior $g = (g_1, g_2)$ is regular if

(a). [independence] $g_1(\pi_2) = \prod_{s \in S} g_1^{(s)}(\pi_2(\cdot|s))$ and $g_2(\pi_1) = \prod_{\theta \in \Theta} g_2^{(\theta)}(\pi_1(\cdot|\theta))$.

(b). [rationalizability] $g_1$ puts probability 1 on $\Pi_2^*$ and $g_2$ puts probability 1 on $\Pi_1^*$.

(c). [$g_1$ non-doctrinaire] $g_1$ is continuous and strictly positive on the interior of $\Pi_2^*$.

(d). [$g_2$ nice] For each type $\theta$, there are positive constants $(\alpha_s^{(\theta)})_{s \in S}$ such that

$$
\pi_1(\cdot|\theta) \mapsto \frac{g_2^{(\theta)}(\pi_1(\cdot|\theta))}{\prod_{s \in S} \pi_1(s|\theta)^{\alpha_s^{(\theta)} - 1}}
$$

is uniformly continuous and bounded away from zero on the relative interior of $\Pi_\theta^*$, the set of rational behavior strategies of type $\theta$.

This assumption bears the same name as the regularity assumption in our previous paper (Fudenberg and He, 2017), and is identical except that agents now believe their opponents will not play dominated strategies, so that beliefs are constrained by the rationalizability condition.

Recall that $\Pi_\theta^* := \times_{\theta \in \Theta} (S_\theta^*)$, the product over types of the undominated signals for each type. Thus the assumption that the receiver’s prior is rationalizable is compatible with the assumption that the prior is regular. Moreover, rationalizable receiver priors can assign positive probability on sender behavior strategies that are supported on ex-ante dominated sender strategies. This fits with the learning model because types are fixed at birth and different types can have different histories and thus hold different beliefs. For instance, in Example 2 of Fudenberg and He (2017) (reproduced below),
there is a learning outcome for infinitely long-lived agents where both types play $s''$. This sender strategy is not a best response to any receiver strategy. The signal choices of the two types $\theta'$ and $\theta''$ in that example do belong to $\Pi_{\theta'}$ and $\Pi_{\theta''}$, but the receiver strategies to which they respectively best respond form disjoint sets.

4.3 History and Aggregate Play

Let $Y_\theta[t] := (\cup_{s \in S}(s \times A_s))^t$ represent the set of histories for a type $\theta$ with age $t$. Note that a valid history encodes the signal that $\theta$ sent each period and the rational action that her opponent played in response. Let $Y_\theta := \bigcup_{t=0}^\infty Y_\theta[t]$ be the set of all histories for type $\theta$.

Similarly, write $Y_2[t] := (\Theta \times S_2)^t$ for the set of histories for a receiver with age $t$. Each period, his history encodes the type of the matched sender and the rational signal she sent. The union $Y_2 := \bigcup_{t=0}^\infty Y_2[t]$ then stands for the set of all receiver histories.

The agents’ dynamic optimization problems discussed in Subsection 4.2 give rise to optimal policy functions $\sigma_\theta : Y_\theta \rightarrow S_\theta$ and $\sigma_2 : Y_2 \rightarrow \times_s(A_s)$. Here, $\sigma_\theta(y_\theta)$ is the signal that a type $\theta$ with history $y_\theta$ would send the next time she plays the signalling game. Since $\theta$ starts with a prior satisfying rationalizability, $\sigma_\theta$ would never prescribe a signal outside of $S_\theta$. Analogously, $\sigma_2(y_2)$ is the pure strategy that a receiver with history $y_2$ would commit to next time he plays the game.

A state $\psi$ of the learning model is a demographic description of how many agents have each possible history. It can be viewed as a distribution

$$\psi \in (\times_{\theta \in \Theta} \Delta(Y_\theta)) \times \Delta(Y_2)$$

and its components are denoted by $\psi_\theta \in \Delta(Y_\theta)$ and $\psi_2 \in \Delta(Y_2)$.

Given the agents’ optimal policies, each history determines how that agent will play in
their next match. Therefore, each state \( \psi \) induces a rational behavior strategy in the signalling game, given by

\[
\sigma_\theta(\psi_\theta)(s) := \psi_\theta \{y_\theta \in Y_\theta : \sigma_\theta(y_\theta) = s\}.
\]

(2)

and

\[
\sigma_2(\psi_2)(a|s) := \psi_2 \{y_2 \in Y_2 : \sigma_2(y_2)(s) = a\}.
\]

Here, \( \sigma_\theta(\psi_\theta) \) and \( \sigma_2(\psi_2) \) are the aggregate behaviors of the type \( \theta \) and receiver populations in state \( \psi \), respectively.

Of particular interest are steady state, to be defined more precisely in Section 6. Loosely speaking, a steady state induces a time-invariant distribution over how the signalling game is played in the society.

5 Aggregate Responses and Steady State

5.1 The Aggregate Sender Response

Suppose we fix the receivers’ aggregate play at \( \pi_2 \in \Pi^*_2 \) and suppose that all senders use the policy rule \( \sigma_\theta \) to map their observations to their play. The aggregate sender response describes the resulting aggregate play of the sender population.

To answer this question, we first introduce the one-period-forward map.

**Definition 9.** The one-period-forward map for type \( \theta \), \( f_\theta : \Delta(Y_\theta) \times \Pi^*_2 \to \Delta(Y_\theta) \) is

\[
f_\theta[\psi_\theta, \pi_2](y_\theta, (s, a)) := \psi_\theta(y_\theta) \cdot \gamma \cdot 1\{\sigma_\theta(y_\theta) = s\} \cdot \pi_2(a|s)
\]

and \( f_\theta(\emptyset) := 1 - \gamma \).

If the distribution over histories in the type-\( \theta \) population is \( \psi_\theta \) today and the receiver population’s aggregate play is \( \pi_2 \), then \( f_\theta[\psi_\theta, \pi_2] \) describes the distribution over histories that will prevail in the type \( \theta \) population tomorrow.

Write \( f^T_\theta \) for the \( T \)-fold application of \( f_\theta \) on \( \Delta(Y_\theta) \), holding fixed some \( \pi_2 \). As discussed in Fudenberg and He (2017), \( \lim_{T \to \infty} f^T_\theta(\psi_\theta, \pi_2) \) exists and is independent of the initial \( \psi_\theta \). Denote this limit as \( \psi^T_\theta \). It is the distribution over type-\( \theta \) history induced by \( \pi_2 \).

**Definition 10.** The aggregate sender response \( R_1 : \Pi^*_2 \to \Pi^*_1 \) is defined by

\[
R_1[\pi_2](s|\theta) := \psi^T_\theta(y_\theta : \sigma_\theta(y_\theta) = s)
\]
That is, $\mathcal{R}_1[\pi_2](\cdot|\theta)$ describes the asymptotic aggregate play of the type-$\theta$ population when the aggregate play of the receiver population is fixed at $\pi_2$ each period. Note that $\mathcal{R}_1$ maps into $\Pi_1^*$ because no type ever wants to send a dominated signal, even as an experiment, regardless of their beliefs about the receiver’s response.

Remark 1. Technically, $\mathcal{R}_1$ depends on $g_1, \delta, \text{ and } \gamma$, just like $\sigma_\theta$ does. When relevant, we will make these dependencies clear by adding the appropriate parameters as superscripts to $\mathcal{R}_1$, but we will mostly suppress them to lighten notation.

5.2 The Aggregate Receiver Response

We now turn to the receiver population’s learning problem. Since the receivers observe the sender’s type and signal each period regardless which action he takes, the optimal policy $\sigma_2$ best responds to the posterior belief at every history $y_2$ and so maps into $\Pi_2^*$.

Definition 11. The one-period-forward map for the receivers $f_2: \Delta(Y_2) \times \Pi_1^* \to \Delta(Y_2)$ is

$$f_2[\psi_2, \pi_1](y_2, (\theta, s)) := \psi_2(y_2) \cdot \gamma \cdot \lambda(\theta) \cdot \pi_1(s|\theta)$$

and $f_2(\emptyset) := 1 - \gamma$.

As with the one-period-forward maps $f_\theta$ for senders, $f_2[\psi_2, \pi_1]$ describes the distribution over receiver histories tomorrow if the distribution over histories in the receiver population today is $\psi_2$ and sender population’s aggregate play is $\pi_1$. We write $\psi_2^{\pi_1} := \lim_{T \to \infty} f_2^T(\psi_2, \pi_1)$ for the long-run distribution over $Y_2$ induced by fixing sender population’s play at $\pi_1$.

Definition 12. The aggregate receiver response $\mathcal{R}_2: \Pi_1^* \to \Pi_2^*$ is

$$\mathcal{R}_2[\pi_1](a|s) := \psi_2^{\pi_1}(y_2 : \sigma_2(y_2)(s) = a)$$

5.3 Steady States and Patient Stability

A steady state strategy profile is pair of mutual aggregate replies, so it is time-invariant under learning

Definition 13. $\pi^*$ is a steady state strategy profile if $\mathcal{R}_1^{g,\delta,\gamma}(\pi_2^*) = \pi_1^*$ and $\mathcal{R}_2^{g,\delta,\gamma}(\pi_1^*) = \pi_2^*$. Denote the set of all such strategy profiles as $\Pi^*(g, \delta, \gamma)$.

We now state without proof three results about these steady states, all of which follow easily from analogous results in Fudenberg and He (2017).

Proposition 5. $\Pi^*(g, \delta, \gamma)$ is non-empty and compact in the norm topology.
The patiently stable strategy profiles under \( g \), loosely speaking, correspond to the set \( \lim_{\delta \to 1} \lim_{\gamma \to 1} \Pi^*(g, \delta, \gamma) \).

**Definition 14.** For each \( 0 \leq \delta < 1 \), a strategy profile \( \pi^* \) is \( \delta \)-stable under \( g \) if there is a sequence \( \gamma_k \to 1 \) and an associated sequence of steady state strategy profiles \( \pi^{(k)} \in \Pi^*(g, \delta, \gamma_k) \), such that \( \pi^{(k)} \to \pi^* \). Strategy profile \( \pi^* \) is patiently stable under \( g \) if there is a sequence \( \delta_k \to 1 \) and an associated sequence of strategy profiles \( \pi^{(k)} \) where each \( \pi^{(k)} \) is \( \delta_k \)-stable under \( g \) and \( \pi^{(k)} \to \pi^* \). Strategy profile \( \pi^* \) is patiently stable if it is patiently stable under some regular prior \( g \).

In a signalling game, a perfect Bayesian equilibrium with heterogeneous off-path beliefs is a strategy profile \( (\pi^*_1, \pi^*_2) \) such that:

- For each \( \theta \in \Theta \), \( u_1(\theta; \pi^*) = \max_{s \in S} u_1(\theta, s, \pi^*_2(\cdot | s)) \).
- For each on-path signal \( s \), \( u_2(p^*(\cdot | s), s, \pi^*_2(\cdot | s)) = \max_{\hat{a} \in A} u_2(p^*(\cdot | s), s, \hat{a}) \).
- For each off-path signal \( s \) and each \( a \in A \) with \( \pi^*_2(a|s) > 0 \), there exists a belief \( p \in \Delta(\Theta) \) such that \( u_2(p, s, a) = \max_{\hat{a} \in A} u_2(p, s, \hat{a}) \).

**Proposition 6.** If strategy profile \( \pi^* \) is patiently stable, then it is a perfect Bayesian equilibrium with heterogeneous off-path beliefs.

Finally, we record an alternative characterization of the steady state strategy profiles in terms of fixed points of the one-period-forward maps defined earlier.

**Definition 15.** A state \( \psi^* \) is a steady state if \( \psi^*_\theta = f_\theta(\psi^*_\theta, \sigma_2(\psi^*_\theta)) \) for every \( \theta \) and \( \psi^*_2 = f_2(\psi^*_2, (\sigma_\theta(\psi^*_\theta)))_{\theta \in \Theta} \). The set of all steady states for regular prior \( g \) and \( 0 \leq \delta, \gamma < 1 \) is denoted \( \Psi^*(g, \delta, \gamma) \).

**Proposition 7.** \( \pi^* \in \Pi^*(g, \delta, \gamma) \) if and only if \( \mathcal{R}_1^{\delta, \gamma}(\pi^*_1) = \pi^*_1 \) and \( \mathcal{R}_2^{\delta, \gamma}(\pi^*_1) = \pi^*_2 \).

### 6 Patient Stability and Equilibrium Refinements

In this section, we relate the equilibrium refinements proposed in Section 2 to the steady-state learning model. We show that under certain strictness assumptions, RCE is necessary for patient stability while uRCE is sufficient for patient stability.

#### 6.1 RCE Is Necessary for Patient Stability

Here we show that any patiently stable strategy profile that satisfies a strictness assumption must be an RCE. The key lemma is analogous to Lemma 1 from Fudenberg and He (2017):
Lemma 1. Suppose \( \theta' \succ_{s'} \theta'' \). Then for any regular prior \( g_1, 0 \leq \delta, \gamma < 1 \), and any \( \pi_2 \in \Pi_2^\bullet \), we have \( \mathcal{R}_1[\pi_2](s' | \theta') \geq \mathcal{R}_1[\pi_2](s' | \theta'') \).

This result says the rates that different sender types experiment with \( s' \) is monotonic with respect to the rationally-compatible order \( \succ_{s'} \). This follows from the fact that sender types who are more compatible with a signal will play it at least as often, following the same argument as in Fudenberg and He (2017). The rationalizability assumption on \( g_1 \) implies that senders never experiment in the hopes of seeing a response which is highly profitable for the sender but dominated for the receiver, such as the “charity” action in Example 2. This extra assumption leads to a stronger result than Lemma 1 from Fudenberg and He (2017), which is stated in terms of the less-complete compatibility order.

Definition 16. A Nash equilibrium \( \pi^* \) is on-path strict for the receiver if for every on-path signal \( s^* \), \( \pi_2(a^* | s^*) = 1 \) for some \( a^* \in A \) and \( u_2(s^*, a^*, \pi_1) > \max_{a \neq a^*} u_2(s^*, a, \pi_1) \).

We call this condition “on-path” strict for the receiver because we do not make assumptions about the receiver’s incentives after off-path signals. For generic payoffs, all pure-strategy equilibria will be on-path strict for the receiver.

Theorem 1. Every strategy profile that is patiently stable and on-path strict for the receiver is an RCE.

The proof is very similar to the proof Theorem 3 in Fudenberg and He (2017) and therefore omitted.

6.2 Quasi-Strict uRCE Is Sufficient for Patient Stability

We now prove our main result: as a partial converse to Theorem 1, we show that under additional strictness conditions, every uRCE is path-equivalent to a patiently stable strategy profile.

Definition 17. A quasi-strict uRCE \( \pi^* \) is a uRCE that is on-path strict for the receiver, strict for the sender (that is, every type strictly prefers its equilibrium signal to any other), and satisfies \( u_1(\theta; \pi^*) > u_1(\theta, s', a) \) for all \( \theta \), all off-path signals \( s' \) and all \( a \in \text{BR}(\hat{P}(s'), s') \).

The last condition in the definition of quasi-strictness requires that every best response to \( \hat{P}(s') \) strictly deters every type from deviating to \( s' \), whenever \( s' \) is off-path. Every uRCE satisfies the weaker version of this condition where “strictly deters” is replaced with “weakly deters”.

Theorem 2. If \( \pi^* \) is a quasi-strict uRCE, then it is path-equivalent to a patiently stable strategy profile.

This theorem follows from three lemmas on \( \mathcal{R}_1 \) and \( \mathcal{R}_2 \). Indeed, the theorem remains valid in any modified learning model where \( \mathcal{R}_1 \) and \( \mathcal{R}_2 \) satisfy the conclusions of these lemmas.
6.2.1 $\mathcal{R}_1$ under a confident prior

The first lemma shows that under a suitable prior, the aggregate sender response of the dynamic learning model approximates the sender’s static best response function when applied to certain receiver strategies, namely strategies that are “close” to one inducing a unique optimal signal for each sender type. The precise meaning of “close” that we use treats on and off-path responses differently, so it requires some auxiliary definitions.

**Definition 18.** Let $\pi^*$ be a strategy profile where every type plays a pure strategy and the receiver plays a pure action after each on-path signal. Say $\pi^*$ induces a unique optimal signal for each sender type if

$$u_1(\theta, \pi_1^*(\theta), \pi_2^*) > \max_{s \neq \pi_1^*(\theta)} u_1(\theta, s, \pi_2^*)$$

for every type $\theta$.

Starting with a strategy profile $\pi^*$ that induces a unique optimal signal for each sender type, define for each off-path $s$ in $\pi^*$ the set of receiver actions $\tilde{A}(s) := \{a | u_1(\theta; \pi^*) > u_1(\theta, s, a), \forall \theta\}$ that strictly deter every type from deviation. Because $\pi_2^*$ induces a unique optimal signal, each $\tilde{A}(s)$ must contain at least one element in the support of $\pi_2^*(\cdot|s)$, but could also contain other actions. It is clear that if $\pi_2^*$ were modified off-path by changing each $\pi_2^*(\cdot|s)$ to be an arbitrary mixture over $\tilde{A}(s)$, then the resulting strategy profile would continue to induce (the same) unique optimal signal for each sender type.

Next, to formalize the idea of “approximating” a receiver strategy that induces a unique optimal signal for each type, we introduce some more notation.

For $\pi^*$ that induces a unique optimal signal for each sender type, write $B_2^{\text{on}}(\pi^*, \epsilon)$ for the elements of $\Pi_2^*$ no more than $\epsilon$ away from $\pi_2^*$ at the on-path signals in $\pi_1^*$, that is

$$B_2^{\text{on}}(\pi^*, \epsilon) := \{\pi_2 \in \Pi_2^*: |\pi_2(a|s) - \pi_2^*(a|s)| \leq \epsilon, \forall a, \text{ on-path } s \text{ in } \pi^*\}.$$

Similarly, define $B_2^{\text{off}}(\pi^*, \epsilon)$ as the elements of $\Pi_2^*$ putting no more than $\epsilon$ probability on actions outside of $\tilde{A}(s)$ after each off-path $s$, where $\tilde{A}(s)$ is the set of actions that would deter every type from deviating to $s$, as above.

$$B_2^{\text{off}}(\pi^*, \epsilon) := \{\pi_2 \in \Pi_2^*: \pi_2(\tilde{A}(s)|s) \geq 1 - \epsilon, \forall \text{ off-path } s \text{ in } \pi^*\}.$$

**Lemma 2.** Suppose $\pi^*$ induces a unique optimal signal for each sender type. Then there exists a regular prior $g_1$, some $0 < \epsilon_{\text{off}} < 1$, and a function $\gamma(\delta, \epsilon)$ valued in $(0, 1)$, such that for every $0 < \delta < 1$, $0 < \epsilon < \epsilon_{\text{off}}$, and $\gamma(\delta, \epsilon) < \gamma < 1$, if $\pi_2 \in B_2^{\text{on}}(\pi^*, \epsilon) \cap B_2^{\text{off}}(\pi^*, \epsilon_{\text{off}})$, then $|\mathcal{R}_1^{\text{on}, \text{of}}[\pi_2](s|\theta) - \pi_1^*(s|\theta)| < \epsilon$ for every $\theta$ and $s$.

Note that the same $\epsilon$ appears in the hypothesis $\pi_2 \in B_2^{\text{on}}(\pi^*, \epsilon)$ as in the conclusion. That is, the aggregate sender response gets closer to $\pi_1^*$ as receiver’s play gets closer to $\pi_2^*$. 

19
The idea is to specify a sender prior \( g_1 \) that is highly confident and correct about the receiver’s response to on-path signals, and is also confident that the receiver responds to each off-path signal \( s \) with actions in \( \tilde{A}(s) \). Take a signal \( s' \) other than the one that \( \theta \) sends in \( \pi^*_1 \). If \( \theta \) has not experimented much with \( s' \), then her belief is close to the prior and she thinks deviation does not pay. If \( \theta \) has experimented a lot with \( s' \), then by law of large numbers her belief is likely to be concentrated in \( \tilde{A}(s') \), so again she thinks deviation does not pay. Since option value for experimentation eventually goes to 0, at most histories all sender types are playing a myopic best response to their beliefs, meaning they will not deviate from \( \pi^*_1 \). The intuition is similar to that of Lemmas 6.1 and 6.4 from Fudenberg and Levine (2006), which says that we can construct a highly concentrated and correct prior so that in the steady state, most agents have correct beliefs about opponents’ play both on and one step off the equilibrium path.

This lemma requires the assumption that \( \pi^* \) is strict for the sender. If \( s^* \) were only weakly optimal for \( \theta \) in \( \pi^* \), there could be receiver strategies arbitrarily close to \( \pi^*_2 \) that make some other signal \( s' \neq s^* \) strictly optimal for \( \theta \). In that case, we cannot rule out that a non-negligible fraction of the \( \theta \) population will rationally play \( s' \) forever when the receiver population plays close to \( \pi^*_2 \).

### 6.2.2 \( R_2 \) and learning rational compatibility

Let \( C \) be the set of sender strategies that respect the rational compatibility order, that is

\[
C := \{ \pi_1 \in \Pi^*_1 : \pi_1(s|\theta) \geq \pi_1(s|\theta') \text{ whenever } \theta \succ_s \theta' \}.
\]

Lemma 3 shows that under suitable prior, most receivers’ beliefs are consistent with the rational compatibility order after every signal if they in fact face a sender aggregate play from \( C \).

**Lemma 3.** For each \( \epsilon > 0 \), there exists a regular receiver prior \( g_2 \) and \( 0 < \gamma < 1 \) so that for any \( \gamma < \gamma < 1 \), \( 0 < \delta < 1 \), and \( \pi_1 \in C \),

\[
R_{2,\delta,\gamma}^{g_2}(\pi_1)(BR(\hat{P}(s), s)|s) \geq 1 - \epsilon
\]

for each signal \( s \).

To do this, we construct a Dirichlet prior \( g_2 \) so that whenever \( \theta' \succ_s \theta'' \), \( g_2 \) assigns much greater prior weight to \( \theta' \) playing \( s \) than to \( \theta'' \) doing the same.\(^{10}\) In the absence of data, the

\(^{10}\)The Dirichlet prior is the conjugate prior to multinomial data, and corresponds to the updating used in fictitious play Fudenberg and Kreps (1993). It is readily verified that if each of \( g_1^{(\theta)} \) and \( g_2^{(s)} \) is Dirichlet and independent of the other components, then \( g \) is regular. In the proof, we work with Dirichlet priors since they give tractable closed-form expressions for the posterior mean belief of opponent’s strategy after a given history.
receiver strongly believes that $p(\theta''|s)/p(\theta'|s) \leq \lambda(\theta'')/\lambda(\theta')$. This strong prior belief can only be overturned by a very large number of observations to the contrary. But because $\pi_1 \in C$, if the receiver has a very large number of observations, the law of large numbers implies this large sample is unlikely to lead the receiver to have a belief outside of $\hat{P}(s)$. So, we can ensure that sufficiently long-lived receivers play a best response to $\hat{P}(s)$ after the off-path signal, with high probability.\footnote{Note that this lemma does not follow from Lemma 6.1 of Fudenberg and Levine (2006), because that lemma only guarantees that the receivers’ beliefs about type $\theta$ sending signal $s$ are within $\epsilon$ of the true probability. But when a signal is sent with zero probability in a strategy profile, perturbing the strategy of every type by $\epsilon$ can generate arbitrary beliefs about the sender’s type after a previously off-path signal.}

Finally, we state a lemma that says for any Dirichlet receiver prior, when lifetimes are long enough, aggregate receiver response approximates the receiver’s best response function on-path when applied to a sender strategy that provides strict incentive after every on-path signal. Write $B_{1_{\theta}}^\text{on}(\pi^*, \epsilon)$ for the elements of $\Pi_{1_{\theta}}^\ast$ where each type $\theta$ plays $\epsilon$-close to $\pi_1^\ast(\cdot|\theta)$, that is

$$B_{1_{\theta}}^\text{on}(\pi^*, \epsilon) := \{\pi_1 \in \Pi_{1_{\theta}}^\ast : |\pi_1(s|\theta) - \pi_1^\ast(s|\theta)| \leq \epsilon, \forall \theta, s\}.$$

Lemma 4. Fix a strategy profile $\pi^*$ where receiver has strict incentive after every on-path signal. For each regular Dirichlet receiver prior $g_2$, there exists $\epsilon_1 > 0$ and a function $\gamma(\epsilon)$ valued in $(0,1)$, so that whenever $\pi_1 \in B_{1_{\theta}}^\text{on}(\pi^*, \epsilon_1)$, $0 < \delta < 1$, and $\gamma(\epsilon) < \gamma < 1$, we have $R_{2_{\theta}}^{g_2, \delta, \gamma}(\pi_1)(a|s) - \pi_2^\ast(a|s)| < \epsilon$ for every on-path signal $s$ in $\pi^*$ and $a$.

The intuition is that when the aggregate sender strategy is close to $\pi_1^\ast$, the law of large numbers implies that after each signal that $\pi_1^\ast$ gives positive probability, a receiver with enough data is likely to have a belief close to the Bayesian belief assigned by $\pi_1^\ast$. Coupled with the fact that $\pi_1^\ast$ is on-path strict for the receiver, this lets us conclude that long-lived receivers play $\pi_2^\ast(\cdot|s)$ after every on-path $s$ with high probability.

6.2.3 Proof of Theorem 2

Proof. We will construct a regular prior $g$. We will then show that for every $0 < \delta < 1$, there exists convex and compact sets of strategy profiles $E_j \subseteq \Pi^\ast$ with $E_j \downarrow E_* \subseteq B_{1_{\theta}}^\text{on}(\pi^*, 0) \cap B_{2_{\theta}}^\text{on}(\pi^*, 0)$ and a corresponding sequence of survival probabilities $\gamma_j \to 1$ so that $(R_{1_{\theta}}^{g_2, \delta, \gamma_j}(\pi_2), R_{2_{\theta}}^{g_2, \delta, \gamma_j}(\pi_1)) \in E_j$ whenever $\pi \in E_j$. We proved in Fudenberg and He (2017) that $R_1$ and $R_2$ are continuous maps, so a fixed point theorem implies that for each $j$, some strategy profile in $E_j$ is a steady state profile under parameters $(g, \delta, \gamma_j)$. Any convergent subsequence of these $j$-indexed steady state profiles has a limit in $E_*$, so this limit agrees with $\pi^\ast$ on path. This shows that for every $\delta$ there is a $\delta$-stable strategy profile path-equivalent to $\pi^\ast$, so there is a patiently stable strategy profile with the same property.

Step 1: Constructing $g$ and some thresholds.
Since $\pi^*$ induces a unique optimal signal for each sender type, by Lemma 2 find a regular sender prior $g_1$, $0 < \epsilon_{\text{off}} < 0$, and a function $\gamma_{LM1}(\delta, \epsilon)$.

In Lemma 3, substitute $\epsilon = \epsilon_{\text{off}}$ to find a regular receiver prior $g_2$ and $0 < \gamma_{LM2} < 1$.

Finally, in Lemma 4 let $g_2$ be as constructed above to find $\epsilon_{LM3} > 0$ and a function $\gamma_{LM3}(\epsilon)$.

**Step 2**: Constructing the sets $E_j$.

For each $j$, let $E_j := C \cap B_{1}^{\text{on}}(\pi^*, \frac{\epsilon_{\text{off}} \wedge \epsilon_{LM3}}{j}) \cap B_{2}^{\text{on}}(\pi^*, \frac{\epsilon_{\text{off}} \wedge \epsilon_{LM3}}{j}) \cap B_{2}^{\text{off}}(\pi^*, \epsilon_{\text{off}})$.

That is, $E_j$ is the set of strategy profiles that respect rational compatibility, differ by no more than $\epsilon_{\text{off}}/j$ from $\pi^*$ on path, and differ by no more than $\epsilon_{\text{off}}$ from $\pi^*$ off path. It is clear that each $E_j$ is convex and compact, and that $\lim_{j \to \infty} E_j \subseteq B_{1}^{\text{on}}(\pi^*, 0) \cap B_{2}^{\text{on}}(\pi^*, 0)$ as claimed.

We may find an accompanying sequence of survival probabilities satisfying

$$\gamma_j > \gamma_{LM1}(\delta, \frac{\epsilon_{\text{off}} \wedge \epsilon_{LM3}}{j}) \lor \gamma_{LM2} \lor \gamma_{LM3}(\frac{\epsilon_{\text{off}} \wedge \epsilon_{LM3}}{j})$$

with $\gamma_j \uparrow 1$.

**Step 3**: $R^{g, \delta, \gamma_j}$ maps $E_j$ into itself.

Let some $\pi \in E_j$ be given.

By Lemma 1, $R^{g, \delta, \gamma_j}[\pi_2] \in C$.

By Lemma 3, $R^{g, \delta, \gamma_j}[\pi_1] \in B_{2}^{\text{off}}(\pi^*, \epsilon_{\text{off}})$, because uniformity of $\pi^*$ means $BR(\tilde{P}(s), s) \subseteq \tilde{A}(s)$ for each off-path $s$.

By Lemma 4, $R^{g, \delta, \gamma_j}[\pi_1] \in B_{2}^{\text{on}}(\pi^*, \frac{\epsilon_{\text{off}} \wedge \epsilon_{LM3}}{j})$.

Finally, from Lemma 2 and the fact that $\pi_2 \in B_{2}^{\text{on}}(\pi^*, \frac{\epsilon_{\text{off}} \wedge \epsilon_{LM3}}{j}) \cap B_{2}^{\text{off}}(\pi^*, \epsilon_{\text{off}})$, we have $R^{g, \delta, \gamma_j}[\pi_2] \in B_{1}^{\text{on}}(\pi^*, \frac{\epsilon_{\text{off}} \wedge \epsilon_{LM3}}{j})$. \qed

## 7 Patient Stability in Generalized Signalling Games

In this section, we discuss the implications of patient stability in two classes of generalized signalling games. For the ease of notation, we will consider a learning model where agents’ priors put full support on opponents’ behavior strategies (as in Fudenberg and He (2017)).

### 7.1 Endogenous Signalling Games

In and Wright (2017) study a class of endogenous signalling games, where the sender chooses her own type before sending a signal. There are two obvious game trees for this situations: one where the sender first chooses her type and then her signal, another where the sender chooses signal first and type second. The authors note that subgame perfection can make different
predictions in the two game trees, and call the equilibrium outcomes which are subgame perfect in both game trees “reordering invariant.”

We note that in our steady-state learning model, both orderings lead to equivalent learning problems for the senders. Provided we hold constant senders’ prior belief as to how receivers react to various signals and receivers’ prior belief as to sender’s strategy (i.e. their choice of (type, signal) pair), the set of patiently stable outcomes does not depend on reordering.

7.2 Signalling Games with Multiple Receivers

Now consider a signalling game with one sender and $R \geq 1$ receivers as in Farrell and Gibbons (1989) and Goltsman and Pavlov (2011). Index the receivers’ utility functions by $u_{2,1}, u_{2,2}, ..., u_{2,R}$. All of the receivers observe the sender’s signal, then simultaneously choose an action from their action set $A_j$ for $1 \leq j \leq R$. Each player’s payoff depends on the $(R + 2)$-tuple $(\theta, s, a_1, ..., a_R)$.

We now define the relevant refinement concept for signalling games with multiple receivers.

**Definition 19.** In a signalling game with multiple receivers, a *compatible perfect Bayesian equilibrium with heterogeneous off-path beliefs* is a strategy profile $(\pi^*_1, (\pi^*_2, j = 1, R))$ such that:

- For each $\theta \in \Theta$, $u_1(\theta; \pi^*) = \max_{s \in S} u_1(\theta, s, (\pi^*_2, j = 1, R))$.

- For each on-path signal $s$ and each $1 \leq j \leq R$, $u_{2,j}(p^*(\cdot|s), s, \pi^*_2, j = 1, R) = \max_{\hat{a}_j \in A_j} u_2(p^*(\cdot|s), s, \hat{a}_j, \pi^*_2, j = 1, R)$, where $p^*(\cdot|s)$ is the Bayesian posterior belief about sender’s type after signal $s$, under strategy $\pi^*_1$.

- For each off-path signal $s$ such that there is a type for which $s$ is not equilibrium dominated, and each $a_j \in A_j$ with $\pi^*_2, j = 1, R(a_j|s) > 0$, there exists a belief $p \in P(s, \pi^*)$ such that $u_{2,j}(p, s, a_j, \pi^*_2, j = 1, R) = \max_{\hat{a}_j \in A_j} u_2(p, s, \hat{a}_j, \pi^*_2, j = 1, R)$, where $P(s, \pi^*)$ is the set of admissible beliefs defined in Fudenberg and He (2017).

To interpret, every type is best responding to the receivers’ play. After every on-path signal, the receivers are playing a Bayesian Nash equilibrium, where their common belief about the sender’s type is the Bayesian posterior belief induced by $\pi^*_1$. After an off-path signal that is not equilibrium dominated for all types, the receivers play a kind of generalized Bayesian Nash equilibrium, where each action in the support of each receiver’s play corresponds to one sub-population. Each sub-population best responds to other receivers’ play, but with possibly different beliefs about the sender’s type. In particular, two sub-populations of the same receiver role may hold different beliefs: all such beliefs are constrained to fall in $P(s, \pi^*)$. Note that
after an off-path signal that is equilibrium dominated for every type, receivers’ play need not resemble Bayesian Nash equilibrium — some receivers could play actions that do not best respond to other receivers’ play under any belief about the sender’s type.

**Theorem 3.** Every patiently stable strategy profile is a compatible perfect Bayesian equilibrium with heterogeneous off-path beliefs.

**Proof.** The first two bullet points in Definition 19 hold for the same reason as Proposition 6.

To show that the last bullet point holds, suppose $\pi^*$ is patiently stable. The condition is vacuously satisfied if every off-path signal is equilibrium dominated for every type. So suppose there is an off-path signal $s'$ and some $a_j$ in the support of $\pi^*_{2,j}(\cdot|s')$ such that $u_{2,j}(p, s', a_j, \pi^*_{2,j}(\cdot|s')) = \max_{\hat{a}_j \in A_j} u_{2,j}(p, s', \hat{a}_j, \pi^*_{2,j}(\cdot|s'))$ does not hold for any $p \in P(s', \pi^*)$. Then there exists $\epsilon > 0$ so that

$$u_{2,j}(p, s', a_j, \pi^*_{2,j}(\cdot|s')) = \max_{\hat{a}_j \in A_j} u_{2,j}(p, s', \hat{a}_j, \pi^*_{2,j}(\cdot|s'))$$

does not hold for any $p$ no more than $\epsilon$ distance away from $P(s', \pi^*)$ and any $\pi_{2,j}(\cdot|s')$ no more than $\epsilon$ away from $\pi^*_{2,j}(\cdot|s')$.

Let $h > 0$ be given. We show that $\pi^*_{2,j}(a_j|s') < 5h$. Find some $N$ (depending on the prior $g$) such that after $N$ samples of $-j$’s play after $s'$ from any aggregate play $\pi_{2,j}(\cdot|s')$, $j$’s mean belief about $-j$’s play after $s'$ falls in an $\epsilon/2$-ball around $\pi_{2,j}(\cdot|s')$ with probability greater than $1 - h$.

Lemma 3 from Fudenberg and He (2017) continue to apply. So, for any $C > 0$, there exist $\delta_k \uparrow 1$, $\gamma_k \uparrow 1$, $\pi^{(k)} \in \Pi^*(g, \delta_k, \gamma_k)$ for each $k$, such that (1) $\pi^{(k)} \rightarrow \pi^*$; (2) $\pi^{(k)}_{2,j}(\cdot|s')$ less than $\epsilon/2$ away from $\pi^*_{2,j}(\cdot|s')$ for every $k$; (3) there is some $\theta$ such that $\pi^*_{1,j}(s'|\theta) > \frac{CN}{\lambda(\theta)h}(1 - \gamma_k)$ for each $k$. In steady state $k$, receiver $j$ aged $h/(1 - \gamma_k)$ observes on average $CN$ instances of the plays of other receivers after $s'$. We may choose $C$ large enough so that the probability that an age $h/(1 - \gamma_k)$ receiver $j$ has observed fewer than $N$ instances of signal $s'$ is below $h$. As the fraction of receivers aged $h/(1 - \gamma_k)$ or older is no smaller than $1 - 2h$, we conclude that no more than $4h$ of the receiver $j$’s have a belief about other receivers’ play that’s further than $\epsilon$ away from $\pi^*_{2,j}(\cdot|s')$. At the same time, by combining Lemmas 1, 2, and 3 from Fudenberg and He (2017), we see that for large enough $k$, more than $1 - h$ fraction of the receiver $j$’s will have a belief about sender’s type that’s within $\epsilon$ of $P(s', \pi^*)$. This means at least $1 - 5h$ of the receiver $j$’s do not play $a_j$ after $s'$, for large enough $k$. \hfill \Box

**References**


A Appendix – Relegated Proofs

A.1 Proof of Lemma 2

Lemma 2: Suppose $\pi^*$ induces a unique optimal signal for each sender type. Then there exists a regular prior $g_1$, some $0 < \epsilon_{\text{off}} < 1$, and a function $\gamma(\delta, \epsilon)$ valued in $(0, 1)$, such that for
every $0 < \delta < 1$, $0 < \epsilon < \epsilon_{\text{off}}$, and $\gamma(\delta, \epsilon) < \gamma < 1$, if $\pi_2 \in B_2^{\text{on}}(\pi^*, \epsilon) \cap B_2^{\text{off}}(\pi^*, \epsilon_{\text{off}})$, then $|\mathcal{R}_1^{\gamma_1, \delta, \gamma}(\pi_2)(s|\theta) - \pi^*_2(s|\theta)| < \epsilon$ for every $\theta$ and $s$.

**Proof.** Here are three lemmas from Fudenberg and Levine (2006):

**FL06 Lemma A.1:** Suppose $\{X_k\}$ is a sequence of i.i.d. Bernoulli random variables with $\mathbb{E}[X_k] = \mu$, and define for each $n$ the random variable

$$S_n := \left\lfloor \frac{\sum_{k=1}^{n}(X_k - \mu)}{n} \right\rfloor.$$  

Then for any $n, \bar{n} \in \mathbb{N}$,

$$\mathbb{P} \left[ \max_{2 \leq n \leq \bar{n}} S_n > \epsilon \right] \leq \frac{2}{3} \cdot \frac{1}{n} \cdot \frac{\mu}{\epsilon^4}.$$  

**FL06 Lemma A.2:** For all $\epsilon, \epsilon' > 0$, there is an $N > 0$ so that for all $\delta, \gamma, g, \pi$, signal $s$ and action $a \in A$,

$$\psi_{\theta}^{\pi_2(g, \delta, \gamma)} \{y_\theta : |\hat{\pi}_2(a|s; y_\theta) - \pi_2(a|s)| > \epsilon, \#(s|y_\theta) > N \} < \epsilon'.$$

(Here, $\hat{\pi}_2(a|s; y_\theta)$ is the empirical frequency of receiver playing $a$ after signal $m$ in history $y_\theta$, that is to say $\hat{\pi}_2(a|s; y_\theta) = \#((a, s), y_\theta)/\#(s, y_\theta).$)

**FL06 Lemma A.4:** For all $\epsilon, \epsilon' > 0$ and $\delta < 1$, there exists $N$ such that for all $\pi$, $g$, and $\gamma$, we get

$$\psi_{\theta}^{\pi_2(g, \delta, \gamma)} \{y_\theta \notin Y_\theta(\epsilon), \#(\sigma_\theta(y_\theta), y_\theta) > N \} \leq \epsilon'$$

where $Y_\theta(\epsilon) \subseteq Y_\theta$ are those histories $y_\theta$ where

$$\max_{s \in S} u_1(\theta, s|y_\theta) \leq u_1(\sigma_\theta(y_\theta)|y_\theta) + \epsilon$$

that is type $\theta$ is playing a myopic $\epsilon$ best response according to posterior belief after history $y_\theta$.

Now we proceed with our argument.

Since $\pi^*$ is strict on-path , there exist $\xi_1, \xi_2 > 0$ such that whenever $\pi_2$ satisfies $|\pi_2(a|s) - \pi^*_2(a|s)| \leq \xi_1$ for every on-path $s$ and action $a$, while for every off-path $s$ we have $\pi_2(\bar{A}(s)|s) \geq 1 - \xi_1$, then for each type $\theta$ we get

$$u_1(\theta, \pi^*_1(\theta), \pi_2) > \xi_2 + \max_{s \neq \pi^*_1(\theta)} u_1(\theta, s, \pi_R).$$

That is, if receiver plays $\xi_1$-close to $\pi^*$ on-path and $\xi_1$-close to $\bar{A}(s)$ off-path, then for every type of sender, playing the prescribed equilibrium signal is strictly better than any other signal by at least $\xi_2 > 0$.

Following Fudenberg and Levine (2006), consider a prior $g_1$ such that whenever sender has fewer than $\bar{n} := 2^{11}/\xi_1^4$ observations of playing signal $s$, her belief as to receiver’s probability
of taking action \(a\) after signal \(s\) differs from \(\pi_1^*(a|s)\) by no more than \(\xi_1\) if \(s\) is on-path, while her belief as to the probability that receiver strategy assigns to \(\tilde{A}(s)\) is at least \(1 - \xi\) if \(s\) is off-path. Also, let \(\epsilon_{\text{off}} := \xi_1/2\).

Now let \(\delta \in (0, 1)\) and \(0 < \epsilon < \epsilon_{\text{off}}\) be given. We construct \(\gamma(\delta, \epsilon)\) satisfying the conclusion of the lemma.

To do this, in FL06 Lemma A.4 put \(\epsilon = \xi_2\) and \(\epsilon' = \epsilon/6\), to obtain a \(N_1(\epsilon)\). Next, in FL06 Lemma A.2 put \(\epsilon = \xi_1/2, \epsilon' = \epsilon/6\), to obtain \(N_2(\epsilon)\). Let \(N(\epsilon) := N_1(\epsilon) \lor N_2(\epsilon)\). There are 5 classes of exceptional histories for type \(\theta\) that can lead to playing some signal \(\hat{s}\) other than the one prescribed by the equilibrium strategy, \(s^* := \pi_1^*(\theta)\).

**Exception 1:** \(\theta\) has played \(\hat{s}\) fewer than \(N(\epsilon)\) times before, that is \(\sigma_\theta(y_\theta) = \hat{s}\) but \(\#(\hat{s}, y_\theta) < N(\epsilon)\). Such histories can be made to have mass no larger than \(\epsilon/6\) by taking \(\gamma(\delta, \epsilon)\) large enough.

**Exception 2:** \(y_\theta\) is in the exceptional set described in FL06 Lemma A.4. But by choice of \(N(\epsilon) \geq N_1(\epsilon)\), we know that

\[
\psi_{\theta}^{(\pi_2(\cdot), \delta, \gamma)} \{y_\theta \notin Y_\theta(\xi_2), \#(\sigma_\theta(y_\theta), y_\theta) > N(\epsilon)\} \leq \epsilon/6.
\]

**Exception 3:** \(\theta\) has played \(\hat{s}\) more than \(N(\epsilon)\) times, but has a misleading sample. By FL93 Lemma A.2,

\[
\psi_{\theta}^{(\pi_2(\cdot), \delta, \gamma)} \{y_\theta : |\hat{\pi}_2(a|\hat{s}; y_\theta) - \pi_2(a|\hat{s})| > \xi_1/2, \#(\hat{s}|y_\theta) > N(\epsilon)\} < \epsilon/6.
\]

Since we have chosen \(\pi \in B^\text{on}_2(\pi^*, \epsilon) \cap B^\text{off}_2(\pi^*, \epsilon_{\text{off}})\), we know \(\pi_2\) differs from \(\pi_2^*\) by no more than \(\epsilon_{\text{off}} = \xi_1/2\) after every on-path signal, and puts no more weight than \(\xi_1/2\) on actions not in \(\tilde{A}(s)\) after off-path signal \(s\). So in particular,

\[
\psi_{\theta}^{(\pi_2(\cdot), \delta, \gamma)} \left\{y_\theta : \begin{array}{l}
|\hat{\pi}_2(a|\hat{s}; y_\theta) - \pi_2^*(a|\hat{s})| > \xi_1 \text{ if } \hat{s} \text{ on-path, or } \\
\hat{\pi}_2(\tilde{A}(\hat{s})|\hat{s}) < 1 - \xi_1 \text{ if } \hat{s} \text{ off-path } \\
\#(\hat{s}|y_\theta) > N(\epsilon)
\end{array} \right\} < \epsilon/6.
\]

**Exception 4:** \(\theta\) has played the equilibrium signal \(s^*\) more than \(N(\epsilon)\) times, but has a misleading sample. As before, we get

\[
\psi_{\theta}^{(\pi_2(\cdot), \delta, \gamma)} \{y_\theta : |\hat{\pi}_2(a|s^*; y_\theta) - \pi_2^*(a|s^*)| > \xi_1, \#(s^*|y_\theta) > N(\epsilon)\} < \epsilon/6.
\]

**Exception 5:** \(\theta\) has played the equilibrium signal \(s^*\) between \(n\) and \(N(\epsilon)\) times, but has a misleading sample. Let \(X_k \in \{0, 1\}\) denote whether \(\theta\) sees the equilibrium response \(\pi_2^*(s^*)\) the \(k\)-th time she plays \(s^*\) \((X_k = 0)\) or whether she sees instead a different response \((X_k = 1)\). As in FL06 Lemma A.1, define

\[
S_n := \frac{|\sum_{k=1}^{n}(X_k - \mu)|}{n}
\]
where $\mu = 1 - \pi_2(\pi_2^*(s^*)|s^*) < \epsilon $ since $s^*$ is an on-path signal in $\pi^*$.

The probability that the fraction of responses other than $\pi_1^*(s^*)$ exceeds $\xi_1$ between then-th time and $N(\epsilon)$-th time that $\theta$ plays $s^*$ is bounded above by FL06 Lemma A.1,

$$\mathbb{P} \left[ \max_{\theta \leq n \leq N(\epsilon)} S_n > \xi_1/2 \right] \leq \frac{2^7}{3} \cdot \frac{1}{n} \cdot \frac{\mu}{(\xi_1/2)^4} \leq \frac{1}{3} \cdot \mu \text{ (by choice of } n) \leq \epsilon_1/3.$$

Finally, at a history $y_\theta$ that does not belong to those exceptions, we must have $\sigma_\theta(y_\theta) = m^*$.

This is because $y_\theta$ is not in exception 1, so $\theta$ has played $\sigma_\theta(y_\theta)$ at least $N(\epsilon)$ times before, and it is not in exception 2, so $\sigma_\theta(y_\theta)$ is a $\xi_2$ myopic best response to current beliefs. Yet the empirical frequency for response after signal $\sigma_\theta(y_\theta)$ is no more than $\xi_1$ away from $\pi_2^*(\sigma_\theta(y_\theta))$ as $y_\theta$ is not in exception 3. Since the prior is Dirichlet and also has this property, this means the current posterior belief about response after signal $\sigma_\theta(y_\theta)$ also has this property. If $\#(s^*, y_\theta) > n$, then $y_\theta$ not being in exceptions 4 or 5 implies belief as to response after signal $s^*$ is also no more than $\xi_1$ away from $\pi_2^*(s^*)$, while if $\#(s^*, y_\theta) < n$ then choice of prior implies the same. In short, beliefs on both responses after $s^*$ and responses after $\sigma_\theta(y_\theta)$ are no more than $\xi_1$ away from their $\pi_2^*$ counterparts. But in that case, no signal other than $s^*$ can be an $\xi_2$ best response.  

### A.2 Proof of Lemma 3

**Lemma 3:** For each $\epsilon > 0$, there exists a regular receiver prior $g_2$ and $0 < \underline{\gamma} < 1$ so that for any $\underline{\gamma} < \gamma < 1$, $0 < \delta < 1$, and $\pi_1 \in C$,

$$R_{g_2, \delta, \gamma}[\pi_1](BR(\hat{P}(s), s)|s) \geq 1 - \epsilon$$

for each signal $s$.

**Proof.** For each $\xi > 0$, consider the approximation to $P_{\theta' \triangleright \theta''}$,

$$P_{\theta' \triangleright \theta''}^\xi := \left\{ p \in \Delta(\Theta) : \frac{p(\theta'')}{p(\theta')} \leq (1 + \xi) \frac{\lambda(\theta'')}{\lambda(\theta')} \right\}$$

and hence the approximation to $\hat{P}(s)$,

$$\hat{P}_\xi(s) := \Delta(\Theta_{s'}) \cap \left\{ P_{\theta' \triangleright \theta''}^\xi : \theta' \triangleright s', \theta'' \right\}.$$ 

Since the BR correspondence has a closed graph, there is an $\xi > 0$ such that $BR(\hat{P}_\xi(s), s) = BR(\hat{P}(s), s)$. 

28
Take some such $\xi$. Next we will choose a series of constants.

- Pick $0 < h < 1$ such that $\frac{1-h}{1+h} > (1 - \xi)^{1/3}$.
- Pick $G > 0$ such that for every $\theta \in \Theta$, $1/(h^2 \cdot G \cdot (1-h) \cdot \lambda(\theta)) < \epsilon/(4 \cdot |S| \cdot |\Theta|^2)$.
- For each $\theta$, construct a Dirichlet prior on $S_\theta$ with parameters $\alpha(\theta, s) \geq 0$. Pick Dirichlet prior parameters $\alpha(\theta, s) \geq 0$ so that whenever $\theta \succ_s \theta'$, we have
  \[
  \alpha(\theta, s) - \alpha(\theta', s) > (\sqrt{4 \cdot |S| \cdot |\Theta|^2}/\epsilon + 1) \cdot G.
  \] (3)
  In the event that $\theta \succ_s \theta'$ and $\theta' \succ_s \theta$, put $\alpha(\theta, s) = \alpha(\theta', s)$.
- Pick $N \in \mathbb{N}$ so that for any $N > N$, $\theta, \theta' \in \Theta$, we have
  \[
  \mathbb{P}[(1-h) \cdot N \cdot \lambda(\theta) \leq \text{Binom}(N, \lambda(\theta)) \leq (1+h) \cdot N \cdot \lambda(\theta)] > 1 - \frac{\epsilon}{4 \cdot |\Theta|}
  \]
  and
  \[
  \frac{(1-h) \cdot N \cdot \lambda(\theta')}{(1+h) \cdot N \cdot \lambda(\theta) + \max_{\theta} \sum_{s \in S} \alpha(\theta, s)} > (1 - \xi)^{1/3} \frac{\lambda(\theta')}{\lambda(\theta)}
  \]
- Pick $\gamma \in (0, 1)$ such that $1 - (\gamma)^{N+1} < \epsilon/4$.

Suppose the receiver’s prior over the strategy of type $\theta$ is Dirichlet with parameters $(\alpha(\theta, s))_{s \in S}$. We claim that the conclusion of the lemma holds.

Fix some strategy $\pi_1 \in C$. Write $\#(\theta | y_2)$ for the number of times the sender has been of type $\theta$ in history $y_2$, while $\#(\theta, s | y_2)$ counts the number of times type $\theta$ has sent signal $s$ in history $y_2$. Put $\psi_2 = \psi_{2,2}^{\pi_1(g, d, \gamma)}$ and write $E \subseteq Y_2$ for those receiver histories with length at least $N$ satisfying
  \[
  (1-h) \cdot N \cdot \lambda(\theta) \leq \#(\theta | y_2) \leq (1+h) \cdot N \cdot \lambda(\theta)
  \]
for every $\theta \in \Theta$. By the choice of $N$ and $\gamma$, whenever $\gamma > \gamma$ we have $\psi(E) \geq 1 - \epsilon/2$. We now show that given $E$, the conditional probability that the receiver’s posterior belief after every off-equilibrium signal $s$ lies in $\hat{P}_\xi(s)$ is at least $1 - \epsilon/2$. To do this, fix signal $s$ and two types with $\theta \succ_s \theta'$.

If $s$ is strictly dominated for both $\theta$ and $\theta'$, then according to the receivers’ Dirichlet prior, $\theta$ and $\theta'$ each sends $s$ with zero probability. Since $\pi \in \Pi^*_1$, we have $\pi_1(s | \theta) = \pi_1(s | \theta') = 0$. So after every positive-probability history, receiver’s belief falls in $\hat{P}_\xi(s)$ as it puts zero probability on the $s$-sender being $\theta$ or $\theta'$. Henceforth we only consider the case where $s$ is not strictly dominated for both.
After history $y_2$, the receiver’s updated posterior likelihood ratio for types $\theta$ and $\theta'$ upon seeing signal $s$ is
\[
\frac{\lambda(\theta)}{\lambda(\theta')} \cdot \frac{\left( \alpha(\theta, s) + \#(\theta, s|y_2) \right) \left( \alpha(\theta', s) + \#(\theta', s|y_2) \right)}{\left( \#(\theta|y_2) + \sum_{s \in S} \alpha(\theta, s) \right) \left( \#(\theta'|y_2) + \sum_{s \in S} \alpha(\theta', s) \right)}
\]
\[
= \frac{\lambda(\theta)}{\lambda(\theta')} \cdot \frac{\alpha(\theta, s) + \#(\theta, s|y_2)}{\alpha(\theta', s) + \#(\theta', s|y_2)} \cdot \frac{\#(\theta'|y_2)}{\#(\theta|y_2)} + \sum_{s \in S} \alpha(\theta', s).
\]
Since we have $\#(\theta'|y_2) \geq (1 - h) \cdot N \cdot \lambda(\theta')$ while $\#(\theta|y_2) \leq (1 + h) \cdot N \cdot \lambda(\theta)$, we get
\[
\frac{\#(\theta'|y_2)}{\#(\theta|y_2)} + \sum_{s \in S} \alpha(\theta', s) \geq \frac{(1 - h) \cdot N \cdot \lambda(\theta')}{(1 + h) \cdot N \cdot \lambda(\theta) + \sum_{s \in S} \alpha(\theta, s)} > (1 - \xi)^{1/3} \cdot \frac{\lambda(\theta')}{\lambda(\theta)}.
\]
If $s$ is strictly dominant for both $\theta$ and $\theta'$, then $\pi_1 \in \Pi^*_1$ means that $\pi_1(s|\theta) = \pi_1(s|\theta') = 1$. In this case, $\#(\theta, y_2|y_2) = \#(\theta|y_2)$ and $\#(\theta', s|y_2) = \#(\theta'|y_2)$. Since $\#(\theta|y_2) \geq (1 - h) \cdot N \cdot \lambda(\theta)$, $\#(\theta'|y_2) \leq (1 + h) \cdot N \cdot \lambda(\theta')$, we have:
\[
\frac{\alpha(\theta, s) + \#(\theta, s|y_2)}{\alpha(\theta', s) + \#(\theta', s|y_2)} \geq \frac{(1 - h) \cdot N \cdot \lambda(\theta)}{\sum_{s \in S} \alpha(\theta', s) + (1 + h) \cdot N \cdot \lambda(\theta')} \geq (1 - \xi)^{1/3} \frac{\lambda(\theta)}{\lambda(\theta')}.
\]
This shows the product is no smaller than $(1 - \xi)^{2/3} \frac{\lambda(\theta)}{\lambda(\theta')}$, so receiver believes in $P^x_{\theta_0, \theta'}$ after every history in $E$.

Now we analyze the term $\frac{\alpha(\theta, s) + \#(\theta, s|y_2)}{\alpha(\theta', s) + \#(\theta', s|y_2)}$ for the case where $s$ is not strictly dominant for both $\theta$ and $\theta'$. We consider two cases, depending on whether $N$ is “large enough” so that the compatible type $\theta$ experiments enough on average in a receiver history of length $N$ under sender strategy $\pi_1$.

**Case A:** $\pi_1(s|\theta) \cdot N < G$. In this case, since $\pi \in C$ and $\theta \succ s \theta'$, we must also have $\pi_1(s|\theta') \cdot N < G$. Then $\#(\theta', s|y_2)$ is distributed as a binomial random variable with mean smaller than $G$, hence standard deviation smaller than $\sqrt{G}$. By Chebyshev’s inequality, the probability that it exceeds $(\sqrt{(4 \cdot |S| \cdot |\Theta|^2)/\epsilon} + 1) \cdot G$ is no larger than
\[
\frac{1}{G \cdot (4 \cdot |S| \cdot |\Theta|^2)/\epsilon} \leq \frac{\epsilon}{4|S| \cdot |\Theta|^2}.
\]
But in any history $y_2$ where $\#(\theta', s|y_2)$ does not exceed this number, we would have
\[
\alpha(\theta', s) + \#(\theta', s|y_2) \leq \alpha(\theta, s) \leq \alpha(\theta, s) + \#(\theta, s|y_2)
\]
by choice of the difference between prior parameters $\alpha(\theta', s)$ and $\alpha(\theta, s)$. Therefore $\frac{\alpha(\theta, s) + \#(\theta, s|y_2)}{\alpha(\theta', s) + \#(\theta', s|y_2)} \geq 1$. In summary, under Case A, there is probability no smaller than $1 - \frac{\epsilon}{4|S| \cdot |\Theta|^2}$ that $\frac{\alpha(\theta, s) + \#(\theta, s|y_2)}{\alpha(\theta', s) + \#(\theta', s|y_2)} \geq 1$. 

30
Case B: \( \pi_1(s|\theta) \cdot N \geq G \). In this case, we can bound the probability that

\[
\#(\theta, s|y_2)/\#(\theta', s|y_2) \leq \frac{\lambda(\theta)}{\lambda(\theta') \cdot (1 - h)} \cdot \left(1 + \frac{h}{1 + h}\right)^2.
\]

Let \( p := \pi_1(s|\theta) \). Given that \( \#(\theta|y_2) \geq (1 - h) \cdot N \cdot \lambda(\theta) \), the distribution of \( \#(\theta, s|y_2) \) first order stochastically dominates \( \text{Binom}(1 - h) \cdot N \cdot \lambda(\theta), p) \).

On the other hand, given that \( \#(\theta|y_2) \leq (1 + h) \cdot N \cdot \lambda(\theta') \) and furthermore \( \pi_1(s|\theta') \leq \pi_1(s|\theta) = p \), the distribution of \( \#(\theta', s|y_1) \) is first order stochastically dominated by \( \text{Binom}(1 + h) \cdot N \cdot \lambda(\theta'), p) \).

The first distribution has mean \((1 - h) \cdot N \cdot \lambda(\theta) \cdot p \) with standard deviation no larger than \( \sqrt{(1 - h) \cdot N \cdot \lambda(\theta)} \cdot p \). Thus

\[
\mathbb{P}[\text{Binom}((1 - h) \cdot N \cdot \lambda(\theta), p) < (1 - h) \cdot (1 - h) \cdot N \cdot \lambda(\theta) \cdot p] < 1/(h \cdot \sqrt{p(1 - h)N\lambda(\theta)})^2 \leq 1/(h \cdot \sqrt{G \cdot (1 - h) \cdot \lambda(\theta)})^2 < \epsilon / (4 \cdot |S| \cdot |\Theta|^2)
\]

where we used the fact that \( pN \geq G \) in the second-to-last inequality, while the choice of \( G \) ensured the final inequality.

At the same time, the second distribution has mean \((1 + h) \cdot N \cdot \lambda(\theta') \cdot p \) with standard deviation no larger than \( \sqrt{(1 + h) \cdot N \cdot \lambda(\theta')} \cdot p \), so

\[
\mathbb{P}[\text{Binom}((1 + h) \cdot N \cdot \lambda(\theta'), p) > (1 + h) \cdot (1 + h) \cdot N \cdot \lambda(\theta') \cdot p] < 1/(h \cdot \sqrt{p(1 + h)N\lambda(\theta')})^2 \leq 1/(h \cdot \sqrt{G \cdot (1 + h) \cdot \lambda(\theta')})^2 < \epsilon / (4 \cdot |S| \cdot |\Theta|^2)
\]

by the same arguments. Combining the bounds on these two binomial random variables,

\[
\mathbb{P}\left[ \frac{\text{Binom}((1 - h) \cdot N \cdot \lambda(\theta), p)}{\text{Binom}((1 + h) \cdot N \cdot \lambda(\theta'), p)} \leq \frac{\lambda(\theta)}{\lambda(\theta') \cdot (1 - h)} \cdot (1 + \frac{h}{1 + h})^2 \right] < \epsilon / (2 \cdot |S| \cdot |\Theta|^2).
\]

Via stochastic dominance, this shows a fortiori

\[
\mathbb{P}\left[ \#(\theta, s|y_2)/\#(\theta', s|y_2) \leq \frac{\lambda(\theta)}{\lambda(\theta') \cdot (1 - h)} \cdot (1 - h) \cdot (1 + \frac{h}{1 + h})^2 \right] < \epsilon / (2 \cdot |S| \cdot |\Theta|^2).
\]

Therefore, for any \( s, \theta, \theta' \) such that \( \theta \succ_s \theta' \),

\[
\psi\left(y_2 : \frac{\alpha(\theta, s) + \#(\theta, s|y_2)}{\alpha(\theta', s) + \#(\theta', s|y_2)} \geq \frac{\lambda(\theta)}{\lambda(\theta') \cdot (1 - h)} \cdot (1 + \frac{h}{1 + h})^2 \right) \geq 1 - \epsilon / (2 \cdot |S| \cdot |\Theta|^2).
\]

This concludes case B.

In either case, at a history \( y_2 \) with \((1 - h) \cdot N \cdot \lambda(\theta) \leq \#(\theta|y_2) \leq (1 + h) \cdot N \cdot \lambda(\theta) \) for every
\( \theta \), for every pair \( \theta, \theta' \) such that \( \theta \succ_s \theta' \), we get
\[
\frac{\alpha(\theta, s) + \#(\theta, s|y_2)}{\alpha(\theta', s) + \#(\theta', s|y_2)} \geq \frac{\lambda(\theta)}{\lambda(\theta')} \cdot \frac{(1 - h)(1 + h)^2}{1 + h} \]
with probability at least \( 1 - \epsilon/(2 \cdot |S| \cdot |\Theta|^2) \).

But at any history \( y_2 \) where this happens, the receiver’s posterior likelihood ratio for types \( \theta \) and \( \theta' \) after signal \( s \) satisfies
\[
\frac{\lambda(\theta)}{\lambda(\theta')} \cdot \frac{\alpha(\theta, s) + \#(\theta, s|y_2)}{\alpha(\theta', s) + \#(\theta', s|y_2)} \cdot \frac{\#(\theta'|y_2) + \sum_{s \in S} \alpha(\theta', s)}{\#(\theta|y_2) + \sum_{s \in S} \alpha(\theta, s)} \geq \frac{\lambda(\theta)}{\lambda(\theta')} \cdot (1 - \xi)^2/3 \cdot (1 - \xi)^1/3 \geq \frac{\lambda(\theta)}{\lambda(\theta')} \cdot (1 - \xi).
\]

As there are at most \( |\Theta|^2 \) such pairs for each signal \( s \) and \( |S| \) total signals,
\[
\psi \left( y_2 : \frac{\lambda(\theta)}{\lambda(\theta')} \cdot \frac{\alpha(\theta, s) + \#(\theta, s|y_2)}{\alpha(\theta', s) + \#(\theta', s|y_2)} \cdot \frac{\#(\theta'|y_2) + \sum_{s \in S} \alpha(\theta', s)}{\#(\theta|y_2) + \sum_{s \in S} \alpha(\theta, s)} \right) \geq \frac{\lambda(\theta)}{\lambda(\theta')} \cdot (1 - \xi) \forall s, \theta \succ_s \theta' \mid E \right) \geq 1 - \epsilon/2
\]
as claimed. As the event \( E \) has \( \psi \)-probability no smaller than \( 1 - \epsilon/2 \), there is \( \psi \) probability at least \( 1 - \epsilon \) that receiver’s posterior belief is in \( \hat{P}_\xi(s) \) after every off-path \( s \).

\[ \square \]

### A.3 Proof of Lemma 4

**Lemma 4:** Fix a strategy profile \( \pi^* \) where the receiver has strict incentive after every on-path signal. For each regular Dirichlet receiver prior \( g_2 \), there exists \( \epsilon_1 > 0 \) and a function \( \gamma(\epsilon) \) valued in \((0,1)\), so that whenever \( \pi_1 \in B^{\text{on}}_1(\pi^*, \epsilon_1) \), \( \delta \in [0,1) \), and \( \gamma(\epsilon) < \gamma < 1 \), we have \( \mathcal{R}_2^{g_2, \delta, \gamma}[\pi_1](a|s) - \pi^*_2(a|s) < \epsilon \) for every on-path signal \( s \) in \( \pi^* \) and \( a \).

**Proof.** Since \( \pi^* \) is on-path strict for the receiver, there exists some \( \xi > 0 \) such that for every on-path signal \( s \) and every belief \( p \in \Delta(\Theta) \) with
\[
|p(\theta) - p(\theta; s, \pi^*)| < \xi, \forall \theta \in \Theta \tag{4}
\]
(where \( p(\cdot; s, \pi^*) \) is the Bayesian belief after on-path signal \( s \) induced by the equilibrium \( \pi^* \)), we have \( \text{BR}(p, s) = \{ \pi^*_2(s) \} \). For each \( s \), we show that there is a large enough \( N(s, \epsilon) \) and small enough \( \zeta(s) \) so that when receiver observes history \( y_2 \) generated by any \( \pi \in B_{\text{on}}(\pi^*, \epsilon') \) with \( \epsilon' < \zeta(s)/4 \) and length least \( N(s, \epsilon) \), there is probability at least \( 1 - \epsilon/2 \) that receiver’s posterior belief satisfies (4). Hence, conditional on having a history length of at least \( N(s, \epsilon) \), there is \( 1 - \epsilon/2 \) chance that receiver will play as in \( \pi^*_2 \) after \( s \). By taking the maximum \( N^*(\epsilon) := \max_s(N(s, \epsilon_1)) \) and minimum \( \epsilon_1 := \min_s(\zeta(s)) \), we see that whenever history is length \( N^*(\epsilon) \) or more, and \( \pi \in B_{\text{on}}(\pi^*, \epsilon') \) with \( \epsilon' < \epsilon_1 \), there is at least \( 1 - \epsilon/2 \) chance that the receiver’s strategy matches \( \pi^*_2 \) after every on-path signal. Since we can pick \( \gamma(\epsilon) \) large enough that \( 1 - \epsilon/2 \) measure of the receiver population is age \( N^*(\epsilon) \) or older, we are done.
To construct $N(s, \epsilon)$ and $\zeta(s)$, let $\Lambda(s) := \lambda\{\theta : \pi_1^*(s|\theta) = 1\}$. Find small enough $\zeta(s) \in (0, 1)$ so that:

- $\left|\frac{\lambda(\theta)}{\Lambda(s)+(1-\Lambda(s))} - \frac{\lambda(\theta)}{\Lambda(s)}\right| < \xi$
- $\left|\frac{\lambda(\theta)+(1-\Lambda(s))}{\Lambda(s)+(1-\Lambda(s))\cdot\zeta(s)} - \frac{\lambda(\theta)}{\Lambda(s)}\right| < \xi$
- $\frac{\zeta(s)}{1-\zeta(s)} \cdot \frac{\lambda(\theta)}{\Lambda(s)} < \xi$

for every $\theta \in \Theta$. After a history $y_2$, the receiver’s posterior belief as to the type of sender who sends signal $s$ satisfies

$$p(\theta|s; y_2) \propto \lambda(\theta) \cdot \frac{\#(\theta, s|y_2) + \alpha(\theta, s)}{\#(\theta|y_2) + A(\theta)},$$

where $\alpha(\theta, s)$ is the Dirichlet prior parameter on signal $s$ for type $\theta$ and $A(\theta) := \sum_{s \in S} \alpha(\theta, s)$. By the law of large numbers, for long enough history length, we can ensure that if $\pi_1^*(s|\theta) > 1 - \frac{\zeta(s)}{4}$, then

$$\frac{\#(\theta, s|y_2) + \alpha(\theta, s)}{\#(\theta|y_2) + A(\theta)} \geq 1 - \zeta(s)$$

with probability at least $1 - \frac{\epsilon}{2|S|}$, while if $\pi_1^*(s|\theta) < \frac{\zeta(s)}{4}$, then

$$\frac{\#(\theta, s|y_2) + \alpha(\theta, s)}{\#(\theta|y_2) + A(\theta)} < \zeta(s)$$

with probability at least $1 - \frac{\epsilon}{2|S|}$. Moreover there is some $N(s, \epsilon)$ so that there is probability at least $1 - \frac{\epsilon}{2|S|}$ that a history $y_2$ with length at least $N(s, \epsilon)$ satisfies above for all $\theta$. But at such a history, for any $\theta$ such that $\pi_1^*(s|\theta) = 1$,

$$p(\theta|s; y_2) \geq \frac{\lambda(\theta) \cdot (1 - \zeta(s))}{\Lambda(s)+(1-\Lambda(s)) \cdot \zeta(s)}$$

and

$$p(\theta|s; y_2) \leq \frac{\lambda(\theta)}{\Lambda(s) \cdot (1 - \zeta(s))},$$

while for some $\theta$ such that $\pi_1^*(s|\theta) = 0$,

$$p(\theta|s; y_2) \leq \frac{\zeta(s)}{1-\zeta(s)} \cdot \frac{\lambda(\theta)}{\Lambda(s)}.$$ 

Therefore the belief $p(\cdot|s; y_R)$ is no more than $\xi$ away from $p(\theta; s, \pi^*)$, as desired. \qed