Type-Compatible Equilibria in Signalling Games*

Drew Fudenberg[†] Kevin He[‡]

First version: October 12, 2016
This version: February 6, 2017

Abstract

The key issue in selecting between equilibria in signalling games is determining how receivers will interpret deviations from the path of play. We develop a foundation for these off-path beliefs, and an associated equilibrium refinement, in a model where equilibrium arises from non-equilibrium learning by long-lived senders and receivers. In our model, non-equilibrium signals are sent by young senders as experiments to learn about receivers' behavior, and different types of senders have different incentives for these various experiments. Using the Gittins index (Gittins, 1979), we characterize which sender types use each signal more often, leading to a constraint we call the "compatibility criterion" on the receiver's off-path beliefs and to the concept of a "type-compatible equilibrium." We compare type-compatible equilibria to signalling-game refinements such as the Intuitive Criterion (Cho and Kreps, 1987) and divine equilibrium (Banks and Sobel, 1987).

*We thank Laura Doval, Glenn Ellison, Lorens Imhof, Yuichiro Kamada, Robert Kleinberg, David Levine, 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: drewf@mit.edu
[‡]Department of Economics, Harvard University. Email: he02@fas.harvard.edu

1 Introduction

In a signalling game, an informed *sender* (for instance a student) observes their type (e.g. ability) and chooses a signal (for example, an education level) that is observed by a *receiver* (such as an employer), who then chooses an action without observing the sender's type. These signalling games can have many perfect Bayesian equilibria, which are supported by different beliefs of the receivers when observing "off-path" signals that should never be observed if all senders follow the equilibrium strategy. Solution concepts such as perfect Bayesian equilibrium and sequential equilibrium place no restrictions on off-path beliefs, while equilibrium refinements like Cho and Kreps (1987)'s Intuitive Criterion and Banks and Sobel (1987)'s divine equilibrium reduce the set of equilibria by assuming that players undertake complex and even iterated reasoning to figure out the meaning of an off-path signal.

In this paper, we use the theory of learning in games to provide a microfoundation for off-path beliefs, allowing us to determine which of these beliefs — and hence which associated equilibria — are plausible. Specifically, we develop a learning model with large populations of senders and receivers with exponentially distributed lifetimes, who are randomly matched to play the signalling game each period. These agents do not know the strategies used by the opposing population. Instead, they believe they face a constant distribution of opponents' play and are born with a non-doctrinaire prior over these distributions. Young senders rationally experiment with various signals, including some that are off the equilibrium path, to learn how receivers respond. Receivers encounter these experimenting senders throughout their lifetime and learn from personal experience, forming a Bayesian belief about the sender's type after every signal. We view signalling game equilibria as the steady states of this adjustment process, in the spirit of Spence (1973)'s interpretation of signalling game equilibrium as a "nontransitory configuration" of the following "information feedback system":

> "*As successive waves of new applicants come into the market, we can imagine repeated cycles around the loop. Employers' conditional probabilistic beliefs are modified, offered wage schedules are adjusted, applicant behavior with respect to signal choice changes, and after hiring, new data become available to the employer.*"

The key to our results is that different types of senders have more or less to gain from experimenting with a given signal. This generates restrictions on what receivers typically observe, and hence on what receivers believe when they see each signal. We develop a refinement based on the possible steady states of the learning system when the agents are both long lived (so they get enough data to learn about the consequences of frequently played actions) and patient (so senders have an incentive to experiment with any signal that could possibly improve on their steady-state payoff).

To carry out our analysis, we exploit the assumption that each agent's lifetime follows an exponential distribution and use the Gittins index (Gittins, 1979) to characterize how senders experiment and learn. We combine this with law-of-large-numbers arguments and a new result of Fudenberg, He, and Imhof (2016) on updating posteriors after rare events to characterize the limits of steady states as agents become patient and long-lived. For example, in the beer-quiche game studied by Cho and Kreps (1987), the Gittins index shows that the strong type has greater incentive to experiment with beer than the weak type does. We show that this implies that long-lived receivers are unlikely to revise the probability of the strong type downwards following an observation of "beer". Therefore the "both types eat quiche" equilibrium is not a steady state of the learning model, as it requires receivers to interpret "beer" as a signal that the sender is weak.

As a consequence, the steady states with long-lived and patient learners must be "type-compatible equilibria" or "TCE," which are Nash equilibria with restrictions on beliefs that we derive from the Gittins index. Type-compatible equilibria rules out some equilibria that satisfy the Intuitive Criterion. It does not rule out any divine equilibria (Banks and Sobel, 1987), and indeed every equilibrium satisfying a uniform version of TCE is a universally divine equilibrium. Importantly, though, in the learning-based approach we develop here, the restrictions on receiver's beliefs arise from Bayesian updating and the fact that the receiver does sometimes observe play of the non-equilibrium messages. This contrasts with the motivations for the Intuitive Criterion and divine equilibrium, which were justified in terms of deductive reasoning by the players about the equilibrium meaning of messages that the equilibrium says should never be observed.

In Section 5 we say more about how type-compatible equilibrium relates to these other refinements.

Related Work

In addition to the papers referenced above, this paper is closely related to the Fudenberg and Levine (1993) and Fudenberg and Levine (2006) analyses of the steady states of patient rational learning when agents have long but finite lifetimes. In general extensive-form games, Fudenberg and Levine (1993) showed that when agents have long lifespan and high patience, they experiment enough to learn the consequence of deviating from the equilibrium path, so that every steady state must correspond to a Nash equilibrium. Fudenberg and Levine (2006) considered a subclass of perfect-information games and studied whether agents at off-path nodes have an incentive to experiment to learn about the play of subsequent movers, as would be necessary for patient rational learning to imply backward induction. There are no subsequent movers at off-path nodes in signalling games, so that is a moot issue here. Conversely, our question of the relative probabilities of experiments that lead to the same information set does not arise in games of perfect information. Thus these two papers are complementary studies of different aspects of rational experimentation in settings where opponents' strategies are unknown.

Our paper is also related to the literature studying Bayesian learning in repeated games with a "grain of truth" (Kalai and Lehrer, 1993; Esponda and Pouzo, 2016) and the literature on boundedly rational experimentation in extensive-form games, including Fudenberg and Levine (1988); Fudenberg and Kreps (1993, 1995); Jehiel and Samet (2005); Noldeke and Samuelson (1997); Laslier and Walliser (2014) as well as to the Bayesian learning model of Kalai and Lehrer (1993). For most of the paper we assume that each sender's type is fixed at birth, though we also discuss the case of i.i.d. types; Dekel, Fudenberg, and Levine (2004) show some of the differences this can make. Also, we assume that agents assign zero probability to dominated strategies of their opponents, as in the Intuitive Criterion, divine equilibrium, and rationalizable self-confirming equilibrium (Dekel, Fudenberg, and Levine, 1999).

Rabin and Sobel (1996) consider a quasi-dynamic model of deviations from equilibrium in signalling games that starts from an exogenous theory of plausible deviations. By contrast, our work considers an explicit dynamic learning model where deviations from equilibrium arise endogenously. Unlike the myopia assumption implicit in Rabin and Sobel (1996) , our focus is on learning outcomes when agents discount the future very little. This patience is essential to be sure that the agents experiment enough to rule out the non-Nash but self-confirming equilibria.

In our model, senders solve a multi-armed bandit problem, with the different messages in the signalling game corresponding to different arms. Robbins (1952) first formulated the multi-armed bandit model to study the problem of a statistician choosing what experiments to undertake next as a function of past observations. Gittins and Jones (1974) showed that the infinite-horizon, discounted multi-armed bandit is indexable, meaning there exists an index for each arm that depends only that arm's posterior distribution of returns, such that the optimal policy is to pull the arm with the highest index each period. Gittins (1979) characterized this index function, which is now commonly known as the Gittins index. Finally, one of our proofs relies on Fudenberg, He, and Imhof (2016), who study the properties of Bayesian posterior beliefs after rare events. Theorem 2 of Fudenberg, He, and Imhof (2016) lets us conclude that with high probability the receiver's belief updating moves in the direction of the sender's experimentation incentives.

2 Type-Compatible Equilibria in Signalling Games

2.1 Signalling Games

A signalling game has two players, a sender and a receiver. The sender's type is drawn from a finite set Θ according to a prior $\lambda \in \Delta(\Theta)$ with $\lambda(\theta) > 0$ for all θ, where here and subsequently $\Delta(X)$ is the collection of probability distributions on set X. There is a finite set M of messages for the sender and a finite set A of actions for the receiver.[1] The utility functions of the sender and receiver are $u_S : \Theta \times M \times A \to \mathbb{R}$ and $u_R : \Theta \times M \times A \to \mathbb{R}$.

When the game is played, the sender knows her type and sends a message $m \in M$ to the receiver. The receiver observes the message, then responds with an action $a \in A$. Finally, payoffs are realized.

A *behavioral strategy for the receiver* is a collection of probability distributions over actions A, one for each message, $(\pi_R(\cdot|m))_{m\in M}$.

For $P \subseteq \Delta(\Theta)$, we let $\text{BR}(P, m) := \bigcup_{p\in P} \left(\underset{a'\in A}{\arg\max}\ u_R(p, m, a') \right)$; this is the set of best responses to m supported by some belief in P. The receiver action a is *conditionally dominated after message* m if it is not a best response to any belief about the sender's type, that is if

$$a \notin \text{BR}(\Delta(\Theta), m).$$

Thus $\Pi_R := \times_{m\in M}\Delta(BR(\Delta(\Theta), m))$ is the set of mixtures over behavioral strategies that never play a conditionally dominated action after any message. [2]

A *behavioral strategy for the sender* is a collection $\pi_S = (\pi_S(\cdot|\theta))_{\theta\in\Theta}$, with each $\pi_S(\cdot|\theta)$ an element of the set $\Delta(M)$ of probability distributions on M. For a given π_S, message m is *off the path of play* if it has probability 0, that is if $\pi_S(m|\theta) = 0$ for all θ.

Analogous to the definition of Π_R, call a message m *dominated for type* θ if it is not a best response to any belief about receiver's strategy, that is if

$$m \notin \bigcup_{\pi_R\in(\Delta(A))^{|M|}} \left(\underset{m'\in M}{\arg\max}\ u_S(\theta, m', \pi_R(\cdot|m')) \right).$$

We denote the set of undominated messages for type θ by $UD(\theta)$, so $\Pi_S := \times_\theta \Delta\left(UD(\theta)\right)$ is the subset of sender's behavioral strategies where no type ever plays a dominated message.

[1] To lighten notation we assume that the same set of actions is feasible following any message. This is without loss of generality for our results as we could define the receiver to have very negative payoffs when he responds to a message with an "impossible" action.

[2] Recall that the set of mixed best responses need not be convex.

2.2 Type-Compatible Equilibria

We now introduce type-compatible equilibrium, a refinement of Nash equilibrium in signalling games. In Sections 3 and 4, we develop a steady-state learning model where populations of long-lived senders and receivers, initially uncertain as to the 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." Our main result, Theorem 4, shows that only type-compatible equilibria can be patiently stable, and thus provides a learning-based justification for type-compatible equilibrium as a solution concept. We also show later that a uniform version of this solution concept is path-equivalent to universally divine equilibrium.

Definition 1. Type θ' is more *compatible* with message m' than type θ'', written as $\theta' \succ_{m'} \theta''$, if for every $\pi_R \in \Pi_R$ such that

$$u_S(\theta'', m', \pi_R(\cdot|m')) \geq \max_{m'' \neq m'} u_S(\theta'', m'', \pi_R(\cdot|m'')),$$

we have

$$u_S(\theta', m', \pi_R(\cdot|m')) \quad > \quad \max_{m'' \neq m'} u_S(\theta', m'', \pi_R(\cdot|m'')).$$

So, $\theta' \succ_{m'} \theta''$ means that whenever m' is a weak best response for θ'' against some rational receiver strategy π_R, it is a strict best response for θ' against π_R. More generally, it turns out that even if senders play many times and experiment rationally, if $\theta' \succ_{m'} \theta''$ then type θ' will play m' whenever type θ'' does, provided the two types hold the same beliefs. We elaborate on this point later in Proposition 3, as it is a key foundation for our results.

Proposition 1.
(a) $\succ_{m'}$ *is transitive.*
(b) Unless m' *is strictly optimal for both* θ' *and* θ'' *against every* $\pi_R \in \Pi_R$*, or* m' *is never weakly optimal for either* θ' *or* θ'' *against any* $\pi_R \in \Pi_R$*,* $\theta' \succ_{m'} \theta''$ *implies* $\theta'' \not\succ_{m'} \theta'$*,*

Proof. To show (a), suppose $\theta' \succ_{m'} \theta''$ and $\theta'' \succ_{m'} \theta'''$. For any $\pi_R \in \Pi_R$ where m' is weakly optimal for θ''', it must be strictly optimal for θ'', hence also strictly optimal for θ'. This shows $\theta' \succ_{m'} \theta'''$.

To establish (b), partition the set of rational receiver strategies as $\Pi_R = \Pi_R^+ \cup \Pi_R^0 \cup \Pi_R^-$, where the three subsets refer to receiver strategies that make m' strictly better, indifferent, or strictly worse than the best alternative message for θ''. If the set Π_R^0 is nonempty, then $\theta' \succ_{m'} \theta''$ implies $\theta'' \not\succ_{m'} \theta'$. This is because against any $\pi_R \in \Pi_R^0$, message m' is strictly optimal for θ' but only weakly optimal for θ''. At the same time, if both Π_R^+ and Π_R^-

are nonempty, then Π_R^0 is nonempty. This is because both $\pi_R \mapsto u_S(\theta^{''}, m^{'}, \pi_R(\cdot|m^{'}))$ and $\pi_R \mapsto \max_{m'' \neq m'} u_S(\theta^{''}, m^{''}, \pi_R(\cdot|m^{''}))$ are continuous functions, so for any $\pi_R^+ \in \Pi_R^+$ and $\pi_R^- \in \Pi_R^-$, there exists $\alpha \in (0,1)$ so that $\alpha\pi_R^+ + (1-\alpha)\pi_R^- \in \Pi_R^0$. If only Π_R^+ is nonempty and $\theta^{'} \succ_{m'} \theta^{''}$, then $m^{'}$ is strictly dominant for both $\theta^{'}$ and $\theta^{''}$ when the receiver is restricted to strategies in Π_R. If only Π_R^- is nonempty, then we can have $\theta^{''} \succ_{m'} \theta^{'}$ only when $m^{'}$ is never a weak best response for $\theta^{'}$ against any $\pi_R \in \Pi_R$. $\square$

To check the compatibility condition one must consider all strategies in Π_R, just as the belief restrictions in divine equilibrium involve all the possible mixed best responses to various beliefs. However, when the sender's utility function is separable in the sense that $u_S(\theta, m, a) = v(\theta, m) + z(a)$, as in Spence (1973)'s job market signalling game and in Cho and Kreps (1987)'s beer-quiche game (given below), a sufficient condition for $\theta^{'} \succ_{m'} \theta^{''}$ is

$$v(\theta^{'}, m^{'}) - v(\theta^{''}, m^{'}) > \max_{m'' \neq m'} v(\theta^{'}, m^{''}) - v(\theta^{''}, m^{''}).$$

This can be interpreted as saying $m^{'}$ is the least costly message for $\theta^{'}$ relative to $\theta^{''}$. In Appendix A, we present a general sufficient condition for $\theta^{'} \succ_{m'} \theta^{''}$ for general payoff functions.

Example 1. In the beer-quiche game, the sender is either strong (θ_{strong}) or weak (θ_{weak}), with prior probability $\lambda(\theta_{\text{strong}}) = 0.9$. The sender chooses to either drink beer or eat quiche for breakfast. The receiver, observing this breakfast choice but not the sender's type, chooses whether to fight the sender. If the sender if θ_{weak}, receiver prefers fighting. If the sender if θ_{strong}, receiver prefers not fighting. Also, θ_{strong} prefers beer for breakfast while θ_{weak} prefers quiche for breakfast. Both types prefer not being fought over having their favorite breakfast.

beer (B)	fight (F)	not fight (NF)
$\theta_{\texttt{strong}}$	1,0	3,1
$\theta_{\texttt{weak}}$	0,1	2,0

quiche (Q)	fight (F)	not fight (NF)
$\theta_{\texttt{strong}}$	0,0	2,1
$\theta_{\texttt{weak}}$	1,1	3,0

This game has separable sender utility with $v(\theta_{\text{strong}}, B) = v(\theta_{\text{weak}}, Q) = 1$, $v(\theta_{\text{strong}}, Q) = v(\theta_{\text{weak}}, B) = 0$, $z(F) = 0$ and $z(NF) = 2$. So, we have $\theta_{\text{strong}} \succ_B \theta_{\text{weak}}$. ♦

Recall that $UD(\theta)$ is the set of undominated messages for type θ, so $UD^{-1}(m^{'})$ is the set of types for which $m^{'}$ is not dominated. For a fixed strategy profile π^*, let $u_S(\theta; \pi^*)$ denote the payoff to type θ under π^*, and let

$$J(m', \pi^*) \;\; := \;\; \left\{ \theta \in \Theta : \max_{a \in \mathrm{BR}(\Delta(\Theta), m')} u_S(\theta, m', a) > u_S(\theta; \pi^*) \right\}$$

be the set of types for which *some* best response to message m is better than their payoff under π^*.[3]

Definition 2. The *compatible beliefs at message m' under profile* π^* is the set

$$P(m', \pi^*) := \left\{ p \in \Delta(UD^{-1}(m')) : \; \frac{p(\theta'')}{p(\theta')} \le \frac{\lambda(\theta'')}{\lambda(\theta')} \;\text{ whenever }\; \begin{array}{c} \text{(i) } \theta' \succ_{m'} \theta'' \\ \text{and} \\ \text{(ii) } \theta' \in J(m', \pi^*) \end{array} \right\}.$$

If $UD^{-1}(m') = \emptyset$, then m' is never a tempting deviation for any type, and the receiver's beliefs and actions after m' are irrelevant; here we set $P(m', \pi^*) := \Delta(\Theta)$. Note that $P(m', \pi^*)$ is always non-empty, since for every (m', π^*), whenever $UD^{-1}(m') \neq \emptyset$, the prior λ conditioned on $UD^{-1}(m')$ is always in $P(m', \pi^*)$.

The motivation for this definition comes from our learning model, where the more compatible type θ' will experiment with m' more often than the less compatible type θ'' does, so that seeing m' should not make the receiver increase the odds ratio of θ'' to θ'.

Definition 3. Strategy profile π^* *satisfies the compatibility criterion at* m' if $\pi_R^*(\cdot|m') \in \Delta(\mathrm{BR}(P(m', \pi^*), m'))$.

Definition 4. Strategy profile π^* a *type-compatible equilibrium* if it is a Nash equilibrium and satisfies the compatibility criterion for every off-path message m'.

Like divine equilibrium and unlike the Intuitive Criterion or Cho and Kreps (1987)'s $D1$ criterion, the compatibility criterion says only that some messages should not increase the relative probability of "implausible" types, as opposed to requiring that these types have probability 0.

2.3 Intuitions for the Compatibility Criterion

To help build some intuition for our definitions and results, we now examine some of the implications and properties of the compatibility criterion. The next result shows that in every

[3] The reverse strict inequality would mean that m' is "equilibrium dominated" for θ in the sense of Cho and Kreps (1987).

perfect Bayesian equilibrium[4] the relative frequencies that types $\theta^{'}$ and $\theta^{''}$ play message $m^{'}$ respect compatibility. By Bayes' rule, this implies that the receiver's equilibrium belief after every *on-path* message $m^{'}$ satisfies the compatibility criterion.

Proposition 2. *If π^* is a perfect Bayesian equilibrium and $\theta^{'} \succ_{m'} \theta^{''}$, then $\pi_S^*(m^{'}|\theta^{'}) \geq \pi_S^*(m^{'}|\theta^{''})$, so if m' is on the equilibrium path with $\pi_S^*(m^{'}|\theta') > 0$, the receiver's posterior beliefs $p(\theta|m)$ satisfy*

$$\frac{p(\theta^{''}|m')}{p(\theta'|m')} \leq \frac{\lambda(\theta^{''})}{\lambda(\theta')}.$$

Proof. It suffices to show that if $\pi_S^*(m^{'}|\theta^{''}) > 0$, then $\pi_S^*(m^{'}|\theta^{'}) = 1$. But since π^* is a PBE, then $\pi_S^*(m^{'}|\theta^{''}) > 0$ implies $m^{'}$ is weakly optimal for type $\theta^{''}$, that is

$$u_S(\theta^{''}, m^{'}, \pi_R^*(\cdot|m^{'})) \geq \max_{m^{''} \neq m^{'}} u_S(\theta^{''}, m^{''}, \pi_R^*(\cdot|m^{''})).$$

By the definition of $\theta^{'} \succ_{m'} \theta^{''}$, this implies

$$u_S(\theta^{'}, m^{'}, \pi_R^*(\cdot|m^{'})) > \max_{m^{''} \neq m^{'}} u_S(\theta^{'}, m^{''}, \pi_R^*(\cdot|m^{''})),$$

so we have $\pi_S^*(m^{'}|\theta^{'}) = 1$, as otherwise the sender could strictly gain by deviating to playing $m^{'}$ all the time when her type is $\theta^{'}$. ☐

Type-compatible equilibrium differs from perfect Bayesian equilibrium in requiring that receiver's beliefs after *off-path* messages also satisfy the compatibility criterion. To gain more intuition for why this is an implication of rational experimentation by the senders, we relate our definition of compatibility to the Gittins index. Suppose type θ knows that the receiver is playing the same behavioral strategy π_R^* every period, but is uncertain as to what π_R^* is. The sender wishes to maximize her expected discounted utility, where in each period she chooses a message m, observes one draw from $\pi_R^*(\cdot|m)$, and receives the associated payoffs that period. If her belief about π_R^* is independent across messages, then she effectively faces a discounted multi-armed bandit problem, where the different arms are the different messages.

Write $\nu_m \in \Delta(\Delta(\mathrm{BR}(\Delta(\Theta), m)))$ for a belief over rational receiver strategies after message m and $\nu = (\nu_m)_{m \in M}$ is a profile of such beliefs. Write $I(\theta, m, \nu, \beta)$ for the Gittins index of message m for type θ, with beliefs ν over receiver's strategies after various messages, so that

$$I(\theta, m, \nu, \beta) := \sup_{\tau > 0} \frac{\mathbb{E}_{\nu_m}\left\{\sum_{t=0}^{\tau-1} \beta^t \cdot u_S(\theta, m, a_m(t))\right\}}{\mathbb{E}_{\nu_m}\left\{\sum_{t=0}^{\tau-1} \beta^t\right\}}.$$

[4]In signalling games, both perfect Bayesian equilibrium (Fudenberg and Tirole, 1991) and sequential equilibrium (Kreps and Wilson, 1982) reduce to Nash equilibrium in conditionally undominated strategies.

Here, $a_m(t) \in \text{BR}(\Delta(\Theta), m)$ is the receiver's response t-th time message m is sent, and the expectation $\mathbb{E}_{\nu_m}$ over the sequence of responses $\{a_m(t)\}_{t\geq 0}$ depends on the prior ν_m. The next Proposition relates the compatibility definition to the Gittins index, connecting the payoff functions in the signalling game to optimal experimentation.

Proposition 3. *$\theta^{'} \succ_{m^{'}} \theta^{''}$ if and only if for every $\beta \in [0,1)$ and every ν, $I(\theta^{''}, m^{'}, \nu, \beta) \geq \max_{m^{''}\neq m^{'}} I(\theta^{''}, m^{''}, \nu, \beta)$ implies $I(\theta^{'}, m^{'}, \nu, \beta) > \max_{m^{''}\neq m^{'}} I(\theta^{'}, m^{''}, \nu, \beta)$.*

The proof of this result is in Appendix B.1. In outline, the idea is that every stopping time τ for sequential experiments with message m induces a distribution $\sigma_m(\tau, \nu, \beta)$ over the (expected discounted) receiver actions that will be observed before stopping. We can view this distribution as a mixed strategy of the receiver, so that the optimal stopping problem that defines the Gittins index, evaluated at τ, yields the sender's one-period payoff against the receiver strategy induced by τ. Moreover, when message $m^{'}$ has the highest Gittins index for type $\theta^{''}$, it is also better for $\theta^{''}$ than using the stopping rule of type $\theta^{'}$ on any other message $m^{''}$, so $I(\theta^{''}, m^{'}, \nu, \beta) \geq \max_{m^{''}\neq m^{'}} u_S(\theta^{''}, m^{''}, \sigma_{m^{''}}(\nu_{m^{''}}, \tau^{\theta^{'}}_{m^{''}}, \beta))$. When $\theta^{'}$ and $\theta^{''}$ share the same beliefs, this lets us apply the compatibility definition.

Lemma 1 below uses an inductive argument to extend the conclusion of Proposition 3 to the histories and beliefs that arise under the optimal policies in a steady-state learning model, where the two types need not have the same beliefs about the receiver's play.

2.4 Uniform Type-Compatible Equilibria

We now define a subset of TCE. Write

$$\hat{P}(m^{'}) := \left\{ p \in \Delta(UD^{-1}(m^{'})) : \ \frac{p(\theta^{''})}{p(\theta^{'})} \leq \frac{\lambda(\theta^{''})}{\lambda(\theta^{'})} \text{ whenever } \theta^{'} \succ_{m^{'}} \theta^{''} \right\}.$$

The difference between $\hat{P}$ and P is that $\hat{P}$ applies to all pairs $\theta^{'} \succ_{m^{'}} \theta^{''}$, whether or not $\theta^{'} \in J(m^{'}, \pi^*)$, so it imposes more restrictions on the receiver's beliefs.

Definition 5. A Nash equilibrium strategy profile π^* is called a *uniform type-compatible equilibrium (uniform TCE)* if for all θ, all off-path messages $m^{'}$ and all $a \in \text{BR}(\hat{P}(m^{'}), m^{'})$, we have $u_S(\theta; \pi^*) \geq u_S(\theta, m^{'}, a)$.

Uniform TCE does allow types to randomize on the path of play, and in particular equilibria where all messages are on-path are vacuously uniform TCE.

The "uniformity" in uniform TCE comes from the requirement that *every* best response to *every* belief in $\hat{P}(m^{'}) \subseteq P(m^{'}, \pi^*)$ deters every type from deviating to the off-path $m^{'}$. By contrast, a (regular) TCE is a Nash equilibrium where *some* best response to $P(m^{'}, \pi^*)$ deters every type from deviating to $m^{'}$.

The following example illustrates this difference.

Example 2. Suppose a worker can have either high ability (θ_H) or low ability (θ_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 separable sender payoffs, with $z(\text{low wage}) = 0$, $z(\text{medium wage}) = 6$, $z(\text{high wage}) = 9$ 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 e PhD is costly.) The employer's payoffs reflect a desire to pay a wage corresponding to the worker's ability and increased productivity with education.

N	low	med	high
θ_H	0,-2	6,0	9,1
θ_L	0,1	6,0	9,-2

C	low	med	high
θ_H	2,-1	8,1	11,2
θ_L	1,2	7,1	10,-1

D	low	med	high
θ_H	-2,0	4,2	7,3
θ_L	-4,3	2,2	5,0

Since $v(\theta_H, \cdot) - v(\theta_L, \cdot)$ is maximized at D, θ_H is more compatible with D than θ_L is. Similarly, θ_L is more compatible with N than θ_H is. There is no compatibility relation at message C.

When the prior is $\lambda(\theta_H) = 0.5$, the strategy profile where the employer always pays a medium salary and both types of worker choose C is a uniform TCE. 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 salary and medium salary, deters every type from deviating. At the same time, no type wants to deviate to D, even if she gets paid the best salary. On the other hand, the equilibrium where the employer pays a low salary for N and C, a medium salary for D, and both types choose D is a TCE but not a uniform TCE. The receiver's play satisfies the compatibility criterion after every off-path information set, but medium salary is also a best response to $\hat{P}(N)$ and it tempts type θ_L to deviate to N. ⧫

As a partial converse to our result that every patiently stable strategy profile is a TCE, we show in Theorem 6 that under additional strictness conditions, every uniform TCE is path-equivalent to a patiently stable strategy profile. We also show in Corollary 1 that every uniform TCE is universally divine, up to path equivalence. [5]

[5]We do not know whether the non-uniform TCE ("both types play D", low wage to any choice) is patiently stable.

3 The Steady State Learning Model

3.1 The Aggregate Model

In the aggregate model, there is a continuum of agents, with a unit mass in the role of receivers and mass $\lambda(\theta)$ in the role of type θ for each $\theta \in \Theta$. Time is doubly infinite and generations overlap. In each period, each agent has $\gamma \in [0, 1)$ chance of surviving and complementary chance $(1 - \gamma)$ of leaving the system. To preserve population sizes, $(1 - \gamma)$ new receivers and $\lambda(\theta)(1 - \gamma)$ new type θ are born into system every period. Each sender learns her type upon birth, which is fixed for life. All agents are rational Bayesians who discount future utility flows by $\delta \in [0, 1)$, so their objective is to maximize the expected value of $\sum_{t=0}^{\infty}(\gamma\delta)^t \cdot u_t$, where $\gamma\delta \in [0, 1)$ is the effective discount factor and u_t is payoff t periods from today. Each period all agents are randomly matched to play the signalling game. Each sender has probability $(1-\gamma)\gamma^t$ of meeting a receiver of age t, while each receiver has $\lambda(\theta)(1-\gamma)\gamma^t$ chance of meeting a type θ of age t. At the end of the period, agents observe the outcomes of their own match – namely, the message sent, the action played in response, and the sender's type. They update their beliefs (as described in Subsection 3.3) and (if still active) play again. Importantly, the sender does not observe receiver's extensive-form strategy, because a sender who plays m in a match does not observe how the receiver would have reacted had she played $m' \neq m$ instead.

3.2 Beliefs about Opponents' Strategies

Agents are rational Bayesians who believe they face a fixed but unknown distribution of opponents' play. Each sender is born with a prior g_S, which is a density function over receiver's behavioral strategies — that is, a Lebesgue-measurable function $g_S : (\Delta(A))^{|M|} \to \mathbb{R}_+$ that integrates to 1. Similarly, each receiver is born with a prior density over the sender's behavioral strategies, $g_R : (\Delta(M))^{|\Theta|} \to \mathbb{R}_+$. We denote the m component of g_S as $g_S^{(m)}$, so that $g_S^{(m)} : \Delta(A) \to \mathbb{R}_+$ is the prior of new senders over how receivers respond to message m. Similarly, we denote the θ component of g_R as $g_R^{(\theta)}$, so that $g_R^{(\theta)} : \Delta(M) \to \mathbb{R}_+$ is the receivers' prior over how type θ plays.

Definition 6. Call priors (g_S, g_R) **regular** if

(a). [*independence*] $g_S = \underset{m \in M}{\times} g_S^{(m)}$ and $g_R = \underset{\theta \in \Theta}{\times} g_R^{(\theta)}$.

(b). [*rationalizability*] g_S puts probability 1 on Π_R while g_R puts probability 1 on Π_S.

(c). [*g_S non-doctrinaire*] g_S is continuous and strictly positive on the relative interior of Π_R.

(d). [g_R *nice*] For each type θ, there are positive constants $\left(\alpha_m^{(\theta)}\right)_{m\in UD(\theta)}$ such that

$$\frac{g_R^{(\theta)}(p)}{\prod_{m\in UD(\theta)} p_m^{\alpha_m^{(\theta)}-1}}$$

is uniformly continuous and bounded away from zero on the relative interior of $\Pi_S^{(\theta)}$.

Independence ensures that the receiver does not learn how type θ plays by observing the behavior of some other type $\theta' \neq \theta$, and that the sender does not learn how receiver reacts to message m by experimenting with some other message $m' \neq m$, so that for example the sender doesn't learn about how receivers respond to beer by sending quiche.[6] Rationalizability says players know each other's payoff structures and anticipate that their opponent will not play dominated strategies. The non-doctrinaire nature of g_S and g_R allows a large enough data set to outweigh prior beliefs. (An agent who assigns probability 0 to some neighborhood of mixed actions may not even have a convergent posterior belief when facing a data-generating process in that neighborhood, see for example Berk (1966)).

The technical assumption about the boundary behavior of g_R in (d) ensures that the prior density function $g_R^{(\theta)}$ behaves like a power function near the boundary of $\Pi_S^{(\theta)}$. Any density that is strictly positive on $\Pi_S^{(\theta)}$ satisfies this condition, as does the Dirichlet distribution, which is the prior associated with fictitious play (see Fudenberg and Kreps (1993)).

3.3 Individual Learning and Type Compatibility

The time-t history of a type θ belongs to the set

$$Y_\theta[t] := \left(\bigcup_{m\in UD(\theta)} \{m\} \times \mathrm{BR}(\Delta(\Theta), m) \right)^t,$$

where each period the history records the message that the sender chose and the response of the receiver in her match. Note that the updating and optimization of the agents is well-defined at each history in $Y_\theta[t]$, because the set rules out histories with prior probability 0, where either type θ sent a dominated message or the receiver played a conditionally dominated response. The set of all histories for type θ is the union $Y_\theta := \bigcup_{t=0}^{\infty} Y_\theta[t]$.

Given prior g_S and effective discount factor $\gamma\delta$, the sender's dynamic optimization problem has an optimal policy function $s_\theta : Y_\theta \to M$ that maps each history to a message[7].

[6] One could imagine learning environments where the senders believe that the responses to various messages are correlated, but independence is a natural special case.

[7] Of course, the optimal policy function s_θ depends on prior g_S, patience δ, and survival chance γ. Where no confusion arises, we suppress these dependencies. If multiple optimal policies exist due to ties in payoffs, pick one optimal policy arbitrarily. The same remarks apply to the receiver's optimization described below.

Analogously, each receiver is born with the same regular prior g_R. He believes he is facing a time-invariant distribution over sender's strategies Π_S and maximizes expected discounted utility with effectively discount factor $\gamma\delta$. The time t history of the receiver is an element of

$$Y_R[t] \coloneqq \left((\times_{m\in M} \mathrm{BR}(\Delta(\Theta), m)) \times \left(\bigcup_{\theta\in\Theta} \{\theta\} \times UD(\theta) \right) \right)^t .$$

That is, in each period the history records the pure (message-contingent) strategy that the receiver commits to, the type of the sender in his match (which is revealed at the end of the period), and the message that the sender played. The set of all histories of the receiver is the union $Y_R \coloneqq \bigcup_{t=0}^{\infty} Y_R[t]$. The receiver's problem also admits some optimal policy function $s_R : Y_R \to \Pi_R$.

To state the next lemma, we introduce the concept of a *response sequence*.

Definition 7. A *response sequence* $\mathfrak{a} = (a_{1,m}, a_{2,m}, ...,)_{m\in M}$ is an element in $\times_{m\in M} (\mathrm{BR}(\Delta(\Theta), m)^{\infty})$.

Each response sequence induces an infinite history $y_\theta(\mathfrak{a})$ for each type θ, defined in the following way.

Definition 8. The *history induced by* $\mathfrak{a}$ *for type* θ , $y_\theta(\mathfrak{a})$, is defined iteratively through its time-t truncations, with $y_\theta^t(\mathfrak{a}) \in Y_\theta[t]$.

In step 0, initialize $y_\theta^0(\mathfrak{a}) \coloneqq \emptyset$ and $\#(m; 0) \coloneqq 0$ for all $m \in M$.

Then iteratively, in step t put $m^t \coloneqq s_\theta(y_\theta^{t-1}(\mathfrak{a}))$, $a^t \coloneqq a_{\#(m^t;t-1)+1,m^t}$, $y_\theta^t(\mathfrak{a}) \coloneqq (y_\theta^{t-1}(\mathfrak{a}), m^t, a^t)$ and $\#(m^t; t) \coloneqq \#(m^t; t-1) + 1$, $\#(m; t) \coloneqq \#(m; t-1)$ for $m \neq m^t$.

A response sequence is an $|M|$-tuple of infinite sequences of receiver actions, one sequence for each message, with $a_{j,m}$ describing how the receiver would respond to the j-th instance of message m.[8] (If the sender sends m'' 5 times and then sends $m' \neq m''$, the response she gets to m' under response sequence $\mathfrak{a} = (a_{1,m}, a_{2,m}, ...)_{m\in M}$ is $a_{1,m'}$, not $a_{6,m'}$.) Fixing any regular prior g_S of the sender, a response sequence $\mathfrak{a}$ together with s_θ, the optimal policy of type θ generates a (deterministic) infinite history of experiments and responses, which we defined above as $y_\theta(\mathfrak{a})$.

The history $y_\theta(\mathfrak{a})$ is defined iteratively through its truncations $y_\theta^{t-1}(\mathfrak{a})$ and the counter $\#(m; t-1)$, which keeps track of how many times θ has played message m as of the end of period $t-1$. If we know $y_\theta^{t-1}(\mathfrak{a})$ and $\#(\cdot; t-1)$, we can deduce what will happen in period t. Sender will choose message m^t according to the optimal policy applied to her current history. This message will be met with a^t, which is element number $\#(m^t; t-1) + 1$ in the response sequence for m^t. So, her new history is $(y_\theta^{t-1}(\mathfrak{a}), m^t, a^t)$. Finally, we update the counter so that the count of message m^t is incremented by one, while all other counts stay the same.

[8]We restricted response sequences to never produce a conditionally dominated response.

The following lemma is the keystone of our results. It extends Proposition 3 to show that if $\theta^{'} \succ_{m'} \theta^{''}$, then along the two sequences $y_{\theta'}(\mathfrak{a})$ and $y_{\theta''}(\mathfrak{a})$ generated by the optimal policies of types θ' and $\theta^{''}$, the total discounted number of times that θ' plays $m^{'}$ is larger than the total discounted number of times that $\theta^{''}$ plays it, for any effective discount factor $\beta \in [0, 1)$. Later, this lemma lets us use response sequences to couple the play of $\theta^{'}$ and $\theta^{''}$ under rational experimentation.

Lemma 1. *If $\theta^{'} \succ_{m'} \theta^{''}$, then for every response sequence $\mathfrak{a}$ and every $\beta \in [0, 1)$ and any regular prior g, we have*

$$\sum_{t=0}^{\infty} \beta^t \cdot 1\{s_{\theta'}(y^t_{\theta'}(\mathfrak{a})) = m^{'}\} \geq \sum_{t=0}^{\infty} \beta^t \cdot 1\{s_{\theta''}(y^t_{\theta''}(\mathfrak{a})) = m^{'}\}.$$

Remark 1. The proof establishes the stronger claim that along each response sequence, at each point in time type $\theta^{'}$ will have played $m^{'}$ at least as many times as type $\theta^{''}$ has. Stating this formally requires additional notation developed in the proof, and the statement in the lemma is all that we need in what follows.

Proof. Let $\mathfrak{a}$ and β be given. Write T^{θ}_j for the period in which type θ sends message $m^{'}$ for the j-th time in the induced history $y_\theta(\mathfrak{a})$. If no such period exists because $\#(m^{'}, y_\theta(\mathfrak{a})) < j$, then set $T^{\theta}_j = \infty$. We use induction on the sequence of statements:

`Statement` j: Provided $T^{\theta^{''}}_j$ is finite, $\#\left(m^{''}, y_{\theta'}^{T^{\theta^{'}}_j}(\mathfrak{a})\right) \leq \#\left(m^{''}, y_{\theta''}^{T^{\theta^{''}}_j}(\mathfrak{a})\right)$ for all $m^{''} \neq m^{'}$.

`Statement 1` is the base case. By way of contradiction, suppose $T^{\theta^{''}}_1 < \infty$ and

$$\#\left(m^{''}, y_{\theta'}^{T^{\theta^{'}}_1}(\mathfrak{a})\right) > \#\left(m^{''}, y_{\theta''}^{T^{\theta^{''}}_1}(\mathfrak{a})\right)$$

for some $m^{''} \neq m^{'}$. Then there is some earliest period $t^* < T^{\theta^{'}}_1$ where

$$\#\left(m^{''}, y_{\theta'}^{t^*}(\mathfrak{a})\right) > \#\left(m^{''}, y_{\theta''}^{T^{\theta^{''}}_1}(\mathfrak{a})\right),$$

where type $\theta^{'}$ played $m^{''}$ in period t^*, $s_{\theta'}(y_{\theta'}^{t^*-1}(\mathfrak{a})) = m^{''}$.

But by construction by the end of period $t^* - 1$ type θ' has sent $m^{''}$ exactly as many times as type $\theta^{''}$ has sent it by period $T^{\theta^{''}}_1 - 1$,

$$\#\left(m^{''}, y_{\theta'}^{t^*-1}(\mathfrak{a})\right) = \#\left(m^{''}, y_{\theta''}^{T^{\theta^{''}}_1 - 1}(\mathfrak{a})\right).$$

Furthermore, neither type has sent $m^{'}$ yet, so also

$$\#\left(m^{'}, y_{\theta^{'}}^{t^{*}-1}(\mathfrak{a})\right) = \#\left(m^{'}, y_{\theta^{''}}^{T_{1}^{\theta^{''}}-1}(\mathfrak{a})\right).$$

Therefore, type $\theta^{'}$ holds the same posterior over the receiver's reaction to messages $m^{'}$ and $m^{''}$ at period $t^{*}-1$ as type $\theta^{''}$ does at period $T_{1}^{\theta^{''}}-1$. So[9] by Proposition 3,

$$m^{'} \in \operatorname*{arg\,max}_{\hat{m}\in M} I\left(\theta^{''}, \hat{m}, y_{\theta^{''}}^{T_{1}^{\theta^{''}}-1}(\mathfrak{a})\right) \implies I(\theta^{'}, m^{'}, y_{\theta^{'}}^{t^{*}-1}(\mathfrak{a})) > I(\theta^{'}, m^{''}, y_{\theta^{'}}^{t^{*}-1}(\mathfrak{a})). \quad (1)$$

However, by construction of $T_{1}^{\theta^{''}}$, we have $s_{\theta^{''}}\left(y_{\theta^{''}}^{T_{1}^{\theta^{''}}-1}(\mathfrak{a})\right) = m^{'}$. By the optimality of the Gittins index policy, the left-hand side of (1) is satisfied. But, again by the optimality of the Gittins index policy, the right-hand side of (1) contradicts $s_{\theta^{'}}(y_{\theta^{'}}^{t^{*}-1}(\mathfrak{a})) = m^{''}$. Therefore we have proven `Statement 1`.

Now suppose `statement` j holds for all $j \leq K$. We show `statement` $K+1$ also holds. If $T_{K+1}^{\theta^{''}}$ is finite, then $T_{K}^{\theta^{''}}$ is also finite. The inductive hypothesis then shows

$$\#\left(m^{''}, y_{\theta^{'}}^{T_{K}^{\theta^{'}}}(\mathfrak{a})\right) \leq \#\left(m^{''}, y_{\theta^{''}}^{T_{K}^{\theta^{''}}}(\mathfrak{a})\right)$$

for every $m^{''} \neq m^{'}$. Suppose there is some $m^{''} \neq m^{'}$ such that

$$\#\left(m^{''}, y_{\theta^{'}}^{T_{K+1}^{\theta^{'}}}(\mathfrak{a})\right) > \#\left(m^{''}, y_{\theta^{''}}^{T_{K+1}^{\theta^{''}}}(\mathfrak{a})\right).$$

Together with the previous inequality, this implies type $\theta^{'}$ played $m^{''}$ for the $\left[\#\left(m^{''}, y_{\theta^{''}}^{T_{K+1}^{\theta^{''}}}(\mathfrak{a})\right)+1\right]$-th time sometime between playing $m^{'}$ for the K-th time and playing $m^{'}$ for the $(K+1)$-th time. That is, if we put

$$t^{*} := \min\left\{\tau : \#(m^{''}, y_{\theta^{'}}^{\tau}(\mathfrak{a}))) > \#\left(m^{''}, y_{\theta^{''}}^{T_{K+1}^{\theta^{''}}}(\mathfrak{a})\right)\right\},$$

then $T_{K}^{\theta^{'}} < t^{*} < T_{K+1}^{\theta^{'}}$. By the construction of t^{*},

$$\#\left(m^{''}, y_{\theta^{'}}^{t^{*}-1}(\mathfrak{a})\right) = \#\left(m^{''}, y_{\theta^{''}}^{T_{K+1}^{\theta^{''}}-1}(\mathfrak{a})\right)$$

and also

[9] In the following equation and elsewhere in the proof, we abuse notation and write $I(\theta, m, y)$ to mean $I(\theta, m, g_S(\cdot|y))$, that is the Gittins index of type θ for message m at the posterior obtained from updating the prior g_S using history y.

$$\#\left(m', y_{\theta'}^{t^*-1}(\mathfrak{a})\right) = K = \#\left(m', y_{\theta''}^{T_{K+1}^{\theta''}-1}(\mathfrak{a})\right).$$

Therefore, type θ' holds the same posterior over the receiver's reaction to messages m' and m'' at period $t^* - 1$ as type θ'' does at period $T_{K+1}^{\theta''} - 1$. As in the base case, we can invoke Proposition 3 to show that it is impossible for θ' to play m'' in period t^* while θ'' plays m' in period $T_{K+1}^{\theta''}$. This shows `statement` j is true for every j, by induction.

To conclude the proof we show that

$$\sum_{t=0}^{\infty} \beta^t \cdot 1\{s_{\theta'}(y_{\theta'}^t(\mathfrak{a})) = m'\} \geq \sum_{t=0}^{\infty} \beta^t \cdot 1\{s_{\theta''}(y_{\theta''}^t(\mathfrak{a})) = m'\}.$$

Since $\beta \leq 1$, it suffices that $T_j^{\theta'} \leq T_j^{\theta''}$ for every j. But for every j where $T_j^{\theta''} < \infty$, `statement` j implies that $\#\left(m'', y_{\theta'}^{T_j^{\theta'}}(\mathfrak{a})\right) \leq \#\left(m'', y_{\theta''}^{T_j^{\theta''}}(\mathfrak{a})\right)$ for each $m'' \neq m'$. The number of periods that type θ' spent sending each message $m'' \neq m'$ before sending m' for the j-th time is fewer than the number of periods θ'' spent doing the same. Therefore it follows θ' sent m' for the j-th time sooner than θ'' did, that is $T_j^{\theta'} \leq T_j^{\theta''}$. Finally, if $T_j^{\theta''} = \infty$, then evidently $T_j^{\theta'} \leq \infty = T_j^{\theta''}$. $\square$

3.4 States and the One-Period-Forward Map

Now that we have defined the aggregate learning model and the associated individual dynamic optimization problems, we will next describe the states of the learning model.

A *state* ψ is a profile of distributions over histories – one distribution on Y_θ for each type θ and one distribution on Y_R for the receiver population

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

Given a state ψ, we refer to its components by $\psi_\theta \in \Delta(Y_\theta)$ and $\psi_R \in \Delta(Y_R)$. Using the optimal policies s_θ for each θ, each state ψ gives rise to a behavioral strategy $\bar{\psi}_S(\cdot|\theta) \in \Delta(M)$ for each type θ,

$$\bar{\psi}_S(m|\theta) := \psi_\theta \left\{ y_\theta \in Y_\theta : s_\theta(y_\theta) = m \right\}. \tag{2}$$

Similar, ψ and the optimal receiver policy s_R induce a behavioral strategy $\bar{\psi}_R$ of the receiver, where

$$\bar{\psi}_R(a|m) := \psi_R \left\{ y_R \in Y_R : s_R(y_R)(m) = a \right\}.$$

In the spirit of the law of large numbers, we assume that the matching process of the continuum of agent model exactly follows its probability distribution. We can now define the deterministic one-period-forward map

$$f : (\times_{\theta\in\Theta}\Delta(Y_\theta)) \times \Delta(Y_R) \to (\times_{\theta\in\Theta}\Delta(Y_\theta)) \times \Delta(Y_R).$$

which returns the state $f[\psi]$ that results tomorrow when starting at state ψ today.

The map f is defined as follows. First, new receivers and new senders of every type enter the system,

$$\begin{aligned} f[\psi]_R(\emptyset) &\coloneqq 1-\gamma \\ f[\psi]_\theta(\emptyset) &\coloneqq \lambda(\theta)\cdot(1-\gamma). \end{aligned}$$

Then, existing agents update their history. For the receivers, we have

$$f[\psi]_R(y_R,(s,\theta,m)) \coloneqq \begin{cases} \psi_R(y_R)\cdot\gamma\cdot\lambda(\theta)\cdot\bar{\psi}_S(m|\theta) & \text{if } s = s_R(y_R) \\ 0 & \text{otherwise.} \end{cases}$$

In this updating rule, we set $f[\psi]_R(y_R,(s,\theta,m)) = 0$ for all impossible histories where $s \neq s_R(y_R)$. When $s = s_R(y_R)$, the fraction of receivers who will have history $(y_R,(s,\theta,m))$ tomorrow is the product of four terms: $\psi_R(y_R)$ is the fraction of receivers who have history y_R today, γ is the probability that each such receiver survives until tomorrow, $\lambda(\theta)$ is the probability of being matched with a type θ tomorrow, and finally $\bar{\psi}_S(m|\theta)$ is the probability that this sender will play message m.

Analogously, the existing type θ update according to

$$f[\psi]_\theta(y_\theta,(m,a)) \coloneqq \begin{cases} \psi_\theta(y_\theta)\cdot\gamma\cdot\bar{\psi}_R(a|m) & \text{if } m = s_\theta(y_\theta) \\ 0 & \text{otherwise.} \end{cases}$$

Here a typeθ sender with history y_θ today must play the message $s_\theta(y_\theta)$ tomorrow. The fraction of typeθ receivers who will have history $(y_\theta,(m,a))$ tomorrow with $m = s_\theta(y_\theta)$ is given by the product of three terms: $\psi_\theta(y_\theta)$ is the fraction of typeθ senders who have history y_θ today, γ is the probability that each such sender survives until tomorrow, and $\bar{\psi}_R(a|m)$ is the probability that the matched receiver will respond to message m with action a.

3.5 Steady States

A state ψ^* such that $f[\psi^*] = \psi^*$ is called a *steady state*. The learning system is stationary at a steady state. Its distribution over histories does not change with time, so neither do the induced behavioral strategies of the agents. Denote the set of all steady states with regular priors $g = (g_S, g_R)$, patience $\delta \in [0, 1)$, and survival chance $\gamma \in [0, 1)$ as $\Psi^*(g, \delta, \gamma)$. Note the dependence of this set on the prior as well as on the parameters δ and γ: In games with multiple equilibria, which ones are selected can depend on the prior.

Proposition 4. $\Psi^*(g, \delta, \gamma)$ *is non-empty and compact in the* ℓ_1 *norm.*

The proof is in the Online Appendix. Intuitively, if lifetimes are finite, then set of histories is finite, so the set of states is of finite dimension. Here the one-period-forward map f is continuous, so the usual version of Brower's fixed-point theorem applies. With exponential lifetimes, very old agents are rare, so truncating the agent's lifetimes at some large T yields a good approximation. Instead of using these approximations directly, our proof shows that under the ℓ_1 norm f is continuous and the feasible states form a compact locally convex Hausdorff space, so we can use a fixed-point theorem for that domain. Note that in the steady state, the information lost when agents exit the system exactly balances the information agents gain through learning.

Let $\bar{\Psi}^*(g, \delta, \gamma)$ denote the set of strategy profiles induced by the steady states in $\Psi^*(g, \delta, \gamma)$.

Definition 9. For each $\delta \in [0, 1)$, say a strategy profile π^* is *δ-stable under g* if there is a sequence $\gamma_k \to 1$ and an associated sequence of steady state strategy profiles, $\pi^{(k)} \in \bar{\Psi}^*(g, \delta, \gamma_k)$, such that $\pi^{(k)} \to \pi^*$. Strategy profile π^* 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 δ_k-stable and $\pi^{(k)} \to \pi^*$. Say π^* is *patiently stable* if it is patiently stable under some regular prior g.

Fix any regular prior g and $\delta \in [0, 1)$. Since each $\bar{\Psi}^*(g, \delta, \gamma)$ is non-empty, to any sequence $\gamma_k \to 1$ we may associate a sequence of steady states strategy profiles $\pi^{(k)} \in \bar{\Psi}^*(g, \delta, \gamma_k)$. This sequence of strategy profiles has a convergent subsequence since the space of behavioral strategy profiles may be viewed as a compact subset of finite-dimensional Euclidean space. This shows δ-stable strategy profiles always exist for every regular prior g. The same arguments establish that patiently strategy profiles always exist for every regular prior g.

Heuristically speaking, patiently stable strategy profiles are the limits of learning outcomes when agents become infinitely patient and long lived, but note the order of limits involved: first we send γ to 1 holding δ fixed, and then send δ to 1. As in past work on steady state learning (Fudenberg and Levine, 1993, 2006), the reason for this is to ensure that when agents have enough data they eventually stop experimenting and play myopic best responses.

3.6 An alternative description of the steady states

Given any state ψ, we can compute $\bar{\psi}_S(\cdot|\theta)$, the behavioral strategy of type θ induced by ψ, directly from its definition in (2). We establish below an alternative expression for $\bar{\psi}_S(\cdot|\theta)$ that holds in steady states; we use this alternative in the proofs of Lemma 3.

Suppose the receiver population plays $\bar{\psi}_R$ and a newborn type θ is matched with an i.i.d. draw from the receiver population each period. We may equivalently think of the newborn sender drawing (but not observing) a response sequence $\mathfrak{a}$ at birth, which then governs how her opponents react to her messages throughout her lifetime. The distribution over the set of response sequences $\times_{m \in M} (\mathrm{BR}(\Delta(\Theta), m)^\infty)$ that makes these two situations equivalent is denoted ν_{ψ_R}, which is defined on finite truncations as

$$\nu_{\psi_R}((a_{1,m}, a_{2,m}, ..., a_{L,m})_{m \in M}) \coloneqq \prod_{m \in M} \prod_{j=1}^{L} \bar{\psi}_R(a_{j,m}|m),$$

then extended to the infinite Cartesian product.

From the perspective of a receiver who matches with a type θ, there is $(1-\gamma)\gamma^t$ chance that this sender is of age t. So, the probability density function for encountering "an age t sender who drew $\mathfrak{a}$ at birth" is $(1-\gamma)\gamma^t \cdot d\nu_{\psi_R}(\mathfrak{a})$. This sender sends message $s_\theta(y_\theta^t(\mathfrak{a}))$ in period t, where $y_\theta^t(\mathfrak{a})$ was defined in Definition 8. Therefore, the probability that this randomly matched type θ plays message m is

$$\int (1-\gamma) \cdot \sum_{t=0}^{\infty} \gamma^t \cdot 1\{s_\theta(y_\theta^t(\mathfrak{a})) = m\} d\nu_{\psi_R}(\mathfrak{a}).$$

We restate this conclusion as a lemma to facilitate later references.

Lemma 2. *If ψ is a steady state, then*

$$\bar{\psi}_S(m|\theta) = \int (1-\gamma) \cdot \sum_{t=0}^{\infty} \gamma^t \cdot 1\{s_\theta(y_\theta^t(\mathfrak{a})) = m\} d\nu_{\psi_R}(\mathfrak{a}).$$

4 Characterizing the Steady States

4.1 Steady States for Fixed δ

When γ is small, agents expect to live only a short time, so their prior beliefs drive their play. When γ is near 1, agents correctly learn the consequences of the strategies they play frequently, but for a fixed patience level they may choose to rarely or never experiment, and so can maintain incorrect beliefs about the consequences of strategies that they do not play.

The next result formally states this, which parallels Fudenberg and Levine (1993)'s result that δ-stable strategy profiles are self-confirming equilibria.

Theorem 1. *Suppose strategy profile π^* is δ-stable under a regular prior. Then for every type θ and message m with $\pi_S^*(m|\theta) > 0$, m is a best response to some $\pi_R \in \Pi_R$ for type θ, and furthermore $\pi_R(\cdot|m) = \pi_R^*(\cdot|m)$. Also, for any message m such that $\pi_S^*(m|\theta) > 0$ for at least one type θ, $\pi_R^*(\cdot|m)$ is supported on pure best responses to the Bayesian belief generated by π_S^* after m.*

We prove Theorem 1 in the Online Appendix. The idea of the proof is the following: If message m has positive probability in the limit, then it is played many times by the senders, so they eventually learn correct posterior distribution for θ given m. As the receivers have no incentive to experiment, their actions after m will be a best response to this correct posterior belief. For the senders, suppose $\pi_S^*(m|\theta) > 0$, but m is not a best response for type θ to any $\pi_R \in \Pi_R$ that matches $\pi_R^*(\cdot|m)$. Then there exists $\xi > 0$ such that m is not a ξ best response to any strategy that differs by no more than ξ from $\pi_R^*(\cdot|m)$ after m. Yet, by the law of large numbers and the Diaconis and Freedman (1990) result that with non-doctrinaire priors the posteriors converge to the empirical distribution at a rate that depends only on the sample size, with high probability θ's posterior belief about receiver's strategy after m is ξ-close to $\pi_R^*(\cdot|m)$. So when a type θ who has played m many times chooses to play it again, she is not doing so to maximize her current period's expected payoff. This implies that type θ has persistent option value for message m, which contradicts the fact that this option value must converge to 0 with the sample size.

Remark 2. This theorem says that each type is playing a best response to a belief about the receiver's play that is (i) correct on the equilibrium path and (ii) assigns probability 0 to dominated replies by the receiver, and that the receiver is playing a best response to the aggregate play of the senders. Thus the δ-stable outcomes are a version of Dekel, Fudenberg, and Levine (1999)'s rationalizable self-confirming equilibrium where different types of sender are allowed to have different beliefs.[10]

Example 3. Consider the following game:

m_1	a_1	a_2
θ_1	2, 0	-1, 0
θ_2	-1, 0	2, 0

m_2	a_1	a_2
θ_1	0,0	0, 0
θ_2	0,0	0,0

[10] Dekel, Fudenberg, and Levine (2004) define type-heterogeneous self-confirming equilibrium in static Bayesian games. To extend their definition to signalling games, we can define the "signal functions" $y_i(a, \theta)$ from that paper to respect the extensive form of the game. See also Fudenberg and Kamada (2016).

Note the receiver is indifferent between all responses. Fix any regular prior g_R for the receiver and let the sender's prior $g_S^{(m_1)}$ be given by a Dirichlet distribution with weights 1 and 3 on a_1 and a_2 respectively. Fix any regular prior $g_S^{(m_2)}$. We claim that it is δ-stable when $\delta = 0$ for both types of senders to play m_2 and for the receiver to play a_1 after every message, which is a type-heterogeneous rationalizable self-confirming equilibrium. However, the behavior of "pooling on m_2" cannot occur even in the usual self-confirming equilibrium, where both types of the sender must hold the same beliefs about the receiver's response to m_1. *A fortiori*, this pooling behavior cannot occur in a Nash equilibrium.

To establish this claim, note that since $\delta = 0$ each sender plays a myopically optimal message after every history. For any γ, there is a steady state where the receiver's policy responds to every message with a_1 after every history, type θ_1 plays m_2 after every history and never updates her prior belief about how receivers react to m_1, while type θ_2 with fewer than 6 periods of experience play m_1 but switch to playing m_2 forever starting at age 7. The behavior of θ_2 comes from the fact that after k periods of playing m_1 and seeing a response of a_1 every period, the sender's expected payoff from playing m_1 next period is

$$\frac{1+k}{4+k}(-1) + \frac{3}{4+k}(2).$$

This expression is positive when $0 \leq k \leq 5$ but negative when $k = 6$. The fraction of type θ_2 aged 6 and below approaches 0 as $\gamma \to 1$, hence we have constructed a sequence of steady state strategy profiles converging to the strategy profile where the two types of senders both play m_2.

This example illustrates that even though all types of senders start with the same prior g_S, their learning is endogenously determined by their play, which is in turn determined by their payoff structures. Since the two different types of senders play differently, their beliefs regarding how the receiver will react to m_1 eventually diverge. ⧫

4.2 Patiently Stable Strategy Profiles are Nash Equilibria

Theorem 2. *Every patiently stable strategy profile is a Nash equilibrium.*

We follow the proof strategy of Fudenberg and Levine (1993), which derived a contradiction via excess option values. The value function of the dynamic optimization problem evaluated at a sufficiently long history should be not much higher than the expected current period payoff of the strategy played at that history — that is, the option value of the agents goes to 0. But if the steady state is a non-Nash outcome, then a non-negligible fraction of the agents of some population i would gain by deviating to some strategy $s_i^{'}$. Moreover, these agents' prior assigned a non-negligible chance to $s_i^{'}$ yielding a strictly higher payoff, and their observations are unlikely to falsely convince them that it does not. Thus if the agents are patient enough they perceive an option value to experimenting with $s_i^{'}$, a contradiction.

In Fudenberg and Levine (1993), this argument relies on the finite lifetime only insofar as to ensure "almost all" histories are long enough, by picking a large enough lifetime. We can achieve the analogous effect in the infinite-horizon model by picking γ close to 1. The proof details may be found in the Online Appendix.

4.3 Patiently Stable Strategy Profiles are Type-Compatible Equilibria

In this subsection, we prove that all patiently stable strategy profiles are type-compatible equilibria.

The next lemma extends the conclusion of Lemma 2 to apply to steady states instead of response sequences. It shows that if $\theta^{'}$ is more compatible with $m^{'}$ than $\theta^{''}$, then in every steady state strategy profile, $\theta^{'}$ must play $m^{'}$ at least as much as $\theta^{''}$ does.

Lemma 3. *Suppose there are types $\theta^{'}, \theta^{''}$ and message $m^{'}$ such that $\theta^{'} \succ_{m^{'}} \theta^{''}$.*

Then for any regular prior g, parameters $\delta, \gamma \in [0, 1)$ and any steady state $\psi \in \Psi^*(g, \delta, \gamma)$, we have $\bar{\psi}_S(m^{'}|\theta^{'}) \geq \bar{\psi}_S(m^{'}|\theta^{''})$.

Proof. By Lemma 2, we may rewrite

$$\bar{\psi}_S(m^{'}|\theta^{'}) = \int (1-\gamma) \cdot \sum_{t=0}^{\infty} \gamma^t \cdot 1\{s_{\theta^{'}}(y_{\theta^{'}}^t(\mathfrak{a})) = m^{'}\} d\nu(\mathfrak{a})$$

and

$$\bar{\psi}_S(m^{'}|\theta^{''}) = \int (1-\gamma) \cdot \sum_{t=0}^{\infty} \gamma^t \cdot 1\{s_{\theta^{''}}(y_{\theta^{''}}^t(\mathfrak{a})) = m^{'}\} d\nu(\mathfrak{a}).$$

So, it suffices to show

$$\sum_{t=0}^{\infty} \gamma^t \cdot 1\{s_{\theta^{'}}(y_{\theta^{'}}^t(\mathfrak{a})) = m^{'}\} \geq \sum_{t=0}^{\infty} \gamma^t \cdot 1\{s_{\theta^{''}}(y_{\theta^{''}}^t(\mathfrak{a})) = m^{'}\}.$$

for every response sequence $\mathfrak{a}$. But this has been established by Lemma 1. ☐

The next lemma says that given a strategy profile π° where a type's best possible payoff to message m' exceeds her payoff under the profile, in any steady state strategy profile ϵ-close to π° this type will experiment "many times" with $m^{'}$, provided expected lifetimes are long and agents are patient. Its proof appears in the Online Appendix.

Lemma 4. *Fix a regular prior g and a strategy profile π° where for some type θ' and message m', $\theta^{'} \in J(m^{'}, \pi^\circ)$.*

There exist number ϵ and functions $\delta(N)$ and $\gamma(N,\delta)$, all valued in $(0,1)$, such that whenever:

- $\delta \geq \delta(N)$, $\gamma \geq \gamma(N,\delta)$
- $\psi \in \Psi^*(g,\delta,\gamma)$
- *$\bar{\psi}$ is no further away than ϵ from π° in L_1 norm,*

we have $\bar{\psi}_S(m^{'}|\theta^{'}) \geq (1-\gamma)\cdot N$.

To gain an intuition for this result, suppose that not only is $m^{'}$ not equilibrium dominated in π°, but furthermore that $m^{'}$ can lead to the highest signalling game payoff for type $\theta^{'}$ under some receiver response $a^{'} \in \mathrm{BR}(\Delta(\Theta), m^{'})$. Holding the prior constant, the Gittins index of message m approaches its highest possible payoff as the sender becomes infinitely patient. Therefore, for every $N \in \mathbb{N}$, when γ and δ are close enough to 1, a newborn type $\theta^{'}$ will play $m^{'}$ in each of the first N periods of her life, regardless of what responses she receives during that time. These N periods account for roughly $(1-\gamma)\cdot N$ fraction of her life. Moreover, even if $m^{'}$ does not lead to the highest potential payoff in the signalling game, long-lived players will have a good estimate of their steady state payoff. So, type θ' will still play any $m^{'}$ that is equilibrium undominated in strategy profile π° at least N times in any steady states that are sufficiently close to π°, though these N periods may not occur at the beginning of her life.

The rate condition implicit in the lemma is important in what follows: Since senders and receivers have the same survival probabilities, a receiver who lives to his expected life length of $1/(1-\gamma)$ will have seen on average at least $\lambda(\theta')N$ instances of type $\theta^{'}$ playing $m^{'}$. As γ grows towards 1, the steady state frequency of experimentation with $m^{'}$ may shrink to 0 for type $\theta^{'}$, but this is offset by the fact the typical receiver lives longer, so we have a constant lower bound on the instances where $\theta^{'}$ plays $m^{'}$ that the receiver sees on average. This will let us apply Theorem 2 of Fudenberg, He, and Imhof (2016) to conclude that a typical receiver believes that $\theta^{'}$ plays m' more frequently than less compatible types do.

Recall from Section 2 the set of compatible beliefs after message $m^{'}$,

$$P(m^{'},\pi^*) := \left\{ p \in \Delta(UD^{-1}(m^{'})) : \ \frac{p(\theta^{''})}{p(\theta^{'})} \leq \frac{\lambda(\theta^{''})}{\lambda(\theta^{'})} \ \text{ whenever } \begin{array}{c} \text{(i) } \theta^{'} \succ_{m^{'}} \theta^{''} \\ \text{and} \\ \text{(ii) } \theta^{'} \in J(m^{'},\pi^*) \end{array} \right\}.$$

We now show that the receiver best responds to $P(m^{'},\pi^*)$ in every patiently stable strategy profile, even when $m^{'}$ is off-path.

Theorem 3. *If g is regular, then for $\pi^* = (\pi_S^*, \pi_R^*)$ to be patiently stable under g it is necessary that $\pi_R^*(\cdot|m^{'}) \in \Delta(BR(P(m^{'}, \pi^*), m^{'})$ for every $m^{'} \in M$.*

In outline the proof has three parts: Lemma 3 shows that types that are more compatible with m' play it more often, Lemma 4 says that types for whom m' is not equilibrium dominated will play it "many times," and finally the "many times" here is sufficiently large that most receivers correctly believe that more compatible types play m' more than less compatible types do, so their posterior odds ratio for more versus less compatible types exceeds the prior ratio.

Proof. Suppose $\hat{a} \notin \mathrm{BR}(P(m^{'}, \pi^*), m^{'})$. We will show that $\pi_R^*(\hat{a}|m^{'}) = 0$ if (π_S^*, π_R^*) is patiently stable. As a first step we will show that there is $\bar{\xi} \in (0, 1)$ such that $\hat{a} \notin \mathrm{BR}(P_\xi(m^{'}, \pi^*), m^{'})$ whenever $\xi < \bar{\xi}$, where we define the "ξ-approximation" to $P(m^{'}, \pi^*)$,

$$P_\xi(m^{'}, \pi^*) := \left\{ p \in \Delta(UD^{-1}(m^{'})) : \frac{p(\theta^{''})}{p(\theta^{'})} \le (1+\xi)\frac{\lambda(\theta^{''})}{\lambda(\theta^{'})} \text{ whenever } \begin{matrix} \text{(i) } \theta^{'} \succ_{m'} \theta^{''} \\ \text{and} \\ \text{(ii) } \theta^{'} \in J(m^{'}, \pi^*) \end{matrix} \right\}.$$

It is clear that each $P_\xi(m^{'}, \pi^*)$ as well as $P(m^{'}, \pi^*)$ itself is closed. The approximations converge down in terms of set inclusion, $P_\xi(m^{'}, \pi^*) \to P(m^{'}, \pi^*)$ as $\xi \to 0$.

If for all $\bar{\xi} > 0$ there is $\xi < \bar{\xi}$ s.t. $\hat{a} \in \mathrm{BR}(P_\xi(m^{'}, \pi^*), m^{'})$,then because the BR correspondence has closed graph, we would have the $\hat{a} \in P(m^{'}, \pi^*)$. So, there exists $\bar{\xi} > 0$ such that $\hat{a} \notin \mathrm{BR}(P_\xi(m^{'}, \pi^*), m^{'})$ for every $\xi \in (0, \bar{\xi})$.

Let some $\xi \in (0, \bar{\xi})$ be fixed. Now apply[11] Theorem 2 from Fudenberg, He, and Imhof (2016), with $\mu = g_R^{(\theta)}$, $\nu = g_R^{(\theta^{'})}$ and $\epsilon = \xi$. We obtain some $N \in \mathbb{N}$ such that whenever a receiver faces any sender strategy $\bar{\psi}_S$ satisfying

$$\bar{\psi}_S(m^{'}|\theta^{'}) \ge \psi_S(m^{'}|\theta^{''}) \tag{3}$$

over n periods, such that

$$n \cdot \bar{\psi}_S(m^{'}|\theta^{'}) \ge N, \tag{4}$$

then there is at least $1 - \xi$ chance that his posterior belief p when seeing $m^{'}$ again satisfies $\dfrac{p(\theta^{''})}{p(\theta^{'})} \le (1+\xi)\dfrac{\lambda(\theta^{''})}{\lambda(\theta^{'})}$ at the end of n periods.

[11] We may appeal to that theorem since g is a regular prior. Theorem 2 of Fudenberg, He, and Imhof (2016) assumes that the priors $g_R^{(\theta)}$ and $g_R^{(\theta^{'})}$ have full support over all strategies, including those that put nonzero probabilities on dominated messages. However, it is straightforward to generalize that result to the case where $g_R^{(\theta)}$ and $g_R^{(\theta^{'})}$ only have full support over $\Pi_S^{(\theta)}$ and $\Pi_S^{(\theta^{'})}$ respectively (and satisfy the conditions in Definition 6).

Next, take a sequence of δ-stable strategies converging to π^*, say $\bar{\psi}^{(k)} \to \pi^*$ where $\bar{\psi}^{(k)}$ is δ_k-stable with $\delta_k \to 1$. Each of these δ-stable strategies can be written as a limit of steady state strategy profiles. That is, we have the array of steady states $\psi^{(k,j)} \in \Psi^*(g, \delta_k, \gamma_{k,j})$, where $\gamma_{k,j} \to 1$ and $\bar{\psi}^{(k)} = \lim_{j\to\infty} \bar{\psi}^{(k,j)}$. We will argue that for large enough k and j, $\bar{\psi}_S^{(k,j)}$ will satisfy (3) and (4), provided $\theta' \succ_{m'} \theta''$ and $\theta' \in J(m', \pi^*)$.

In fact, by Lemma 3, $\theta' \succ_{m'} \theta''$ implies (3) is satisfied in every steady state strategy profile. To see that (4) eventually holds too, find $\epsilon, \delta(N/\xi)$, and $\gamma(N/\xi, \delta)$ by Lemma 4. Now we diagonalize the array $(\bar{\psi}^{(k,j)})$, finding $j_0 \in \mathbb{N}$ and function $k(j)$ such that whenever $j > j_0$ and $k > k(j)$, we get $\bar{\psi}^{(j,k)}$ is no more than ϵ away from from π^*. Also, we may define j_0 and $k(j)$ so that $j \geq j_0$ implies $\delta_j \geq \delta(N/\xi)$, and also so that $k > k(j)$ implies $\gamma_{j,k} > \gamma(N/\xi, \delta_j)$.

Therefore, whenever $j > j_0$ and $k > k(j)$, a receiver who faces the sender strategy $\bar{\psi}_S^{(j,k)}$ for more than $\frac{\xi}{1-\gamma_{j,k}}$ periods has at least $1-\xi$ chance of seeing a sample that leads to a posterior belief with $\dfrac{p(\theta'')}{p(\theta')} \leq (1+\xi)\dfrac{\lambda(\theta'')}{\lambda(\theta')}$. In other words, after $\frac{\xi}{1-\gamma_{j,k}}$ periods, there is at least $1-\xi$ chance that such a receiver has belief in $P_\xi(m', \pi^*)$ as to the type of someone who sends m'. As the fraction of receivers whose age is at most $\frac{\xi}{1-\gamma_{j,k}}$ is $1 - \left(\gamma_{j,k}^{\frac{\xi}{1-\gamma_{j,k}}}\right) \approx \xi$, we have therefore shown whenever $j > j_0, k > k(j)$, we have

$$\bar{\psi}_R^{(j,k)}(y_R : p(\cdot|m; y_R) \in P_\xi(m, \pi^*)) \geq 1 - 2\xi.$$

Here, $p(\cdot|m; y_R) \in \Delta(\Theta)$ stands for the receiver's posterior belief on the sender's type, after history y_R and after seeing message m today. The term "2ξ" comes from ξ of the receivers not being older than age $\frac{\xi}{1-\gamma_{j,k}}$, while ξ of the receivers older than $\frac{\xi}{1-\gamma_{j,k}}$ may have an exceptional history. But $\hat{a} \notin \text{BR}(P_\xi(m, \pi^*), m)$, so $\bar{\psi}_R^{(j,k)}(\hat{a}|m) < 2\xi$ whenever $j > j_0$, $k > j(k)$, and thus $\pi_R^*(\hat{a}|m) < 2\xi$. As the choice of $\xi \in (0, \bar{\xi})$ was arbitrary, we conclude that $\pi_R^*(\hat{a}|m) = 0$. □

Theorem 4. *Under any regular prior, patiently stable strategy profiles exist and must be type-compatible equilibria.*

Proof. This follows from Proposition 4, Theorem 2, and Theorem 3. □

4.4 An Additional Implication of Patient Stability

In generic games, pure strategy equilibria must satisfy a stronger condition to be patiently stable: the set $P(m, \pi^*)$ of allowed beliefs after an out-of-equilibrium message can be reduced to the smaller set

$$\widetilde{P}(m,\pi^*) := \left\{ p \in \Delta(\widetilde{J}(m',\pi^*)) : \ \frac{p(\theta'')}{p(\theta')} \leq \frac{\lambda(\theta'')}{\lambda(\theta')} \ \text{ whenever } \theta' \succ_{m'} \theta'' \right\},$$

where

$$\widetilde{J}(m,\pi^*) := \left\{ \theta \in \Theta : \max_{a \in \mathrm{BR}(\Delta(\Theta),m)} u_S(\theta,m,a) \geq u_S(\theta;\pi^*) \right\}$$

is the set of types for which *some* best response to message m is at least as good as their payoff under π^*. If $\widetilde{J}(m,\pi^*) = \emptyset$, then define $\widetilde{P}(m,\pi^*) := \Delta(\Theta)$. Note that $\widetilde{P}$, unlike P, assigns probability 0 to equilibrium dominated types, which is the belief restriction of the Intuitive Criterion.

Definition 10. A Nash equilibrium π^* is *on-path strict for the receiver* if for every on-path message m^*, $\pi_R^*(a^*|m^*) = 1$ for some $a^* \in A$ and $u_R(m^*,a^*,\pi_S^*) > \max_{a \neq a^*} u_R(m^*,a,\pi_S^*)$.

Of course, the receiver cannot have strict ex-ante preferences over play at unreached information sets; this condition is called "on-path strict" because we do not place restrictions on receiver's incentives after off-path messages. In generic signalling games, all pure-strategy equilibria are on-path strict for the receiver, but the same is not true for mixed-strategy equilibria.

Definition 11. A Nash equilibrium π^* is a *strong type-compatible equilibrium* if it is on-path strict for the receiver and, for every off-path message m' the receiver's strategy $\pi_R^*(\cdot|m')$ satisfies the *strong compatibility criterion*,

$$\pi_R^*(\cdot|m') \in \Delta(\mathrm{BR}(\widetilde{P}(m',\pi^*),m')).$$

It is immediate that every strong TCE is a TCE, since the latter places less stringent restrictions on receiver's off-path behavior and does not require on-path strictness for receiver. It is also immediate that every strong TCE satisfies the Intuitive Criterion.

Theorem 5. *Suppose π^* is on-path strict for the receiver and patiently stable. Then it is a strong type-compatible equilibrium.*

The proof of this theorem appears in Appendix B.2. Here we provide an outline of the arguments.

We first show there is a sequence of steady state strategy profiles $\bar{\psi}^{(k)} \in \Psi^*(g,\delta_k,\gamma_k)$ with $\gamma_k \to 1$ and $\bar{\psi}^{(k)} \to \pi^*$, where the rate of on-path convergence of $\bar{\psi}_R^{(k)}$ to π_R^* is of order $(1-\gamma_k)$. That is, there exists some $N^{\text{wrong}} \in \mathbb{N}$ so that $\bar{\psi}_R^{(k)}(\cdot|m^*)$ plays actions other than the equilibrium response to m^* less than $(1-\gamma_k)\cdot N^{\text{wrong}}$ of the time for each k and each on-path message m^*. Next, we consider a type θ^{D} for whom m^* equilibrium dominates the off-path

m'. We show the probability that a very patient θ^{D} ever switches away from m^* after trying it for the first time is bounded by a multiple of the weight that $\bar{\psi}_R^{(k)}(\cdot|m^*)$ assigns to non-equilibrium responses to m^*. Together with the fact that $\bar{\psi}_R^{(k)}(\cdot|m^*)$ converges to $\pi_R^*(\cdot|m^*)$ at the rate of $(1-\gamma_k)$, this lets us find some $N \in \mathbb{N}$ so that $\bar{\psi}_S^{(k)}(m'|\theta^{\mathrm{D}}) < N \cdot (1-\gamma_k)$ for every k. On the other hand, for each $\theta' \in \tilde{J}(m', \pi^*)$, Lemma 4 shows for any $N' \in \mathbb{N}$, for large enough k we will have $\bar{\psi}_S^{(k)}(m'|\theta') > N' \cdot (1-\gamma_k)$. So by choosing N' sufficiently large relative to N, we can show that $\lim_{k\to\infty} \frac{\bar{\psi}_S^{(k)}(m'|\theta')}{\bar{\psi}_S^{(k)}(m'|\theta^{\mathrm{D}})} = \infty$. Finally, we apply Theorem 2 of Fudenberg, He, and Imhof (2016) to deduce that a typical receiver has enough data to conclude someone who sends m' is arbitrarily more likely to be θ' than θ^{D}, thus eliminating completely any belief in equilibrium dominated types after m'.

Remark 3. As noted by Fudenberg and Kreps (1988) and Sobel, Stole, and Zapater (1990), it seems "intuitive" that learning and rational experimentation should lead receivers to assign probability 0 to types that are equilibrium dominated, so it might seem surprising that this theorem needs the additional assumption that the equilibrium is on-path strict for the receiver. However, in our model senders start out initially uncertain about the receivers' play, and so even types for whom a message is equilibrium dominated might initially experiment with it. Showing that these experiments do not lead to "perverse" responses by the receivers requires some arguments about the *relative* probabilities with which equilibrium-dominated types and non-equilibrium-dominated types play off-path messages. When the equilibrium involves on-path receiver randomization, a non-trivial fraction of receivers could play an action that a type finds strictly worse than her worst payoff under an off-path message. In this case, we do not see how to show that the probability she ever switches away from her equilibrium message tends to 0 with patience, since the event of seeing a large number of these unfavorable responses in a row has probability bounded away from 0 even when the receiver population plays exactly their equilibrium strategy. However, we do not have a counterexample to show that the conclusion of the theorem fails without on-path strictness for the receiver.

Example 4. In the following modified beer-quiche game, we still have $\lambda(\theta_{\text{strong}}) = 0.9$ but the payoffs of fighting a θ_{weak} who drinks beer have been substantially increased:

beer	fight	not fight
$\theta_{\texttt{strong}}$	1,0	3,1
$\theta_{\texttt{weak}}$	0,1000	2,0

quiche	fight	not fight
$\theta_{\texttt{strong}}$	0,0	2,1
$\theta_{\texttt{weak}}$	1,1	3,0

Consider the Nash equilibrium of "both types eat quiche", supported by receiver fighting anyone who drinks beer. Since fight is a best response to the prior λ, it is not ruled out by the compatibility criterion.

This pooling equilibrium is on-path strict for the receiver, because receiver has a strict preference for "not fight" at the only on-path message, "quiche". Moreover, it is not a strong TCE, because $\widetilde{J}(\text{beer}, \pi^*) = \{\text{strong}\}$ implies in every TCE the receiver must assign probability 1 to sender being θ_{strong} after seeing "beer", so "not fight" is the only allowable off-path response by strong compatibility. Thus Theorem 5 implies that this equilibrium is not patiently stable.. ♦

4.5 A Sufficient Condition for Patient Stability

We now show that under some additional strictness conditions, every uniform TCE is patiently stable for some regular prior.[12] We prove the following result in Appendix B.3.

Definition 12. A *quasi-strict uniform TCE* π^* is a uniform TCE that is on-path strict for the receiver, strict for the sender (that is, every type strictly prefers its equilibrium message to any other), and satisfies $u_S(\theta; \pi^*) > u_S(\theta, m', a)$ for all θ, all off-path messages m' and all $a \in \text{BR}(\hat{P}(m'), m')$.

Theorem 6. *If π^* is a quasi-strict uniform type-compatible equilibrium, then it is path-equivalent to a patiently stable strategy profile.*

To prove Theorem 6, we construct a Dirichlet prior for the receiver such that in every steady state, the receiver has a high probability of holding a belief in $\hat{P}(m')$ after m'.[13] To do this, we construct the prior g_R so that whenever $\theta' \succ_{m'} \theta''$, g_R assigns much greater prior weight to θ' playing m' than to θ'' playing m'. In the absence of data, the receiver strongly believes that $p(\theta''|m')/p(\theta'|m') \leq \lambda(\theta'')/\lambda(\theta')$. This strong prior belief can only be overturned by a very large number of observations to the contrary. But if the receiver has a very large number of observations, then since θ experiments more with m than θ' by Lemma 3, the law of large numbers implies this large sample is unlikely to lead the receiver to have a belief outside of $\hat{P}(m')$. So, we can ensure that sufficiently long-lived receivers play a best response to $\hat{P}(m')$ after the off-path m', with high probability. Also, provided that the sender population is playing close enough to π_S^*, the law of large numbers implies that after every message m on-path in π^*, a receiver with enough data is likely to have a belief close to the Bayesian belief after m assigned by π^*. Coupled with the fact that π^* is on-path

[12] Note that the steady state of our learning model depends on the priors even in static games like battle of the sexes.

[13] 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 $g_R^{(\theta)}$ and $g_S^{(m)}$ 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.

strict for the receiver, this lets us conclude that long-lived receivers play $\pi_R^*(\cdot|m)$ after every on-path m with high probability.

Finally, we specify a sender prior g_S that is highly confident and correct about the receiver's response to on-path messages, and is also confident that the receiver responds to off-path messages m' with actions in $BR(\hat{P}(m'), m')$. The fact that sender's option value for experimentation eventually goes to 0, together with the assumption that all of receiver's best responses to $\hat{P}(m^{'})$ lead to strictly less than the equilibrium payoff for every type, shows sufficiently long-lived senders behaves similar to π_S^* when the receiver population plays close to $\pi_R^*(\cdot|m^{'})$ after every on-path $m^{'}$ and plays a best response to $\hat{P}(m^{'})$ after every off-path $m^{'}$.

This last step uses the assumption that π^* is strict for the sender. If m^* were only weakly optimal in π^*, there could be receiver strategies arbitrarily close to π^* that make some other message $m^{'} \neq m^*$ strictly optimal for θ. In that case, we cannot rule out that θ will play m' forever with non-negligible probability in some steady states where the receiver population plays close to π_R^*.

5 Comparison to Other Equilibrium Refinements

This section compares compatibility and type-compatible equilibrium to other equilibrium refinement ideas in the literature.

We begin by relating compatibility to a form of iterated dominance in the ex-ante strategic form of the game, where the sender chooses a message as function of her type. We show that every sender strategy that specifies playing message $m^{'}$ 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 Π_R: if the less compatible $\theta^{''}$ does not have a profitable deviation, then the more compatible type strictly prefers deviating to $m^{'}$.

Proposition 5. *Suppose $\theta' \succ_{m'} \theta^{''}$. Then any ex-ante strategy of the sender π_S with $\pi_S(m^{'}|\theta^{''}) > 0$ but $\pi_S(m^{'}|\theta^{'}) < 1$ is removed by strict dominance once the receiver is restricted to using strategies in Π_R.*

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

We next relate TCE and strong TCE to the Intuitive Criterion. Note first that the quiche-pooling equilibrium in Example 4, which is a TCE but not a strong TCE, fails the Intuitive Criterion because beer is equilibrium dominated for θ_{weak}. The next example has an equilibrium that satisfies the Intuitive Criterion but is not a TCE, so that these solution concepts are not nested. As noted above, every strong TCE satisfies the Intuitive Criterion. Since they are also TCE, we see that the set of strong TCEs is strictly smaller than the set of equilibria that pass the Intuitive Criterion.

Example 5. 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:

m_1	a_1	a_2
θ_1	4, 1	0, 0
θ_2	6, 0	2, 1

m_2	a_1	a_2
θ_1	7, 1	3, 0
θ_2	7, 0	3, 1

Against any receiver strategy, the two types θ_1 and θ_2 get the same payoffs from m_2, but θ_2 gets strictly higher payoffs than θ_1 from m_1. So, whenever m_2 is weakly optimal for θ_2, it is strictly optimal for θ_1, so $\theta_1 \succ_{m_2} \theta_2$.

Consider now the Nash equilibrium in which the types pool on m_1, i.e. $\pi_S^*(m_1|\theta_1) = \pi_S^*(m_1|\theta_2) = 1$, $\pi_R^*(a_1|m_1) = 1$, and $\pi_R^*(a_2|m_2) = 1$. Since $\theta_1 \in J(m_2, \pi^*)$, the compatibility criterion requires that every action played with positive probability in $\pi_R^*(\cdot|m_2)$ best responds to some belief p about sender's type satisfying $\frac{p(\theta_2)}{p(\theta_1)} \leq \frac{\lambda(\theta_2)}{\lambda(\theta_1)} = \frac{1}{3}$. But action a_2 does not best respond to any such belief, so π^* is not a type-compatible equilibrium. On the other hand, it passes the Intuitive Criterion because the off-path message m_2 is not equilibrium dominated for either type. ♦

Now we compare divine equilibrium with type-compatible equilibrium and uniform type-compatible equilibrium. For a strategy profile π^*, let

$$D(\theta, m; \pi^*) := \{\alpha \in \text{MBR}(m) \text{ s.t. } u_S(\theta; \pi^*) < u_S(\theta, m, \alpha)\}$$

be the subset of mixed best responses to m that would make type θ strictly prefer deviating from the strategy $\pi_S^*(\cdot|\theta)$. Similarly let

$$D^\circ(\theta, m; \pi^*) := \{\alpha \in \text{MBR}(m) \text{ s.t. } u_S(\theta; \pi^*) = u_S(\theta, m, \alpha)\}$$

be the set of mixed best responses that would make θ indifferent to deviating.

Proposition 6.

(a). If π^* is a Nash equilibrium where $m^{'}$ is off-path and furthermore $\theta^{'} \succ_{m'} \theta^{''}$, then $D(\theta^{''}, m^{'}; \pi^*) \cup D^{\circ}(\theta^{''}, m^{'}; \pi^*) \subseteq D(\theta^{'}, m^{'}; \pi^*)$.

(b). Every divine equilibrium is a type-compatible equilibrium.

Proof. To show (a), note first that if $D(\theta^{''}, m^{'}; \pi^*) \cup D^{\circ}(\theta^{''}, m^{'}; \pi^*) = \emptyset$ the conclusion holds vacuously. If $D(\theta^{''}, m^{'}; \pi^*) \cup D^{\circ}(\theta^{''}, m^{'}; \pi^*)$is not empty, take any $\alpha^{'} \in D(\theta^{''}, m^{'}; \pi^*) \cup D^{\circ}(\theta^{''}, m^{'}; \pi^*)$ and define $\pi_R^{'} \in \Pi_R$ by $\pi_R^{'}(\cdot|m^{'}) = \alpha^{'}$, $\pi_R^{'}(\cdot|m) = \pi_R^*(\cdot|m)$ for $m \neq m^{'}$. Then

$$u_S(\theta^{"}; \pi^*) = \max_{m \neq m^{'}} u_S(\theta^{"}, m, \pi_R^{'}(\cdot|m)) \leq u_S(\theta^{"}, m, \pi_R^{'}(\cdot|m^{'})) = u_S^{'}(\theta^{"}, m^{'}, \alpha^{'}),$$

and when $\theta^{'} \succ_{m'} \theta^{''}$, this implies that

$$\max_{m \neq m^{'}} u_S(\theta^{'}, m, \pi_R^{'}(\cdot|m)) < u_S(\theta^{'}, m, \pi_R^{'}(\cdot|m^{'})) = u_S^{'}(\theta^{'}, m, \alpha^{'}).$$

Hence $\alpha^{'} \in D(\theta^{'}, m^{'}; \pi^*)$.

To show (b), suppose π^* is a divine equilibrium. Then it is a Nash equilibrium, and furthermore for any off-path message $m^{'}$ where $\theta^{'} \succ_{m'} \theta^{''}$, Proposition 6 (a) implies that

$$D(\theta^{''}, m^{'}; \pi^*) \cup D^{\circ}(\theta^{''}, m^{'}; \pi^*) \subseteq D(\theta^{'}, m^{'}; \pi^*).$$

Since π^* is a divine equilibrium, $\pi_R^*(\cdot|m^{'})$ must then best respond to some belief $p \in \Delta(\Theta)$ with $\frac{p(\theta^{''})}{p(\theta')} \leq \frac{\lambda(\theta^{''})}{\lambda(\theta')}$. Considering all $(\theta^{'}, \theta^{''})$ pairs, we see that in a divine equilibrium $\pi_R^*(\cdot|m^{'})$ best responds to some belief in $P(m^{'}, \pi^*)$. $\square$

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

Example 6. (*A type-compatible equilibrium that is not divine*[14]*.)* Consider the following signalling game with two types and three messages, with prior $\lambda(\theta_1) = 2/3$.

m_1	a_1	a_2
θ_1	0, 1	-1, 0
θ_2	0, 0	-1, 1

m_2	a_1	a_2
θ_1	2, 1	-1, 0
θ_2	1, 0	-1, 1

m_3	a_1	a_2
θ_1	5, 0	-3, 1
θ_2	0, 1	-2, 0

[14] As noted by Van Damme (1987), it may seem more natural to replace the set $\alpha \in \text{MBR}(m)$ in the definitions of D and D^0 with the larger set $\alpha \in \text{co}(\text{BR}(m))$, which leads to the weaker equilibrium refinement that Sobel, Stole, and Zapater (1990) call "co-divinity". This example also shows that TCE need not be co-divine.

We claim that the following is a pure-strategy type-compatible equilibrium: $\pi_S(m_1|\theta_1) = \pi_S(m_1|\theta_2) = 1, \pi_R(a_1|m_1) = 1, \pi_R(a_2|m_2) = 1, \pi_R(a_2|m_3) = 1$. Evidently π is a Nash equilibrium. It suffices now to check that the receiver's off-path beliefs do not violate type compatibility, that is we do not have $\theta_1 \succ_{m_2} \theta_2$ or $\theta_2 \succ_{m_3} \theta_1$.

Observe that against the receiver strategy $\tilde{\pi}_R(a_1|m) = \frac{1}{2}$ for every m, m_2 is strictly optimal for θ_2 but m_3 is strictly optimal for θ_1, so $\theta_1 \not\succ_{m_2} \theta_2$. And for the receiver strategy $\hat{\pi}_R(a_1|m) = 1$ for every m, m_3 is strictly optimal for θ_1 but m_2 is strictly optimal for θ_2, so $\theta_2 \not\succ_{m_3} \theta_1$.

However, $D(\theta_2, m_2; \pi) \cup D^\circ(\theta_2, m_2; \pi)$ is the set of distributions on $\{a_1, a_2\}$ that put at least weight 0.5 on a_1. Any such distribution is in $D(\theta_1, m_2; \pi)$. So in every divine equilibrium, the receiver plays a best response to a belief that puts weight no less than 2/3 on θ_1 after message m_2, which can only be a_1. The difference here arises because under divine equilibrium, the beliefs after message m_2 only depend on the comparison between the payoffs to m_2 with those of the equilibrium message m_1, while the compatibility criterion also considers the payoffs to m_3. In the learning model, this corresponds to the possibility that θ_1 chooses to play m_3 at beliefs that induce θ_2 to play m_2. ♦

Finally, we show that every uniform type-compatible equilibrium 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),[15] which comes from iterative applications of the following pruning procedure: after message m the receiver is required to put 0 probability on those types θ such that

$$D^\circ(\theta, m; \pi^*) \subseteq \cup_{\theta' \neq \theta} D(\theta', m; \pi^*).$$

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 uniform type-compatible equilibrium into one that passes the NWBR test. Note that here, unlike in Theorem 6, we do not restrict to on-path strict equilibria. Since every such equilibrium is universally divine Cho and Kreps (1987), this implies that every uniform TCE is path-equivalent to a universally divine equilibrium.

Proposition 7. *Every uniform type-compatible equilibrium is path-equivalent to a uniform type-compatible equilibrium that passes the NWBR test.*

Proof. Consider a uniform TCE π^*. For every off-path m, perform the following modifications on $\pi_R^*(\cdot|m)$: if the first-round application of the NWBR procedure would have deleted

[15] This is closely related to, but not the same as, the NWBR property of Kohlberg and Mertens (1986).

every type, then do not modify $\pi_R^*(\cdot|m)$. Otherwise, find some θ_m not deleted by the iterated NWBR procedure, then change $\pi_R^*(\cdot|m)$ to some action in $\mathrm{BR}(\theta_m, m)$, i.e. a best response to the belief putting probability 1 on θ_m.

This modified strategy profile passes the NWBR test. We now establish that it remains a uniform TCE by checking that for those off-path m where $\pi_R^*(\cdot|m)$ was modified, the modified version is still a best response to $\hat{P}(m)$. (By uniformity, this would ensure that the modified strategy profile remains a Nash equilibrium.)

Type θ_m satisfies $\theta_m \in UD^{-1}(m)$. Otherwise, $D^\circ(\theta_m, m; \pi^*) = \emptyset$ and θ_m would be deleted by NWBR in the first round. Now it suffices to argue there is no θ' such that $\theta' \succ_m \theta_m$, which implies the belief putting probability 1 on θ_m is in $\hat{P}(m)$. But if there were such θ', by Proposition 6(a) we would have $D^\circ(\theta_m, m; \pi^*) \subseteq D(\theta', m; \pi^*)$, so θ_m should have been deleted by NWBR in the first round, contradicting the fact that θ_m survives all iterations of the NWBR procedure. $\square$

Corollary 1. *Every uniform type-compatible equilibrium is path-equivalent to a universally divine equilibrium.*

Proof. This is follows from Proposition 7 because every NWBR equilibrium is a universally divine equilibrium. $\square$

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

uniform TCEs $\subsetneq$ universally divine equilibria $\subsetneq$ strong TCEs $\subsetneq$ Intuitive Criterion $\subsetneq$ Nash equilibria.

6 Discussion and Future Work

The key modeling device that enabled us to derive most of our results is the use of agents with exponentially distributed lifetimes, in contrast to the fixed lifetimes used in (Fudenberg and Levine, 1993, 2006). Under our assumptions, the sender's optimization problem is equivalent to a discounted, infinite-horizon multi-armed bandit problem that can be solved using the Gittins index, allowing us to compare the experimentation of different types of senders. By contrast, if agents were to have finite lifetimes, their optimization problem would not be stationary. For this reason, the finite-horizon analog of the Gittins index is only approximately optimal for the finite-horizon multi-armed bandit problem (Nino-Mora, 2011). Applying the exponential lifetime framework to steady state learning models for other classes of extensive-form games could prove fruitful, especially for games where we need to compare the behaviors of various players or player types.

It is useful to contrast the necessity of the compatibility criterion here with Fudenberg and Levine (2006)'s Theorem 5.1, which roughly states that in a class of games of perfect information, there are no restrictions on the beliefs of off-path players about what would happen if they themselves deviate. This is because, as Fudenberg and Levine (2006) show, most of the time agents play a myopic best response at off-path nodes, so they do not learn the payoffs of other actions. At a formal level, the only players who move off-path in a signalling game are the receivers, and no other players act after them, so the issue raised by Theorem 5.1 is moot. That said, receivers *do* need to track and update beliefs about the senders, but no matter what a receiver plays after an off-equilibrium message m, he still learns the type who sent m at the end of the match. It is not the case that he only learns about the expected payoff to the one action he used; he fact revises his expected payoff to all of his possible actions. Moreover, the same would be true if the receivers were initially uncertain about the distribution of types, and because the updating result of Fudenberg, He, and Imhof (2016) covers this case the results here extend immediately. This contrasts with the situation in Esponda (2008), where the potential buyers do not observe the quality of cars they do not buy and thus can maintain incorrect beliefs about the quality of the cars unless they experiment.

Our results show how various sorts of TCE provide upper and lower bounds on the set of patiently stable strategy profiles in a signalling game, but they do not give an exact characterization of the set of patiently stable profiles. The gap between TCE and patient stability arises because the compatibility criterion only asks that *some* belief in $P(m', \pi^*)$ leads to a receiver best response that deters types from playing the off-path m'. There might exist types $\theta' \succ_{m'} \theta''$ such that for beliefs p with odds ratio $\frac{p(\theta'')}{p(\theta')}$ slightly below $\frac{\lambda(\theta'')}{\lambda(\theta')}$, the receiver's best response deters every type from the off-path m', but when $\frac{p(\theta'')}{p(\theta')}$ is close to 0, the receiver's best response is strictly better than some type's equilibrium payoff. Uniform TCE responds to the indeterminacy of compatible beliefs by requiring that *all* compatible beliefs lead to receiver actions that deter every type. but this requirement is too stringent. Nevertheless, our results do show how the theory of learning in games provides a foundation for equilibrium refinements in signalling games.

We hope to pursue the following extensions in future papers:

(1) **Temporary sender types**. Instead of the sender's type being assigned at birth and fixed for life, at the start of each period each sender takes an i.i.d. draw from λ to discover her type for that period. When the players are impatient, this yields different steady states than the fixed-type model here, as noted by Dekel, Fudenberg, and Levine (2004). This model will require different tools to analyze, since the sender's problem now becomes a restless bandit.

(2) **Application to supermodular signalling games**. The compatibility criterion as stated places restrictions on the two most extreme signals in a supermodular game. A more

careful analysis of the learning system should reveal restrictions on a given type's relative frequencies of experimenting across multiple messages.

(3) **Application to misspecified learning**. So far, we have considered situations where the steady state strategy profile $\bar{\psi}$ that players are learning about falls within the support of their prior g. Along the lines of Fudenberg, Romanyuk, and Strack (2017), we could consider the outcome of Bayesian learning when the prior excludes some actions that actually do get played in the steady state.

(4) **Adding passive learning**. In our model, agents only observe the outcome of their own play. In may cases agents receive some additional information, perhaps through observing the outcomes of some matches other than their own each period. We plan to consider the models featuring this kind of passive "background learning" in the future.

References

Banks, J. S. and J. Sobel (1987): "Equilibrium Selection in Signaling Games," *Econometrica*, 55, 647–661.

Berk, R. H. (1966): "Limiting Behavior of Posterior Distributions When the Model Is Incorrect," *Annals of Mathematical Statistics*, 37, 51–58.

Billingsley, P. (1995): *Probability and Measure*, John Wiley and Sons.

Cho, I.-K. and D. M. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179–221.

Dekel, E., D. Fudenberg, and D. K. Levine (1999): "Payoff information and self-confirming equilibrium," *Journal of Economic Theory*, 89, 165–185.

——— (2004): "Learning to play Bayesian games," *Games and Economic Behavior*, 46, 282–303.

Diaconis, P. and D. Freedman (1990): "On the Uniform Consistency of Bayes Estimates for Multinomial Probabilities," *Annals of Statistics*, 18, 1317–1327.

Esponda, I. (2008): "Behavioral equilibrium in economies with adverse selection," *American Economic Review*, 98, 1269–1291.

Esponda, I. and D. Pouzo (2016): "Berk-Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models," *Econometrica*, 84, 1093–1130.

Fudenberg, D., K. He, and L. Imhof (2016): "Bayesian Posteriors For Arbitrarily Rare Events," *Mimeo.*

Fudenberg, D. and Y. Kamada (2016): "Rationalizable Partition-Confirmed Equilibrium with Heterogeneous Beliefs," *Mimeo*.

Fudenberg, D. and D. M. Kreps (1988): "A Theory of Learning, Experimentation, and Equilibrium in Games," *Mimeo*.

——— (1993): "Learning mixed equilibria," *Games and Economic Behavior*, 5, 320–367.

——— (1995): "Learning in extensive-form games I. Self-confirming equilibria," *Games and Economic Behavior*, 8, 20–55.

Fudenberg, D. and D. K. Levine (1988): "Open-loop and closed-loop equilibria in dynamic games with many players," *Journal of Economic Theory*, 44, 1–18.

——— (1993): "Steady State Learning and Nash Equilibrium," *Econometrica*, 61, 547–573.

——— (2006): "Superstition and Rational Learning," *American Economic Review*, 96, 630–651.

Fudenberg, D., G. Romanyuk, and P. Strack (2017): "Active Learning with Misspecified Beliefs," *Theoretical Economics (forthcoming)*.

Fudenberg, D. and J. Tirole (1991): "Perfect Bayesian equilibrium and sequential equilibrium," *Journal of Economic Theory*, 53, 236–260.

Gittins, J. and D. Jones (1974): "A Dynamic Allocation Index for the Sequential Design of Experiments," in *Progress in Statistics*, ed. by J. Gani, Amsterdam, NL: North-Holland, 241–266.

Gittins, J. C. (1979): "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society. Series B (Methodological)*, 148–177.

Jehiel, P. and D. Samet (2005): "Learning and Bounded RationalityLearning to play games in extensive form by valuation," *Journal of Economic Theory*, 124, 129–148.

Kalai, E. and E. Lehrer (1993): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61, 1019–1045.

Kohlberg, E. and J.-F. Mertens (1986): "On the Strategic Stability of Equilibria," *Econometrica*, 54, 1003–1037.

Kreps, D. M. and R. Wilson (1982): "Sequential equilibria," *Econometrica*, 863–894.

LASLIER, J.-F. AND B. WALLISER (2014): "Stubborn learning," *Theory and Decision*, 79, 51–93.

NINO-MORA, J. (2011): "Computing a classic index for finite-horizon bandits," *INFORMS Journal on Computing*, 23, 254–267.

NOLDEKE, G. AND L. SAMUELSON (1997): "A Dynamic Model of Equilibrium Selection in Signaling Markets," *Journal of Economic Theory*, 73, 118–156.

RABIN, M. AND J. SOBEL (1996): "Deviations, dynamics, and equilibrium refinements," *Journal of Economic Theory*, 68, 1–25.

ROBBINS, H. (1952): "Some aspects of the sequential design of experiments," *Bulletin of the American Mathematical Society*, 58, 527–536.

SOBEL, J., L. STOLE, AND I. ZAPATER (1990): "Fixed-equilibrium rationalizability in signaling games," *Journal of Economic Theory*, 52, 304–331.

SPENCE, M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355–374.

VAN DAMME, E. (1987): *Stability and Perfection of Nash Equilibria*, Springer-Verlag.

Appendix

A A Sufficient Condition for Compatibility

The definition of $\theta^{'} \succ_{m'} \theta^{''}$ is phrased in terms of the weakly and strictly optimal messages for types $\theta^{'}$ and $\theta^{''}$ against some $\pi_R \in \Pi_R$, without making direct reference to the types' payoff structures. In this appendix, we present sufficient condition for compatibility that we can directly check from the signalling game payoff matrices.

Definition 13. For $h \in [0,1]$, the *maximum* and *minimum payoff wedges between types* $\theta', \theta^{''}$ *at message* $m^{'}$ *with* h *scaling are*

$$\overline{W}_h(\theta', \theta^{''}; m^{'}) := \max_{a \in \mathrm{BR}(\Delta(\Theta), m^{'})} \Big((1-h)u_S(\theta', m^{'}, a) - hu_S(\theta^{''}, m^{'}, a)\Big)$$

$$\underline{W}_h(\theta', \theta^{''}; m^{'}) := \min_{a \in \mathrm{BR}(\Delta(\Theta), m^{'})} \Big((1-h)u_S(\theta', m^{'}, a) - hu_S(\theta^{''}, m^{'}, a)\Big).$$

Proposition 8. *If there existsh* $\in [0,1]$ *with*

$$\underline{W}_h(\theta^{'}, \theta^{''}; m^{'}) > \max_{m^{''} \neq m^{'}} \overline{W}_h(\theta', \theta^{''}; m^{''}),$$

then $\theta^{'} \succ_{m'} \theta^{''}$.

Proof. **Case 1**: $h = 0$.

Then $\underline{W}_h(\theta^{'}, \theta^{''}; m^{'}) > \max_{m^{''} \neq m^{'}} \overline{W}_h(\theta', \theta^{''}; m^{''})$ is equivalent to

$$\min_{a \in \mathrm{BR}(\Delta(\Theta), m^{'})} u_S(\theta^{'}, m^{'}, a) > \max_{m^{''} \neq m^{'}} \max_{a \in \mathrm{BR}(\Delta(\Theta), m^{''})} u_S(\theta^{'}, m^{''}, a).$$

This means for any $\pi_R \in \Pi_R$, $m^{'}$ is always strictly optimal for $\theta^{'}$. This shows $\theta^{'} \succ_{m'} \theta^{''}$.

Case 2: $h = 1$.

Then $\underline{W}_h(\theta^{'}, \theta^{''}; m^{'}) > \max_{m^{''} \neq m^{'}} \overline{W}_h(\theta', \theta^{''}; m^{''})$ is equivalent to

$$\min_{a \in \mathrm{BR}(\Delta(\Theta), m^{'})} -u_S(\theta^{''}, m^{'}, a) > \max_{m^{''} \neq m^{'}} \max_{a \in \mathrm{BR}(\Delta(\Theta), m^{''})} -u_S(\theta^{''}, m^{''}, a),$$

which can be rearranged to say

$$\max_{a \in \mathrm{BR}(\Delta(\Theta), m^{'})} u_S(\theta^{''}, m^{'}, a) < \min_{m^{''} \neq m^{'}} \min_{a \in \mathrm{BR}(\Delta(\Theta), m^{''})} u_S(\theta^{''}, m^{''}, a).$$

Then we vacuously have $\theta^{'} \succ_{m'} \theta^{''}$, since $m^{'}$ is never weakly optimal for $\theta^{''}$ against any $\pi_R \in \Pi_R$.

Case 3: $0 < h < 1$.

Let any $\pi_R \in \Pi_R$ be given that makes m' weakly optimal for θ''. For any $m'' \neq m'$, we show

$$u_S(\theta', m', \pi_R(\cdot|m')) > u_S(\theta', m'', \pi_R(\cdot|m'')).$$

From $\underline{W}_h(\theta', \theta''; m') > \max_{m'' \neq m'} \overline{W}_h(\theta', \theta''; m'')$ and the fact that $\pi_R(\cdot|m)$ is supported on $\mathrm{BR}(\Delta(\Theta), m)$ for every $m \in M$, we get

$$(1-h)u_S(\theta', m', \pi_R(\cdot|m')) - hu_S(\theta'', m', \pi_R(\cdot|m')) > (1-h)u_S(\theta', m'', \pi_R(\cdot|m'')) - hu_S(\theta'', m'', \pi_R(\cdot|m'')).$$

Using the fact that $0 < h < 1$, we can rearrange this inequality to say

$$u_S(\theta', m', \pi_R(\cdot|m')) - u_S(\theta', m'', \pi_R(\cdot|m'')) > \frac{h}{1-h} \cdot \left[u_S(\theta'', m', \pi_R(\cdot|m')) - u_S(\theta'', m'', \pi_R(\cdot|m''))\right].$$

When m' is weakly optimal for θ'', $u_S(\theta'', m', \pi_R(\cdot|m')) - u_S(\theta'', m'', \pi_R(\cdot|m'')) \geq 0$. This shows $u_S(\theta', m', \pi_R(\cdot|m')) - u_S(\theta', m'', \pi_R(\cdot|m'')) > 0$, that is m' is strictly better than m'' for θ'. Since the choice of $m'' \neq m'$ was arbitrary, m' must be strictly optimal for θ'. We therefore conclude $\theta' \succ_{m'} \theta''$. □

To understand the sufficient condition in Proposition 8, suppose we take $h = \frac{1}{2}$. Then the condition is equivalent to requiring that

$$\min_{a \in \mathrm{BR}(\Delta(\Theta), m')} \left(u_S(\theta', m', a) - u_S(\theta'', m', a)\right) > \max_{m'' \neq m'} \left\{ \max_{a \in \mathrm{BR}(\Delta(\Theta), m'')} \left(u_S(\theta', m'', a) - u_S(\theta'', m'', a)\right) \right\}. \tag{5}$$

This says that the minimum payoff difference between type θ' and type θ'' at message m' is larger than the maximum payoff difference at any other message m'', where the minimum and maximum are taken over all rational receiver responses. In signalling games with separable sender payoffs $u_S(\theta, m, a) = v(\theta, m) + z(a)$, equation (5) reduces to the sufficient condition stated in the main text,

$$v(\theta', m') - v(\theta'', m') > \max_{m'' \neq m'} \left(v(\theta', m'') - v(\theta'', m'')\right).$$

Different values of h correspond to different rescalings of sender's payoffs. For each collection of $\{\alpha_\theta, \beta_\theta\}_{\theta \in \Theta}$ with $\alpha_\theta > 0$ for each θ, the rescaling

$$\tilde{u}_S(\theta, m, a) := \alpha_\theta \cdot u_S(\theta, m, a) + \beta_\theta$$

does not change any type's preference on lotteries over (m, a) pairs or experimentation incentives. Substituting the rescaled payoffs into (5), we get

$$\min_{a\in\mathrm{BR}(\Delta(\Theta),m')}\left(\alpha_{\theta'}u_S(\theta',m',a)-\alpha_{\theta''}u_S(\theta'',m',a)\right)>\max_{m''\neq m'}\max_{a\in\mathrm{BR}(\Delta(\Theta),m'')}\left(\alpha_{\theta'}u_S(\theta',m'',a)-\alpha_{\theta''}u_S(\theta'',m'',a)\right).$$

This is equivalent to requiring $\underline{W}_h(\theta',\theta'';m')>\max_{m''\neq m'}\overline{W}_h(\theta',\theta'';m'')$ for $h=\frac{\alpha_{\theta'}}{\alpha_{\theta'}+\alpha_{\theta''}}$.

B Relegated Proofs

B.1 Proof of Proposition 3

Proposition 3: $\theta'\succ_{m'}\theta''$ if and only if for every $\beta\in[0,1)$ and every ν, $I(\theta'',m',\nu,\beta)\geq\max_{m''\neq m'}I(\theta'',m'',\nu,\beta)$ implies $I(\theta',m',\nu,\beta)>\max_{m''\neq m'}I(\theta',m'',\nu,\beta)$.

Proof. **Step 1: (If)**

For every ν, define the *induced average receiver strategy* $\bar{\pi}_R^\nu\in\Pi_R$ as

$$\bar{\pi}_R^\nu(a|m):=\int_{\sigma\in\Delta(\mathrm{BR}(\Delta(\Theta),m))}\sigma(a)d\nu(\sigma),$$

where the domain of integration is the set of all rational mixed responses σ to m, distributed according to ν.

If $\theta'\not\succ_{m'}\theta''$, then there is $\pi_R\in\Pi_R$ such that

$$u_S(\theta'',m',\pi_R(\cdot|m'))\geq\max_{m''\neq m'}u_S(\theta'',m'',\pi_R(\cdot|m''))$$

and

$$u_S(\theta',m',\pi_R(\cdot|m'))\leq\max_{m''\neq m'}u_S(\theta',m'',\pi_R(\cdot|m'')).$$

But when $\beta=0$, the Gittins index of message m is just its myopic payoff, $I(\theta,m,\nu,\beta)=u_S(\theta,m,\bar{\pi}_R^\nu(\cdot|m))$, so by choosing a prior ν such that $\bar{\pi}_R^\nu=\pi_R$ we have the contradiction $I(\theta'',m',\nu,\beta)\geq\max_{m''\neq m'}I(\theta'',m'',\nu,\beta)$ yet $I(\theta',m',\nu,\beta)\leq\max_{m''\neq m'}I(\theta',m'',\nu,\beta)$.

Step 2: (Only if)

Step 2.1: Synthetic receiver strategy.

A belief ν_m and a stopping time τ_m together define a stochastic process $(A_t)_{t\geq0}$ over the space $\mathrm{BR}(\Delta(\Theta),m)\cup\{\emptyset\}$, where $A_t\in\mathrm{BR}(\Delta(\Theta),m)$ corresponds to the receiver action seen in period t if τ_m has not yet stopped ($\tau_m>t$), and $A_t:=\emptyset$ if τ_m has stopped ($\tau_m\leq t$). Enumerating $\mathrm{BR}(\Delta(\Theta),m)=\{a_1,...,a_n\}$, we write $p_{t,i}:=\mathbb{P}_{\nu_m}[A_t=a_i]$ for $1\leq i\leq n$ to record the probability of seeing receiver action a_i in period t and $p_{t,0}:=\mathbb{P}_{\nu_m}[A_t=\emptyset]=$

$\mathbb{P}_{\nu_m}[\tau_m \leq t]$ for the probability of seeing no receiver action in period t due to τ_m having stopped.

Given ν_m and τ_m, we define the *synthetic receiver strategy* $\sigma_m(\nu_m, \tau_m, \beta)$,

$$\sigma_m(\nu_m, \tau_m, \beta)(a) := \begin{cases} \frac{\sum_{t=0}^{\infty} \beta^t p_{t,i}}{\sum_{t=0}^{\infty} \beta^t (1-p_{t,0})} & \text{if } a = a_i \\ 0 & \text{else} \end{cases}.$$

As $\sum_{i=1}^{n} p_{t,i} = 1 - p_{t,0}$ for each $t \geq 0$, it is clear that $\sigma_m(\nu_m, \tau_m, \beta)$ puts non-negative weights on actions in $\mathrm{BR}(\Delta(\Theta), m)$ that sum to 1, so $\sigma_m(\nu_m, \tau_m, \beta) \in \Delta(\mathrm{BR}(\Delta(\Theta), m))$ may indeed be viewed as a rational receiver response to message m.

Step 2.2: Synthetic receiver strategy and per-period payoff.

We now show that, for any β and any stopping time τ_m for message m, the utility of playing against $\sigma_m(\nu_m, \tau_m, \beta)$ is exactly the corresponding normalized payoff under τ_m,

$$u_S(\theta, m, \sigma_m(\nu_m, \tau_m, \beta)) = \mathbb{E}_{\nu_m}\left\{\sum_{t=0}^{\tau_m - 1} \beta^t \cdot u_S(\theta, m, a_m(t))\right\} \Big/ \mathbb{E}_{\nu_m}\left\{\sum_{t=0}^{\tau_m - 1} \beta^t\right\}.$$

To see why this is true, rewrite the denominator of the right-hand side as

$$\begin{aligned}\mathbb{E}_{\nu_m}\left\{\sum_{t=0}^{\tau_m - 1} \beta^t\right\} &= \mathbb{E}_{\nu_m}\left\{\sum_{t=0}^{\infty} [1_{\tau_m > t}] \cdot \beta^t\right\} \\ &= \sum_{t=0}^{\infty} \beta^t \cdot \mathbb{P}_{\nu_m}[\tau_m > t] = \sum_{t=0}^{\infty} \beta^t (1 - p_{t,0}),\end{aligned}$$

and rewrite the numerator as

$$\begin{aligned}\mathbb{E}_{\nu_m}\left\{\sum_{t=0}^{\tau_m - 1} \beta^t \cdot u_S(\theta, m, a_m(t))\right\} &= \sum_{t=0}^{\infty} \beta^t \cdot \left(\underbrace{p_{t,0} \cdot 0}_{\text{get 0 if already stopped}} + \underbrace{\sum_{i=1}^{n} p_{t,i} \cdot u_S(\theta, m, a_i)}_{\text{else, take average expected payoff}}\right) \\ &= \sum_{i=1}^{n} \left(\sum_{t=0}^{\infty} \beta^t \cdot p_{t,i}\right) \cdot u_S(\theta, m, a_i).\end{aligned}$$

So overall,

$$\mathbb{E}_{\nu_m}\left\{\sum_{t=0}^{\tau_m-1}\beta^t\cdot u_S(\theta,m,a_m(t))\right\}\ /\ \mathbb{E}_{\nu_m}\left\{\sum_{t=0}^{\tau_m-1}\beta^t\right\}=\sum_{i=1}^{n}\left[\frac{(\sum_{t=0}^{\infty}\beta^t\cdot p_{t,i})}{\sum_{t=0}^{\infty}\beta^t(1-p_{t,0})}\right]\cdot u_S(\theta,m,a_i)$$
$$=u_S(\theta,m,\sigma_m(\nu_m,\tau_m,\beta)).$$

Thus under the optimal stopping time τ_m^θ for the stopping problem of type θ, message m,

$$u_S(\theta,m,\sigma_m(\nu_m,\tau_m^\theta,\beta))=\mathbb{E}_{\nu_m}\left\{\sum_{t=0}^{\tau_m^\theta-1}\beta^t\cdot u_S(\theta,m,a_m(t))\right\}\ /\ \mathbb{E}_{\nu_m}\left\{\sum_{t=0}^{\tau_m^\theta-1}\beta^t\right\}=I(\theta,m,\nu,\beta)$$

by the definition of $I(\theta,m,\nu,\beta)$ as the value of the optimal stopping problem.

Step 2.3: Applying the definition of $\theta^{'}\succ_{m^{'}}\theta^{''}$.

Suppose now $\theta^{'}\succ_{m^{'}}\theta^{''}$ and fix some $\beta\in[0,1)$ and prior belief ν. Suppose $I(\theta^{''},m^{'},\nu,\beta)\geq\max_{m^{''}\neq m^{'}}I(\theta^{''},m^{''},\nu,\beta)$. We show that $I(\theta^{'},m^{'},\nu,\beta)>\max_{m^{''}\neq m^{'}}I(\theta^{'},m^{''},\nu,\beta)$.

On any arm $m^{''}\neq m^{'}$ type $\theta^{''}$ could use the (suboptimal) stopping time $\tau_{m^{''}}^{\theta^{'}}$, so

$$I(\theta^{''},m^{''},\nu,\beta)\geq\mathbb{E}_{\nu_{m^{''}}}\left\{\sum_{t=0}^{\tau_{m^{''}}^{\theta^{'}}-1}\beta^t\cdot u_S(\theta^{''},m^{''},a_{m^{''}}(t))\right\}\ /\ \mathbb{E}_{\nu_{m^{''}}}\left\{\sum_{t=0}^{\tau_{m^{''}}^{\theta^{'}}-1}\beta^t\right\}$$
$$=u_S(\theta^{''},m^{''},\sigma_{m^{''}}(\nu_{m^{''}},\tau_{m^{''}}^{\theta^{'}},\beta)).$$

By the hypothesis $I(\theta^{''},m^{'},\nu,\beta)\geq\max_{m^{''}\neq m^{'}}I(\theta^{''},m^{''},\nu,\beta)$, we get $I(\theta^{''},m^{'},\nu,\beta)\geq\max_{m^{''}\neq m^{'}}u_S(\theta^{''},m^{''},\sigma_{m^{''}}(\nu_{m^{''}},\tau_{m^{''}}^{\theta^{'}},\beta))$.

Now define $\pi_R\in\Pi_R$ by $\pi_R(\cdot|m^{'}):=\sigma_{m^{'}}(\nu_{m^{'}},\tau_{m^{'}}^{\theta^{''}},\beta)$, $\pi_R(\cdot|m^{''}):=\sigma_{m^{''}}(\nu_{m^{''}},\tau_{m^{''}}^{\theta^{'}},\beta)$ for all $m^{''}\neq m^{'}$. Then $u_S(\theta^{''},m^{'},\pi_R(\cdot|m^{'}))\geq\max_{m^{''}\neq m^{'}}u_S(\theta^{''},m^{''},\pi_R(\cdot|m^{''}))$. By the definition of $\theta^{'}\succ_{m^{'}}\theta^{''}$, this implies $u_S(\theta^{'},m^{'},\pi_R(\cdot|m^{'}))>\max_{m^{''}\neq m^{'}}u_S(\theta^{'},m^{''},\pi_R(\cdot|m^{''}))$. But since $\pi_R(\cdot|m^{''})=\sigma_{m^{''}}(\nu_{m^{''}},\tau_{m^{''}}^{\theta^{'}},\beta)$, we get $u_S(\theta^{'},m^{''},\pi_R(\cdot|m^{''}))=I(\theta^{'},m^{''},\nu,\beta)$ for all $m^{''}\neq m^{'}$. This means $u_S(\theta^{'},m^{'},\sigma_{m^{'}}(\nu_{m^{'}},\tau_{m^{'}}^{\theta^{''}},\beta))>\max_{m^{''}\neq m^{'}}I(\theta^{'},m^{''},\nu,\beta)$.

On the left-hand side, $u_S(\theta^{'},m^{'},\sigma_{m^{'}}(\nu_{m^{'}},\tau_{m^{'}}^{\theta^{''}},\beta))$ is attained by taking the suboptimal stopping time $\tau_{m^{'}}^{\theta^{''}}$ in the optimal stopping problem of type $\theta^{'}$, message $m^{'}$, so we get $I(\theta^{'},m^{'},\nu,\beta)\geq u_S(\theta^{'},m^{'},\sigma_{m^{'}}(\nu_{m^{'}},\tau_{m^{'}}^{\theta^{''}},\beta))$. This shows $I(\theta^{'},m^{'},\nu,\beta)>\max_{m^{''}\neq m^{'}}I(\theta^{'},m^{''},\nu,\beta)$. □

B.2 Proof of Theorem 5

Throughout this subsection, we will make use of the following version of Hoeffding's inequality.

Fact. *(Hoeffding's inequality) Suppose $X_1, ..., X_n$ are independent random variables on $\mathbb{R}$ such that $a_i \le X_i \le b_i$ with probability 1 for each i. Write $S_n := \sum_{i=1}^n X_i$. Then,*

$$\mathbb{P}\left[|S_n - \mathbb{E}[S_n]| \ge d\right] \le 2\exp\left(-\frac{2d^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Lemma B.1. *In strategy profile π^*, suppose m^* is on-path and $\pi_R^*(a^*|m^*) = 1$, where a^* is a strict best response to m^* given π_S^*. Then there exists $N \in \mathbb{R}$ so that, for any regular prior and any sequence of steady states strategy profiles $\bar{\psi}^{(k)} \in \bar{\Psi}^*(g, \delta_k, \gamma_k)$ where $\gamma_k \to 1, \bar{\psi}^{(k)} \to \pi^*$, there exists $K \in \mathbb{N}$ such that whenever $k \ge K$, we have $\bar{\psi}_R^{(k)}(a^*|m^*) \ge 1 - (1 - \gamma_k) \cdot N$.*

Proof. Since a^* is a strict best response after m^* for π_S^*, there exists $\epsilon > 0$ so that a^* will continue to be a strict best response after m^* for any $\pi_S^{'} \in \Pi_S$ where for every $\theta \in \Theta$, $|\pi_S^{'}(m^*|\theta) - \pi_S^*(m^*|\theta)| < 3\epsilon$.

Since $\bar{\psi}^{(k)} \to \pi^*$, find large enough K such that $k \ge K$ implies for every $\theta \in \Theta$, $\left|\bar{\psi}_S^{(k)}(m^*|\theta) - \pi_S^*(m^*|\theta)\right| < \epsilon$.

Write $e_{n,\theta}^{\text{obs}}$ for the probability that an age-n receiver has encountered type θ fewer than $\frac{1}{2}n\lambda(\theta)$ times. We will find a number $N^{\text{obs}} < \infty$ so that

$$\sum_{\theta \in \Theta} \sum_{n=0}^{\infty} e_{n,\theta}^{\text{obs}} \le N^{\text{obs}}.$$

Fix some $\theta \in \Theta$. Write $Z_t^{(\theta)} \in \{0, 1\}$ as the indicator random variable for whether the receiver sees a type θ in period t of his life and write $S_n := \sum_{t=1}^n Z_t^{(\theta)}$ for the total number of type θ encountered up to age n. We have $\mathbb{E}[S_n] = n\lambda(\theta)$, so we can use Hoeffding's inequality to bound $e_{n,\theta}^{\text{obs}}$.

$$\begin{aligned} e_{n,\theta}^{\text{obs}} &\le \mathbb{P}\left[|S_n - \mathbb{E}[S_n]| \ge \frac{1}{2}n\lambda(\theta)\right] \\ &\le 2\exp\left(-\frac{2 \cdot [\frac{1}{2}n\lambda(\theta)]^2}{n}\right). \end{aligned}$$

This shows $e_{n,\theta}^{\text{obs}}$ tends to 0 at the same rate as $\exp(-n)$, so

$$\sum_{n=0}^{\infty} e_{n,\theta}^{\text{obs}} \le \sum_{n=0}^{\infty} 2\exp\left(-\frac{2 \cdot [\frac{1}{2}n\lambda(\theta)]^2}{n}\right) =: N_\theta^{\text{obs}} < \infty.$$

So we set $N^{\text{obs}} \coloneqq \sum_{\theta\in\Theta} N_\theta^{\text{obs}}$.

Next, write $e_{n,\theta}^{\text{bias},k}$ for the probability that, after observing $\left\lfloor \frac{1}{2}n\lambda(\theta)\right\rfloor$ i.i.d. draws from $\bar{\psi}_S^{(k)}(\cdot|\theta)$, the empirical frequency of message m^* differs from $\pi_S^*(m^*|\theta)$ by more than 2ϵ. So again, write $Z_t^{\theta,k} \in \{0,1\}$ to indicate if the t-th draw resulted in message m^*, with $\mathbb{E}\left[Z_t^{\theta,k}\right] = \bar{\psi}_S^{(k)}(m^*|\theta)$, and put $S_{n,k} \coloneqq \sum_{t=1}^{\left\lfloor \frac{1}{2}n\lambda(\theta)\right\rfloor} Z_t^{\theta,k}$ for total number of m^* out of $\left\lfloor \frac{1}{2}n\lambda(\theta)\right\rfloor$ draws. We have $\mathbb{E}[S_{n,k}] = \left\lfloor \frac{1}{2}n\lambda(\theta)\right\rfloor \cdot \bar{\psi}_S^{(k)}(m^*|\theta)$, but $\left|\bar{\psi}_S^{(k)}(m^*|\theta) - \pi_S^*(m^*|\theta)\right| < \epsilon$ whenever $k \geq K$. That means,

$$
\begin{aligned}
e_{n,\theta}^{\text{bias},k} \coloneqq & \mathbb{P}\left[\left|\frac{S_{n,k}}{\left\lfloor \frac{1}{2}n\lambda(\theta)\right\rfloor} - \pi_S^*(m^*|\theta)\right| \geq 2\epsilon\right] \\
\leq & \mathbb{P}\left[\left|\frac{S_{n,k}}{\left\lfloor \frac{1}{2}n\lambda(\theta)\right\rfloor} - \bar{\psi}_S^{(k)}(m^*|\theta)\right| \geq \epsilon\right] \text{ if } k \geq K \\
= & \mathbb{P}\left[\left|S_{n,k} - \mathbb{E}[S_{n,k}]\right| \geq \left\lfloor \frac{1}{2}n\lambda(\theta)\right\rfloor \cdot \epsilon\right] \\
\leq & 2\exp\left(-\frac{2\cdot(\left\lfloor \frac{1}{2}n\lambda(\theta)\right\rfloor \cdot \epsilon)^2}{\left\lfloor \frac{1}{2}n\lambda(\theta)\right\rfloor}\right) \text{ by Hoeffding's inequality.}
\end{aligned}
$$

Let $N_\theta^{\text{bias}} \coloneqq \sum_{n=1}^{\infty} 2\exp\left(-\frac{2\cdot(\left\lfloor \frac{1}{2}n\lambda(\theta)\right\rfloor \cdot\epsilon)^2}{\left\lfloor \frac{1}{2}n\lambda(\theta)\right\rfloor}\right)$, with $N_\theta^{\text{bias}} < \infty$ since the summand tends to 0 at the same rate as$\exp(-n)$. This argument shows whenever $k \geq K$, we have $\sum_{n=1}^{\infty} e_{n,\theta}^{\text{bias},k} \leq N_\theta^{\text{bias}}$. Now let $N^{\text{bias}} \coloneqq \sum_{\theta\in\Theta} N_\theta^{\text{bias}}$.

Finally, since g is regular, we appeal to Proposition 1 of Fudenberg, He, and Imhof (2016) to see that there exists some $\underline{N}$ so that whenever the receiver has a data set of size $n \geq \underline{N}$ on type θ's play, his Bayesian posterior as to the probability that θ plays m^* differs from the empirical distribution by no more than ϵ. Put $N^{\text{age}} \coloneqq \frac{2\underline{N}}{\min_{\theta\in\Theta}\lambda(\theta)}$.

Consider any steady state $\psi^{(k)}$ with $k \geq K$. With probability no smaller than $1 - \sum_{\theta\in\Theta} e_{n,\theta}^{\text{bias},k}$, an age-$n$ receiver who has seen at least $\frac{1}{2}n\lambda(\theta)$ instances of type θ for every $\theta \in \Theta$ will have an empirical distribution such that every type's probability of playing m^* differs from $\pi_S^*(m^*|\theta)$ by less than 2ϵ. If furthermore $n \geq N^{\text{age}}$, then in fact $\frac{1}{2}n\lambda(\theta) \geq \underline{N}$ for each θ so the same probability bound applies to the event that the receiver's Bayesian posterior on every type θ playing m^* deviating less than 3ϵ from $\pi_S^*(m^*|\theta)$. By the construction of ϵ, playing a^* after m^* is the unique best response to such a posterior.

Therefore, for $k \geq K$, the probability that sender population plays some action other than a^* after m^* in $\psi^{(k)}$ is bounded by

$$
N^{\text{age}}(1-\gamma_k) + (1-\gamma_k) \cdot \sum_{n=0}^{\infty} \gamma_k^n \cdot \sum_{\theta\in\Theta} \left(e_{n,\theta}^{\text{obs}} + e_{n,\theta}^{\text{bias},k}\right).
$$

To explain this expression, receivers aged N^{age} or younger account for no more than $N^{\text{age}}(1-\gamma_k)$ of the population. Among the age n receivers, no more than $\sum_{\theta\in\Theta} e^{\text{obs}}_{n,\theta}$ fraction has a sample size smaller than $\frac{1}{2}n\lambda(\theta)$ for any type θ, while $\sum_{\theta\in\Theta} e^{\text{bias},k}_{n,\theta}$ is an upper bound on the probability (conditional on having a large enough sample) of having a biased enough sample so that some type's empirical frequency of playing m^* differs by more than 2ϵ from $\pi^*_S(m^*|\theta)$.

But since $\gamma_k \in [0,1)$,

$$\sum_{n=0}^{\infty} \gamma_k^n \cdot \sum_{\theta\in\Theta} e^{\text{obs}}_{n,\theta} < \sum_{n=0}^{\infty} \sum_{\theta\in\Theta} e^{\text{obs}}_{n,\theta} \leq N^{\text{obs}}$$

and

$$\sum_{n=0}^{\infty} \gamma_k^n \cdot \sum_{\theta\in\Theta} e^{\text{bias},k}_{n,\theta} < \sum_{n=0}^{\infty} \sum_{\theta\in\Theta} e^{\text{bias},k}_{n,\theta} \leq N^{\text{bias}}.$$

We conclude that whenever $k \geq K$,

$$\bar{\psi}^{(k)}_R(a^*|m^*) \geq 1 - (1-\gamma_k)\cdot(N^{\text{age}} + N^{\text{obs}} + N^{\text{bias}}).$$

Finally, observe none of $N^{\text{age}}, N^{\text{obs}}, N^{\text{bias}}$ depends on the sequence $\bar{\psi}^{(k)}$, therefore N is chosen independent of the sequence $\bar{\psi}^{(k)}$. □

Lemma B.2. *Assume g is regular. Suppose there is some $a^* \in A$ and $v \in \mathbb{R}$ so that $u_S(\theta, m^*, a^*) > v$. Then, there exist $C_1 \in (0,1)$, $C_2 > 0$ so that in every sender history y_θ, $\#(m^*, a^*; y_\theta) \geq C_1 \cdot \#(m^*; y_\theta) + C_2$ implies $\mathbb{E}\left[u_S(\theta, m^*, \pi_R(\cdot|m^*))|y_\theta\right] > v$.*

Proof. Write $\underline{u} := \min_{a\in A} u_S(\theta, m^*, a)$. There exists $q \in (0,1)$ so that

$$q \cdot u_S(\theta, m^*, a^*) + (1-q)\cdot \underline{u} > v.$$

Find a small enough $\epsilon > 0$ so that $0 < \frac{q}{1-\epsilon} < 1$.

Since g is regular, Proposition 1 of Fudenberg, He, and Imhof (2016) tells us there exists some C_0 so that the posterior mean belief of sender with history y_θ, is no less than

$$(1-\epsilon)\cdot \frac{\#(m^*, a^*; y_\theta)}{\#(m^*; y_\theta) + C_0}.$$

Whenever this expression is at least q, the expected payoff to θ playing m^* exceeds v. That is, it suffices to have

$$(1-\epsilon)\cdot \frac{\#(m^*, a^*; y_\theta)}{\#(m^*; y_\theta) + C_0} \geq q \iff \#(m^*, a^*; y_\theta) \geq \frac{q}{1-\epsilon}\#(m^*; y_\theta) + \frac{q}{1-\epsilon}\cdot C_0.$$

Putting $C_1 := \frac{q}{1-\epsilon}$ and $C_2 := \frac{q}{1-\epsilon} \cdot C_0$ proves the lemma. $\square$

Lemma B.3. *Let Z_t be i.i.d. Bernoulli random variables, where $\mathbb{E}[Z_t] = 1 - \epsilon$. Write $S_n := \sum_{t=1}^{n} Z_t$. For $0 < C_1 < 1$ and $C_2 > 0$, there exist $\bar{\epsilon}, G_1, G_2 > 0$ such that whenever $0 < \epsilon < \bar{\epsilon}$,*

$$\mathbb{P}\left[S_n \geq C_1 n + C_2 \ \forall n \geq G_1\right] \geq 1 - G_2\epsilon.$$

Proof. We make use of a lemma from Fudenberg and Levine (2006), which in turn extends some inequalities from Billingsley (1995):**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*

$$S_n := \frac{|\sum_{k=1}^{n}(X_k - \mu)|}{n}.$$

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

$$\mathbb{P}\left[\max_{\underline{n} \leq n \leq \bar{n}} S_n > \epsilon\right] \leq \frac{2^7}{3} \cdot \frac{1}{\underline{n}} \cdot \frac{\mu}{\epsilon^4}.$$

For every $G_1 \geq 1$ and every $0 < \epsilon < 1$,

$$\begin{aligned}\mathbb{P}[S_n \geq C_1 n + C_2 \ \forall n \geq G_1] &= 1 - \mathbb{P}\left[(\exists n \geq G_1) \sum_{t=1}^{n} Z_t < C_1 n + C_2\right] \\ &= 1 - \mathbb{P}\left[(\exists n \geq G_1) \sum_{t=1}^{n}(X_t - \epsilon) > (1 - \epsilon - C_1)n - C_2\right]\end{aligned}$$

where $X_t := 1 - Z_t$. Let $\bar{\epsilon} := \frac{1}{2}(1 - C_1)$ and $G_1 := 2C_2/\bar{\epsilon}$. Suppose $0 < \epsilon < \bar{\epsilon}$. Then for every $n \geq G_1$, $(1 - \epsilon - C_1)n - C_2 \geq \bar{\epsilon}n - C_2 \geq \frac{1}{2}\bar{\epsilon}n$. Hence,

$$\mathbb{P}\left[S_n \geq C_1 n + C_2 \ \forall n \geq G_1\right] \geq 1 - \mathbb{P}\left[(\exists n \geq G_1) \sum_{t=1}^{n}(X_t - \epsilon) > \frac{1}{2}\bar{\epsilon}n\right]$$

and, by FL06 Lemma A.1, the probability on the right-hand side is at most $G_2\epsilon$ with $G_2 := 2^{11}/(3G_1\bar{\epsilon}^4)$. $\square$

We now prove Theorem 5.

Theorem 5: Suppose π^* is on-path strict for the receiver and patiently stable. Then it is a strong type-compatible equilibrium.

Proof. Let some $a' \notin \mathrm{BR}(\Delta(\tilde{J}(m', \pi^*)), m')$ and $h > 0$ be given. We will show that $\pi^*(a'|m') < 3h$.

Step 1: Defining the constants $\xi, \theta^J, a_\theta, m_\theta, C_1, C_2, G_1, G_2,$ and N^{recv}.

(i) For each $\xi > 0$, define the ξ-approximations to $\Delta(\tilde{J}(m^{'}, \pi^*))$ as the probability distributions with weight no more than ξ on types outside of $\tilde{J}(m^{'}, \pi^*)$,

$$\Delta_\xi(\tilde{J}(m^{'}, \pi^*)) := \left\{p \in \Delta(\Theta) : p(\theta) \leq \xi \ \forall \theta \notin \tilde{J}(m^{'}, \pi^*)\right\}.$$

Because the best-response correspondence has closed graph, there exists some $\xi > 0$ so that $a^{'} \notin \text{BR}(\Delta_\xi(\tilde{J}(m^{'}, \pi^*)), m^{'})$.

(ii) Since $\tilde{J}(m^{'}, \pi^*)$ is non-empty, we can fix some $\theta^J \in \tilde{J}(m^{'}, \pi^*)$.

(iii) For each equilibrium dominated type $\theta \in \Theta \backslash \tilde{J}(m^{'}, \pi^*)$, identify some on-path message m_θ so that $\pi^*_S(m_\theta|\theta) > 0$. By assumption of on-path strictness for receiver, there is some $a_\theta \in A$ so that $\pi^*_R(a_\theta|m_\theta) = 1$ and furthermore a_θ is the strict best response to m_θ in π^*. By the definition of equilibrium dominance,

$$u_S(\theta, m_\theta, a_\theta) > \max_{a \in \text{BR}(\Delta(\Theta), m^{'})} u_S(\theta, m^{'}, a) =: v_\theta.$$

By applying Lemma B.2 to each $\theta \in \Theta \backslash \tilde{J}(m^{'}, \pi^*)$, we obtain some $C_1 \in (0, 1)$, $C_2 > 0$ so for every $\theta \in \Theta \backslash \tilde{J}(m^{'}, \pi^*)$ and in every sender history y_θ, $\#(m_\theta, a_\theta; y_\theta) \geq C_1 \cdot \#(m_\theta; y_\theta) + C_2$ implies $\mathbb{E}\left[u_S(\theta, m_\theta, \pi_R(\cdot|m_\theta))|y_\theta\right] > v_\theta$.

(iv) By Lemma B.3, find $\bar{\epsilon}, G_1, G_2 > 0$ such that if $\mathbb{E}[Z_t] = 1 - \epsilon$ are i.i.d. Bernoulli and $S_n := \sum_{t=1}^n Z_t$, then whenever $0 < \epsilon < \bar{\epsilon}$,

$$\mathbb{P}\left[S_n \geq C_1 n + C_2 \ \forall n \geq G_1\right] \geq 1 - G_2 \epsilon.$$

(v) Because at π^*, a_θ is a strict best response to m_θ for every $\theta \in \Theta \backslash \tilde{J}(m^{'}, \pi^*)$, from Lemma B.1 we may find a N^{recv} so that for each sequence $\bar{\psi}^{(k)} \in \bar{\Psi}^*(g, \delta_k, \gamma_k)$ where $\gamma_k \to 1, \bar{\psi}^{(k)} \to \pi^*$, there corresponds $K^{\text{recv}} \in \mathbb{N}$ so that $k \geq K^{\text{recv}}$ implies $\bar{\psi}^{(k)}_R(a_\theta|m_\theta) \geq 1 - (1 - \gamma_k) \cdot N^{\text{recv}}$ for every $\theta \in \Theta \backslash \tilde{J}(m^{'}, \pi^*)$.

Step 2: Two conditions to ensure that all but $3h$ receivers believe in $\Delta_\xi(\tilde{J}(m^{'}, \pi^*))$.

Consider some steady state $\psi \in \Psi^*(g, \delta, \gamma)$ for g regular, $\delta, \gamma \in [0, 1)$.

In Theorem 2 of Fudenberg, He, and Imhof (2016), put $c = \frac{2}{\xi} \cdot \frac{\max_{\theta \in \Theta} \lambda(\theta)}{\lambda(\theta^J)}$ and $\delta = \frac{1}{2}$. We conclude that there exists some N^{rare} (not dependent on ψ) such that whenever $\bar{\psi}_S(m^{'}|\theta^J) \geq c \cdot \bar{\psi}_S(m^{'}|\theta^{\text{D}})$ for every equilibrium dominated type $\theta^{\text{D}} \notin \tilde{J}(m^{'}, \pi^*)$, an age-$n$ receiver in steady state ψ with

$$n \cdot \bar{\psi}_S(m^{'}|\theta^J) \geq N^{\text{rare}} \tag{6}$$

has probability at least $1 - h$ of holding a posterior belief $g_R(\cdot|y_R)$ such that θ^J is at least $\frac{1}{2}c$ times as like to play $m^{'}$ as θ^{D} is for every $\theta^{\text{D}} \notin \tilde{J}(m^{'}, \pi^*)$. Thus history y_R generates a posterior belief after m', $p(\cdot|m';y_R)$ such that

$$\frac{p(\theta^{\mathrm{D}}|m';y_R)}{p(\theta^J|m';y_R)} \leq \frac{\lambda(\theta^{\mathrm{D}})}{\lambda(\theta^J)} \cdot \xi \cdot \frac{\lambda(\theta^J)}{\max_{\theta\in\Theta}\lambda(\theta)} \leq \xi.$$

In particular, $p(\cdot|m';y_R)$ must assign weight no greater than ξ to each type not in $\tilde{J}(m',\pi^*)$, therefore the belief belongs to $\Delta_\xi(\tilde{J}(m',\pi^*))$. By construction of ξ, a' is then not a best response to m' after history y_R.

A receiver whose age n satisfies Equation (6) plays a' with probability less than h. However, to bound the overall probability of a' in the entire receiver population in steady state ψ, we ensure that Equation (6) is satisfied for all except $2h$ fraction of receivers in ψ. We claim that when γ is large enough, a sufficient condition is $\bar{\psi}_S(m'|\theta^J) \geq (1-\gamma)N^*$ for some $N^* \geq N^{\text{rare}}/h$. This is because under this condition, any agent aged $n \geq \frac{h}{1-\gamma}$ satisfies Equation (6), while the fraction of receivers younger than $\frac{h}{1-\gamma}$ is $1 - \left(\gamma^{\frac{h}{1-\gamma}}\right) \leq 2h$ for γ near enough to 1.

To summarize, in Step 2 we have found a constant N^{rare} and shown that if γ is near enough to 1, then $\bar{\psi}_R(a'|m') \leq 3h$ if ψ satisfies the following two conditions:

(C1) $\bar{\psi}_S(m'|\theta^J) \geq c \cdot \bar{\psi}_S(m'|\theta^{\mathrm{D}})$ for every equilibrium dominated type $\theta^{\mathrm{D}} \notin \tilde{J}(m',\pi^*)$

(C2) $\bar{\psi}_S(m'|\theta^J) \geq (1-\gamma)N^*$ for some $N^* \geq N^{\text{rare}}/h$.

In the following step, we show there is a sequence of steady states $\psi^{(k)} \in \Psi^*(g,\delta_k,\gamma_k)$ with $\delta_k \to 1$, $\gamma_k \to 1$, and $\bar{\psi}^{(k)} \to \pi^*$ such that in every $\bar{\psi}^{(k)}$ the above two conditions are satisfied. Using the fact that $\gamma_k \to 1$, we conclude for large enough k we get $\bar{\psi}_R^{(k)}(a'|m') \leq 3h$, which in turn shows $\pi^*(a'|m') < 3h$ as $\bar{\psi}^{(k)} \to \pi^*$.

Step 3: Extracting a suitable subsequence of steady states.

In the statement of Lemma 4, put $\pi^\circ := \pi^*$, $\theta' := \theta^J$. We obtain some number ϵ and functions $\delta(N)$, $\gamma(N,\delta)$. Put $N^{\text{ratio}} := \frac{2}{\xi}G_2 \cdot N^{\text{recv}} \frac{\max_{\theta\in\Theta}\lambda(\theta)}{\lambda(\theta^J)}$ and $N^* := \max(N^{\text{ratio}}, N^{\text{rare}}/h)$.

Since π^* is patiently stable, it can be written as the limit of some strategy profiles $\pi^* = \lim_{k\to\infty}\pi^{(k)}$, where each $\pi^{(k)}$ is δ_k-stable with $\delta_k \to 1$. By the definition of δ-stable, each $\pi^{(k)}$ is the limit $\pi^{(k)} = \lim_{j\to\infty}\bar{\psi}^{(k,j)}$ with $\psi^{(k,j)} \in \Psi^*(g,\delta_k,\gamma_{k,j})$ with $\lim_{j\to\infty}\gamma_{k,j} = 1$. It is without loss to assume that for every $k \geq 1$, $\delta_k \geq \delta(N^*)$ and that the L_1 distance between $\pi^{(k)}$ and π^* is less than $\epsilon/2$. Now for each k, find a large enough index $j(k)$ so that (i) $\gamma_{k,j(k)} \geq \gamma(N^*,\delta_k)$, (ii) L_1 distance between $\bar{\psi}^{(k,j)}$ and $\pi^{(k)}$ is less than $\min(\frac{\epsilon}{2},\frac{1}{k})$, and (iii) $\lim_{k\to\infty}\gamma_{k,j(k)} = 1$. This generates a sequence of k-indexed steady states, $\psi^{(k,j(k))} \in \Psi^*(g,\delta_k,\gamma_{k,j(k)})$. We will henceforth drop the dependence through the function $j(k)$ and just refer to $\psi^{(k)}$ and γ_k. The sequence $\psi^{(k)} \in \Psi^*(g,\delta_k,\gamma_k)$ satisfies: (1) $\delta_k \to 1, \gamma_k \to 1$; (2) $\delta_k \geq \delta(N^*)$ for each k; (3) $\gamma_k \geq \gamma(N^*,\delta_k)$ for each k; (4) $\bar{\psi}^{(k)} \to \pi^*$; (5) the L_1 distance between $\bar{\psi}^{(k)}$ and π^* is no larger than ϵ. Lemma 4 implies that, for every k,

$\bar{\psi}_S^{(k)}(m'|\theta^J) \geq (1-\gamma_k)N^*$. So, every member of the sequence thus constructed satisfies condition (**C2**).

Step 4: An upper bound on experimentation probability of equilibrium-dominated types.

It remains to show that eventually condition (**C1**) is also satisfied in the sequence constructed in **Step 3**.

We first bound the rate at which receiver's strategy $\bar{\psi}_R^{(k)}$ converges to π_R^*. By Lemma B.1, there exists some K^{recv} so that $k \geq K^{\text{recv}}$ implies

$$\bar{\psi}_R^{(k)}(a_\theta|m_\theta) \geq 1 - (1-\gamma_k) \cdot N^{\text{recv}}$$

for every $\theta \in \Theta \backslash \tilde{J}(m', \pi^*)$.

Find next a large enough K^{error} so that $k \geq K^{\text{error}}$ implies $(1-\gamma_k) \cdot N^{\text{recv}} < \bar{\epsilon}$ (where $\bar{\epsilon}$ was defined in **Step 1**).

We claim that when $k \geq \max(K^{\text{recv}}, K^{\text{error}})$, a type $\theta \notin \tilde{J}(m', \pi^*)$ sender who keeps sending message m_θ forever against a receiver population that plays $\bar{\psi}_R^{(k)}(\cdot|m_\theta)$ has less than $(1-\gamma_k) \cdot N^{\text{recv}} \cdot G_2$ chance of ever having a posterior belief that the expected payoff to m_θ is no greater than v_θ in some period $n \geq G_1$. This is because by Lemma B.3,

$$\mathbb{P}\left[S_n \geq C_1 n + C_2 \ \forall n \geq G_1\right] \geq 1 - G_2 \cdot \bar{\psi}_R^{(k)}(\{a \neq a_\theta\}|m_\theta) \geq 1 - G_2 \cdot (1-\gamma_k) \cdot N^{\text{recv}}$$

where S_n refers to the number of times that the receiver population responded to m_θ with a_θ in the first n times that m_θ was sent. But Lemma B.2 guarantees that provided $S_n \geq C_1 n + C_2$, sender's expected payoff for m_θ is strictly above v_θ, so we have established the claim.

Finally, find a large enough K^{Gittins} so that $k \geq K^{\text{Gittins}}$ implies the effective discount factor $\delta_k \gamma_k$ is so near 1 that for every $\theta \notin \tilde{J}(m', \pi^*)$, the Gittins index for message m_θ cannot fall below v_θ if m_θ has been used no more than G_1 times. Then for $k \geq \max(K^{\text{recv}}, K^{\text{error}}, K^{\text{Gittins}})$, there is less than $G_2 \cdot (1-\gamma_k) \cdot N^{\text{recv}}$ chance that the equilibrium dominated sender $\theta \notin \tilde{J}(m', \pi^*)$ will play m' even once. To see this, we observe that according to the prior, the Gittins index for m_θ is higher than that of m', whose index is no higher than its highest possible payoff v_θ. This means the sender will not play m' until her Gittins index for m_θ has fallen below v_θ. Since $k \geq K^{\text{recv}}$, this will not happen before the sender has played m_θ at least G_1 times, and since $k \geq \max(K^{\text{error}}, K^{\text{recv}})$, the previous claim establishes that the probability of the expected payoff to m_θ (and, *a fortiori*, the Gittins index for m_θ) ever falling below v_θ sometime after playing m_θ for the G_1-th time is no larger than $G_2 \cdot (1-\gamma_k) \cdot N^{\text{recv}}$.

This shows for $k \geq \max(K^{\text{recv}}, K^{\text{error}}, K^{\text{Gittins}})$, $\bar{\psi}_S^{(k)}(m'|\theta) \leq G_2 N^{\text{recv}} \cdot (1-\gamma_k)$ for every $\theta \notin \tilde{J}(m', \pi^*)$. But since $\bar{\psi}_S^{(k)}(m'|\theta^J) \geq N^* \cdot (1-\gamma_k)$ where $N^* \geq N^{\text{ratio}} = \frac{2}{\xi} G_2 \cdot N^{\text{recv}} \frac{\max_{\theta \in \Theta} \lambda(\theta)}{\lambda(\theta^J)}$, we see that condition (**C1**) is satisfied whenever $k \geq \max(K^{\text{recv}}, K^{\text{error}}, K^{\text{Gittins}})$. $\square$

B.3 Proof of Theorem 6

B.3.1 A Sufficient Condition for δ-Stability

In the first half of the proof, we will define a map $\bar{f}(\cdot; g, \delta, \gamma) : \Pi \to \Pi$, whose fixed points are steady state strategy profiles under (g, δ, γ). After establishing the continuity of $\bar{f}$, a fixed-point theorem implies $\bar{f}$ must have a fixed point on any closed, convex subset of Π that $\bar{f}$ maps into itself. So, if there is a decreasing sequence E_j of closed, convex subsets of Π and an associated sequence of survival chances $\gamma_j \to 1$, such that $\bar{f}(E_j; g, \delta, \gamma_j) \subseteq E_j$ for each j, then there is a steady state profile in E_j under (g, δ, γ_j) for each j. By taking a subsequence of these steady state profiles, we see that some strategy in $\cap_{j=1}^{\infty} E_j$ is δ-stable.

To each behavioral strategy profile $\pi = (\pi_S, \pi_R)$ of the signalling game, we may associate a state $\psi(\pi; g, \delta, \gamma)$ of the learning model, which is the distribution over histories that would be generated if a randomly sampled sender of type θ played like $\pi_S(\cdot|\theta)$ while a randomly sample receiver played like π_R. To do this, we inductively define component measures $\psi_R(\pi; g, \delta, \gamma) \in \Delta(Y_R)$, $\psi_\theta(\pi; g, \delta, \gamma) \in \Delta(Y_\theta)$, starting with

$$\psi_R(\pi; g, \delta, \gamma)(\emptyset) := 1 - \gamma$$
$$\psi_\theta(\pi; g, \delta, \gamma)(\emptyset) := 1 - \gamma.$$

Then, inductively,

$$\psi_R(\pi; g, \delta, \gamma)(y_R, s, \theta, m) := \begin{cases} \gamma \cdot \psi_R(\pi; g, \delta, \gamma)(y_R) \cdot \lambda(\theta) \cdot \pi_S(m|\theta) & \text{if } s = s_R(y_R) \\ 0 & \text{else} \end{cases}$$

and

$$\psi_\theta(\pi; g, \delta, \gamma)(y_S, m, a) := \begin{cases} \gamma \cdot \psi_\theta(\pi; g, \delta, \gamma)(y_S) \cdot \pi_R(a|m) & \text{if } m = s_\theta(y_S) \\ 0 & \text{else} \end{cases}.$$

To interpret, suppose we know $\psi_R(\pi; g, \delta, \gamma)(y_R)$ and wish to compute the probability of a history (y_R, s, θ, m), i.e. y_R together with 1 period of additional information. This probability is 0 if s is not what the receiver should have used against history y_R. Otherwise, there is γ chance that a receiver with history y_R survives into the next period. Conditional on survival, we need the receiver to meet a type θ and to observe message $\pi_S(m|\theta)$, which together have probability $\lambda(\theta) \cdot \pi_S(m|\theta)$. The interpretation of the equation for $\psi_\theta(\pi; g, \delta, \gamma)(y_S, m, a)$ is analogous.

The next Lemma gives an alternative characterization of $\bar{\Psi}^*(g, \delta, \gamma)$. Suppose we start with a strategy profile, then compute the state induced by it, and finally write down the strategy profile associated with the resulting state by the learning model. If we get back the strategy profile we started out with, then it is a steady state strategy profile. To this end, define the map $\bar{f}(\cdot; g, \delta, \gamma) : \Pi \to \Pi$ where Π is the collection of all behavioral strategy profiles of the signalling game,

$$\bar{f}(\pi; g, \delta, \gamma) \coloneqq \bar{\psi}(\pi; g, \delta, \gamma)$$

where $\bar{\psi}(\pi; g, \delta, \gamma)$ is the strategy profile associated with state $\psi(\pi; g, \delta, \gamma)$.

Lemma B.4. *If $\bar{f}(\pi; g, \delta, \gamma) = \pi$ then $\pi \in \bar{\Psi}^*(g, \delta, \gamma)$.*

Proof. See Online Appendix. □

Towards applying a fixed-point theorem, we now establish the continuity of the $\bar{f}$ function.

Lemma B.5. *$\bar{f}(\cdot; g, \delta, \gamma)$ is continuous.*

Proof. See Online Appendix. □

We now establish the central Lemma for proving sufficient conditions of patient stability.

Lemma B.6. *Suppose there is a sequence of closed convex sets $E_j \subseteq \Pi$ and a sequence of survival probabilities (γ_j) such that (i) $E_j \downarrow E_\infty$; (ii) $\gamma_j \uparrow 1$, (iii) for every j, $\bar{f}(E_j; g, \delta, \gamma_j) \subseteq E_j$. Then there is some $\pi^\infty \in E_\infty$ which is δ-stable under regular prior g.*

Proof. Since E_j is compact and convex and $\bar{f}$ is continuous by Lemma B.5, there is a fixed point $\pi^{(j)} \in E_j$ for (g, δ, γ_j). By Lemma B.4, this fixed point is a steady state strategy profile, $\pi^{(j)} \in \bar{\Psi}^*(g, \delta, \gamma_j)$. Thus we have a sequence of steady state strategy profiles $\pi^{(j)}$ associated with survival probabilities γ_j that converge to 1. But some subsequence of $\pi^{(j)}$ converges, and furthermore the subsequence must converge to some point in E_∞ as $E_j \downarrow E_\infty$. □

B.3.2 Proof of Theorem 6

In the next half of the proof, we will construct sets of strategy profiles E_j whose intersection only includes strategy profiles that agree with the desired π^* on the equilibrium path. We will define a prior g so that for any $\delta \in (0, 1)$, there exists a sequence $\gamma_j \to 1$ so that $\bar{f}(E_j; g, \delta, \gamma_j) \subseteq E_j$ for each j. By Lemma B.6, this shows some strategy profile path-equivalent to π^* is δ-stable for every δ, hence some such strategy profile must be patiently stable.

Consider a quasi-strict uniform type-compatible equilibrium, π^* that is on-path strict for the receiver and strict for the sender. It is without loss to assume that π^* is also a PBE (if not, we may modify $\pi^*(\cdot|m)$ at off-path m so that it is a pure best response to $\hat{P}(m)$ – this modification will continue to deter all types from deviating to m).

For $\epsilon_1 \geq 0$, define the ϵ_1 closed ball around π^* on-path, $B_{\text{on}}(\pi^*, \epsilon_1)$, as

$$B_{\text{on}}(\pi^*, \epsilon_1) := \left\{ \pi \in \Pi : \begin{array}{c} |\pi_S(m|\theta) - \pi_S^*(m|\theta)| \leq \epsilon_1 \ \forall \theta, m \\ |\pi_R(a|m) - \pi_R^*(a|m)| \leq \epsilon_1, \forall a, \text{ on-path } m \text{ in } \pi^* \end{array} \right\}.$$

Then define strategies that differ no more than $\epsilon_2 \geq 0$ from best responses to $\hat{P}(m)$ after each off-path message m,

$$B_{\text{off}}(\pi^*, \epsilon_2) := \left\{ \pi \in \Pi : \pi_R \left(\text{BR}(\hat{P}(m), m) \mid m \right) \geq 1 - \epsilon_2 \text{ for each off-path } m \right\}.$$

It is clear that both $B_{\text{on}}(\pi^*, \epsilon_1)$ and $B_{\text{off}}(\pi^*, \epsilon_2)$ are closed and convex.

Finally, define the set of "compatible" strategy profiles as

$$C := \{\pi \in \Pi : \pi_S(m|\theta) \geq \pi_S(m|\theta') \text{ whenever } \theta \succ_m \theta'\}.$$

As C consists of finitely many weak linear restrictions on Π, it is also closed and convex. Note also that since π^* is a PBE, we have $\pi^* \in C$.

Theorem 6: If π^* is a quasi-strict uniform type-compatible equilibrium that is on-path strict for the receiver and strict for the sender, then it is path-equivalent to a patiently stable strategy profile.

Proof. We proceed in three steps. In Step 1, we show that for any fixed $\epsilon_2 > 0$, we can construct an independent, non-doctrinaire, Dirichlet prior g_R for the receiver, together with a threshold $\underline{\gamma}_{R,1} \in (0, 1)$, so that for any $\gamma > \underline{\gamma}_{R,1}$, $\delta \in (0, 1)$ and $\pi \in C$, we get

$$\bar{f}_R(\pi; g_R, \delta, \gamma)(\text{BR}(\hat{P}(m), m) \mid m) \geq 1 - \epsilon_2, \forall \text{ off-path } m \text{ in } \pi^*.$$

In Step 2, we find $\epsilon_R \in (0, 1)$ so that for any $\epsilon_1 > 0$, there exists a threshold $\underline{\gamma}_{R,2}(\epsilon_1) \in (0, 1)$ with the property that whenever $\gamma > \underline{\gamma}_{R,2}(\epsilon_1)$, $\delta \in (0, 1)$ and $\pi \in B_{\text{on}}(\pi^*, \epsilon_R)$, we have

$$|\bar{f}_R(\pi; g_R, \delta, \gamma)(a|m) - \pi_R^*(a|m)| \leq \epsilon_1, \forall a, \text{ on-path } m \text{ in } \pi^*$$

(where g_R is as constructed in step 1). Together, these two steps imply that we can fix the receiver's prior g_R such that whenever players have sufficiently long expected lifetimes and the senders play in a way that does not differ too much from π^* and respects compatibility, the receiver population plays a best response to $\hat{P}(m)$ after off-path m with probability at least $1-\epsilon_2$, and plays arbitrarily similarly to π^* after on-path messages.

In Step 3, we construct the sender's prior g_S and pick $\epsilon_S > 0$ (not dependent on the g_R and ϵ_R constructed in steps 1 and 2) such that for any $\delta \in (0,1)$ and $0 < \epsilon_1 < \epsilon_S$, there exists a $\underline{\gamma}_S(\delta, \epsilon_1)$ so that whenever $\gamma > \underline{\gamma}_S(\delta, \epsilon_1)$ and $\pi \in B_{\text{on}}(\pi^*, \epsilon_1) \cap B_{\text{off}}(\pi^*, \epsilon_S)$, we have

$$|\bar{f}_S(\pi; g_S, \delta, \gamma)(m|\theta) - \pi_S^*(m|\theta)| \leq \epsilon_1, \forall m, \theta.$$

To see how these three steps can be used to establish the Proposition, Step 3 lets us find a sender prior g_S, constant ϵ_S, and threshold function $\underline{\gamma}_S(\delta, \epsilon_1)$. Next, in Step 1, set $\epsilon_2 = \epsilon_S$ to find receiver prior g_R and threshold $\underline{\gamma}_{R,1}$. Finally, in step 2, find ϵ_R and threshold function $\underline{\gamma}_{R,2}(\epsilon_1)$.

Letting δ be arbitrary, we show that some strategy profile path-equivalent to π^* is δ-stable under any regular prior. Consider the sequence of decreasing, closed, convex sets of strategy profiles given by

$$E_j := C \cap B_{\text{on}}(\pi^*, \min(\epsilon_R, \epsilon_S)/j) \cap B_{\text{off}}(\pi^*, \epsilon_R).$$

That is, E_j is the set of strategy profiles that respect compatibility, differ by no more than ϵ_R/j from π^* on path, and differ by no more than ϵ_R from π^* off path. We may find an accompanying sequence of survival probabilities satisfying

$$\gamma_j > \max\left(\underline{\gamma}_{R,1}, \underline{\gamma}_{R,2}(\min(\epsilon_R, \epsilon_S)/j), \underline{\gamma}_S(\delta, \min(\epsilon_R, \epsilon_S)/j)\right)$$

with $\gamma_j \uparrow 1$. Since we chose $\epsilon_2 = \epsilon_S$, Step 1 implies that for each $\pi \in E_j$,

$$\bar{f}_R(\pi; g_R, \delta, \gamma_j)(\text{BR}(\hat{P}(m), m) \mid m) \geq 1 - \epsilon_R, \forall \text{ off-path } m \text{ in } \pi^*,$$

so $\bar{f}(E_j; g, \delta, \gamma_j) \subseteq B_{\text{off}}(\pi^*, \epsilon_R)$.

Choose $\epsilon_1 = \min(\epsilon_R, \epsilon_S)/j$ in Step 2. Since $E_j \subseteq B_{\text{on}}(\pi^*, \epsilon_R/j) \subseteq B_{\text{on}}(\pi^*, \epsilon_R)$, Step 2 implies that for every $\pi \in E_j$,

$$|\bar{f}_R(\pi; g_R, \delta, \gamma_j)(a|m) - \pi_R^*(a|m)| \leq \epsilon_1, \forall a, \text{ on-path } m \text{ in } \pi.^*$$

Finally, again choose $\epsilon_1 = \min(\epsilon_R, \epsilon_S)/j$ in Step 3. since $E_j \subseteq B_{\text{on}}(\pi^*, \epsilon_S/j) \subseteq B_{\text{on}}(\pi^*, \epsilon_S)$, Step 3 implies that for every $\pi \in E_j$,

$$|\bar{f}_S(\pi; g_S, \delta, \gamma_j)(m|\theta) - \pi_S^*(m|\theta)| \leq \epsilon_1, \forall m, \theta.$$

This shows that $\bar{f}(E_j; g, \delta, \gamma_j) \subseteq B_{\text{on}}(\pi^*, \epsilon_1) = B_{\text{on}}(\pi^*, \min(\epsilon_R, \epsilon_S)/j)$.

By an application of Lemma 1, we know that $\bar{f}(\pi; g, \delta, \gamma_j) \in C$ for any $\pi \in \Pi$. In conclusion, we have shown

$$\bar{f}(E_j; g, \delta, \gamma_j) \subseteq C \cap B_{\text{on}}(\pi^*, \min(\epsilon_R, \epsilon_S)/j) \cap B_{\text{off}}(\pi^*, \epsilon_R) = E_j.$$

Since $E_j \downarrow B_{\text{on}}(\pi^*, 0) \cap B_{\text{off}}(\pi^*, \epsilon_R)$, which are the strategy profiles that match π^* on path and put weight no more than ϵ_R outside of $\text{BR}(\hat{P}(m), m)$, by Lemma B.6 there exists some $\pi(\delta)$ path-equivalent to π^* such that $\pi(\delta)$ is δ-stable. Since this argument applies to an arbitrary δ, there exists a sequence $\delta_j \uparrow 1$ such that $\pi(\delta_j)$ converge. Since each of $\pi(\delta_j)$ matches π^* on path, there exists some patiently stable strategy profile which is path-equivalent to π^*.

It remains to show that the constructions in Steps 1, 2, 3 are feasible. Since Step 3 involves tedious details and does not depend on the thresholds and priors constructed in Steps 1 and 2, we present it in the Online Appendix.

Step 1: **Constructing g_R and $\underline{\gamma}_{R,1}$.**

For each $\xi > 0$, consider the approximation to $\hat{P}(m)$,

$$\hat{P}_\xi(m) := \left\{ p \in UD^{-1}(m) : \frac{p(\theta')}{p(\theta)} \le (1+\xi) \cdot \frac{\lambda(\theta')}{\lambda(\theta)} \text{ whenever } \theta \succ_m \theta \right\}.$$

There exists some $\xi > 0$ such that $\text{BR}(\hat{P}_\xi(m), m) = \text{BR}(\hat{P}(m), m)$. Else, if exists $\tilde{a} \notin \text{BR}(\hat{P}(m), m))$ such that for every $\xi > 0$, there is $0 < \xi' < \xi$ with $\tilde{a} \in \text{BR}(\hat{P}_{\xi'}(m), m)$, then $\tilde{a} \in \text{BR}(\hat{P}(m), m)$ also. Take some such ξ and next we will choose a series of constants.

- Pick $0 < h < 1$ such that $\frac{1-h}{1+h} > (1-\xi)^{1/3}$.
- Pick $\underline{N} \in \mathbb{N}$ so that for any $N > \underline{N}$, $\theta \in \Theta$, we have

$$\mathbb{P}[(1-h) \cdot N \cdot \lambda(\theta) \le \text{Binom}(N, \lambda(\theta)) \le (1+h) \cdot N \cdot \lambda(\theta)] > 1 - \frac{\epsilon_2}{4 \cdot |\Theta|}.$$

- Pick $G > 0$ such that for every $\theta \in \Theta$, $1/(h \cdot \sqrt{G \cdot (1-h) \cdot \lambda(\theta)})^2 < \epsilon_2/(4 \cdot |M| \cdot |\Theta|^2)$.
- Pick numbers $\alpha(\theta, m) > 0$ so that whenever $\theta \succ_m \theta'$, we have

$$\alpha(\theta, m) - \alpha(\theta', m) > (\sqrt{(4 \cdot |M| \cdot |\Theta|^2)/\epsilon_2} + 1) \cdot G. \tag{7}$$

  Ensure also $\sum_m \alpha(\theta, m) = A$ is the same for all θ.
- Pick $\underline{\gamma}_{R,1} \in (0,1)$ such that $1 - (\underline{\gamma}_{R,1})^{\underline{N}+1} < \epsilon_2/4$.

Suppose the receiver's prior as to the strategy of typeθ is Dirichlet with parameters $(\alpha(\theta, m))_{m \in M}$. We claim that whenever $\gamma \in (\underline{\gamma}_{R,1}, 1)$ and $\pi \in C$,

$$\bar{f}_R(\pi; g_R, \delta, \gamma)\left(\text{BR}(\hat{P}(m), m) \mid m\right) \geq 1 - \epsilon_2, \forall a, \text{ off-path } m \text{ in } \pi^*.$$

To prove this claim, we will show that at receiver histories of length $N > \underline{N}$, it is extremely likely that the receiver holds a belief in $\hat{P}_\xi(m)$, hence will play some action in $\text{BR}(\hat{P}(m), m)$.This will be sufficient to prove step 1 because we have chosen γ large enough so that histories with length less than $\underline{N}$ are rare. To prove that it is extremely likely receiver's belief lies in $\hat{P}_\xi(m)$ for large N, we first ensure the number of times the receiver has seen senders of type θ is roughly proportional to the prior. Next, we branch into two cases. (1) If $\pi_S(m|\theta)$ is small relative to N for even the compatible type θ, then the much bigger prior choice $\alpha(\theta, m) \gg \alpha(\theta', m)$ ensures the number of times the less compatible type θ' has played m is unlikely to overwhelm the prior, preserving receiver's belief to be in $\hat{P}_\xi(m)$. (2) If $\pi_S(m|\theta)$ is large relative to N, then law of large numbers kicks in to ensure that the ratio of number of times that types θ and θ' played m does not fall too far below the prior ratio of the two types, given a behavioral strategy where θ plays m more frequently than θ' does.

To spell out the details, fix some strategy profile $(\pi_S, \pi_R) \in C$. Write $\#(\theta|y_R)$ for the number of times the sender has been of θ type in history y_R, while $\#(\theta, m|y_R)$ counts the number of times type θ has sent message m in history y_R. Put $\psi_R = \psi_R(\pi; g_R, \delta, \gamma)$ and write $E \subseteq Y_R$ for those receiver histories with length at least $\underline{N}$ satisfying

$$(1 - h) \cdot N \cdot \lambda(\theta) \leq \#(\theta|y_R) \leq (1 + h) \cdot N \cdot \lambda(\theta)$$

for every $\theta \in \Theta$. By choice of $\underline{N}$ and $\underline{\gamma}_{R,1}$, whenever $\gamma > \underline{\gamma}_{R,1}$ 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 message m lies in $\hat{P}_\xi(m)$ is at least $1 - \epsilon/2$. To do this, fix message m and two types with $\theta \succ_m \theta'$. After history y_R, the receiver's updated posterior likelihood ratio for types θ and θ' upon seeing message m is

$$\frac{\lambda(\theta)}{\lambda(\theta')} \cdot \left(\frac{\alpha(\theta, m) + \#(\theta, m|y_R)}{\#(\theta|y_R) + A} \Big/ \frac{\alpha(\theta', m) + \#(\theta', m|y_R)}{\#(\theta'|y_R) + A} \right) = \frac{\lambda(\theta)}{\lambda(\theta')} \cdot \frac{\alpha(\theta, m) + \#(\theta, m|y_R)}{\alpha(\theta', m) + \#(\theta', m|y_R)} \cdot \frac{\#(\theta'|y_R) + A}{\#(\theta|y_R) + A}.$$

Since we have $\#(\theta'|y_R) \geq (1 - h) \cdot N \cdot \lambda(\theta')$ while $\#(\theta|y_R) \leq (1 + h) \cdot N \cdot \lambda(\theta)$, we get

$$\frac{\#(\theta'|y_R) + A}{\#(\theta|y_R) + A} \geq \frac{1 - h}{1 + h} \frac{\lambda(\theta')}{\lambda(\theta)} > (1 - \xi)^{1/3} \cdot \frac{\lambda(\theta')}{\lambda(\theta)}.$$

Now we analyze the term $\frac{\alpha(\theta, m) + \#(\theta, m|y_R)}{\alpha(\theta', m) + \#(\theta', m|y_R)}$ by considering two cases, depending on whether

N is "large enough" so that the compatible type θ experiments enough on average in a receiver history of length N under sender strategy π_S.

Case A: $\pi_S(m|\theta) \cdot N < G$. In this case, since $\pi \in C$ and $\theta \succ_m \theta'$, we must also have $\pi_S(m|\theta') \cdot N < G$. Then $\#(\theta', m|y_R)$ 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 $(\sqrt{(4 \cdot |M| \cdot |\Theta|^2)/\epsilon_2} + 1) \cdot G$ is no larger than

$$\frac{1}{G \cdot (4 \cdot |M| \cdot |\Theta|^2)/\epsilon_2} = \frac{\epsilon_2}{4|\Theta|^2 \cdot |M| \cdot G} < \frac{\epsilon_2}{4|M| \cdot |\Theta|^2}.$$

But in any history y_R where $\#(\theta', m|y_R)$ does not exceed this number, we would have

$$\alpha(\theta^{'}, m) + \#(\theta^{'}, m|y_R) \leq \alpha(\theta, m) \leq \alpha(\theta, m) + \#(\theta, m|y_R)$$

by choice of the difference between prior parameters $\alpha(\theta', m)$ and $\alpha(\theta, m)$. Therefore $\frac{\alpha(\theta,m)+\#(\theta,m|y_R)}{\alpha(\theta^{'},m)+\#(\theta^{'},m|y_R)} \geq 1$. In summary, under Case A, there is probability no smaller than $1 - \frac{\epsilon_2}{4|\Theta|^2}$ that $\frac{\alpha(\theta,m)+\#(\theta,m|y_R)}{\alpha(\theta^{'},m)+\#(\theta^{'},m|y_R)} \geq 1$.

Case B: $\pi_S(m|\theta) \cdot N \geq G$. In this case, we can bound the probability that

$$\#(\theta, m|y_R)/\#(\theta^{'}, m|y_R) \leq \frac{\lambda(\theta)}{\lambda(\theta')} \cdot (\frac{1-h}{1+h})^2.$$

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

On the other hand, given that $\#(\theta|y_R) \leq (1+h) \cdot N \cdot \lambda(\theta')$ and furthermore $\pi_S(m|\theta') \leq \pi_S(m|\theta) = p$, the distribution of $\#(\theta', m|y_R)$ 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

$$\begin{aligned}
&\mathbb{P}\left[\text{Binom}((1-h) \cdot N \cdot \lambda(\theta), p) < (1-h) \cdot (1-h) \cdot N \cdot \lambda(\theta) \cdot p\right] \\
&< 1/(h \cdot \sqrt{p(1-h)N\lambda(\theta)})^2 \leq 1/(h \cdot \sqrt{G \cdot (1-h) \cdot \lambda(\theta)})^2 < \epsilon_2/(4 \cdot |M| \cdot |\Theta|^2)
\end{aligned}$$

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

$$\mathbb{P}\left[\text{Binom}((1+h)\cdot N\cdot\lambda(\theta'),p)>(1+h)\cdot(1+h)\cdot N\cdot\lambda(\theta')\cdot p\right]$$
$$<1/(h\cdot\sqrt{p(1+h)N\lambda(\theta')})^2\leq 1/(h\cdot\sqrt{G\cdot(1+h)\cdot\lambda(\theta')})^2<\epsilon_2/(4\cdot|M|\cdot|\Theta|^2)$$

by the same arguments. Via stochastic dominance, this shows

$$\mathbb{P}\left[\#(\theta,m|y_R)/\#(\theta',m|y_R)\leq\frac{\lambda(\theta)}{\lambda(\theta')}\cdot(\frac{1-h}{1+h})^2\right]<\epsilon_2/(2\cdot|M|\cdot|\Theta|^2).$$

So, *a fortiori*,

$$\mathbb{P}\left[\frac{\text{Binom}((1-h)\cdot N\cdot\lambda(\theta),p)}{\text{Binom}((1+h)\cdot N\cdot\lambda(\theta'),p)}\leq\frac{\lambda(\theta)}{\lambda(\theta')}\cdot(\frac{1-h}{1+h})^2\right]<\epsilon_2/(2\cdot|M|\cdot|\Theta|^2).$$

Therefore, for any m,θ,θ' such that $\theta\succ_m\theta'$,

$$\psi\left(y_R:\frac{\alpha(\theta,m)+\#(\theta,m|y_R)}{\alpha(\theta',m)+\#(\theta',m|y_R)}\geq\frac{\lambda(\theta)}{\lambda(\theta')}\cdot(\frac{1-h}{1+h})^2\mid E\right)\geq 1-\epsilon_2/(2\cdot|M|\cdot|\Theta|^2).$$

In either event, at a history y_R with $(1-h)\cdot N\cdot\lambda(\theta)\leq\#(\theta|y_R)\leq(1+h)\cdot N\cdot\lambda(\theta)$ for every θ, for every pair θ,θ' such that $\theta\succ_m\theta'$, we get $\frac{\alpha(\theta,m)+\#(\theta,m|y_R)}{\alpha(\theta',m)+\#(\theta',m|y_R)}\geq\frac{\lambda(\theta)}{\lambda(\theta')}\cdot(\frac{1-h}{1+h})^2$ with probability at least $1-\epsilon_2/(2\cdot|M|\cdot|\Theta|^2)$.

But at any history y_R where this happens, the receiver's posterior likelihood ratio for types θ and θ' after message m satisfies

$$\begin{aligned}\frac{\lambda(\theta)}{\lambda(\theta')}\cdot\frac{\alpha(\theta,m)+\#(\theta,m|y_R)}{\alpha(\theta',m)+\#(\theta',m|y_R)}\cdot\frac{\#(\theta'|y_R)+A}{\#(\theta|y_R)+A} &\geq \frac{\lambda(\theta)}{\lambda(\theta')}\cdot\frac{\lambda(\theta)}{\lambda(\theta')}\cdot(\frac{1-h}{1+h})^2\cdot(1-\xi)^{1/3}\cdot\frac{\lambda(\theta')}{\lambda(\theta)}\\ &\geq \frac{\lambda(\theta)}{\lambda(\theta')}\cdot(1-\xi)^{2/3}\cdot(1-\xi)^{1/3}\geq\frac{\lambda(\theta)}{\lambda(\theta')}\cdot(1-\xi).\end{aligned}$$

As there are at most $|\Theta|^2$ such pairs for each message m and $|M|$ total messages,

$$\psi\left(y_R:\frac{\lambda(\theta)}{\lambda(\theta')}\cdot\frac{\alpha(\theta,m)+\#(\theta,m|y_R)}{\alpha(\theta',m)+\#(\theta',m|y_R)}\cdot\frac{\#(\theta'|y_R)+A}{\#(\theta|y_R)+A}\geq\frac{\lambda(\theta)}{\lambda(\theta')}\cdot(1-\xi)\ \forall m,\theta\succ_m\theta'\ |E\right)\geq 1-\epsilon_2/2$$

as claimed. As the event E has ψ-probability no smaller than $1-\epsilon_2/2$, there is ψ probability at least $1-\epsilon$ that receiver's posterior belief is in $\hat{P}_\xi(m)$ after every off-equilibrium m.

Step 2: **Constructing $\epsilon_R \in (0,1)$ and $\underline{\gamma}_{R,2}(\epsilon_1)$.**

Keep the prior g_R from Step 1. Since π^* is on-path strict for the receiver, there exists some $\xi > 0$ such that for every on-path message m and every belief $p \in \Delta(\Theta)$ with

$$|p(\theta) - p(\theta; m, \pi^*)| < \xi, \ \forall \theta \in \Theta \tag{8}$$

(where $p(\cdot; m, \pi^*)$ is the Bayesian belief after on-path message m induced by the equilibrium π^*), we have $\text{BR}(p, m) = \{\pi_R^*(m)\}$. For each m, we show that there is a large enough $N(m, \epsilon_1)$ and small enough $\zeta(m)$ so that when receiver observes history y_R generated by any $\pi \in B_{\text{on}}(\pi^*, \epsilon_1)$ with $\epsilon_1 < \zeta(m)/4$ and length least $N(m, \epsilon_1)$, there is probability at least $1 - \frac{\epsilon_1}{2|M|}$ that receiver's posterior belief satisfies (8). Hence, conditional on having a history length of at least $N(m, \epsilon_1)$, there is $1 - \frac{\epsilon_1}{2|M|}$ chance that receiver will play as in π_R^* after m. By taking the maximum $N^*(\epsilon_1) := \max_m(N(m, \epsilon_1))$ and minimum $\epsilon_R := \min_m \zeta(m)$, we see that whenever history is length $N^*(\epsilon_1)$ or more, and $\pi \in B_{\text{on}}(\pi^*, \epsilon_1)$ with $\epsilon_1 < \epsilon_R$, there is at least $1 - \epsilon_1/2$ chance that the receiver's strategy matches π_R^* after every on-path message. Since we can pick $\underline{\gamma}_{R,2}(\epsilon_1)$ large enough that $1 - \epsilon_1/2$ measure of the receiver population is age $N^*(\epsilon_1)$ or older, we are done.

To construct $N(m, \epsilon_1)$ and $\zeta(m)$, let $\Lambda(m) := \lambda\{\theta : \pi_S^*(m|\theta) = 1\}$. Find small enough $\zeta(m) \in (0,1)$ so that:

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

for every $\theta \in \Theta$. After a history y_R, the receiver's posterior belief as to the type of sender who sends message m satisfies

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

where $\alpha(\theta, m)$ is the Dirichlet prior parameter on message m for type θ and $A(\theta) := \sum_{m \in M} \alpha(\theta, m)$, as defined in Step 1. By the law of large numbers, for long enough history length, we can ensure that if $\pi_S(m|\theta) > 1 - \frac{\zeta(m)}{4}$, then

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

with probability at least $1 - \frac{\epsilon_1}{2|M|^2}$, while if $\pi_S(m|\theta) < \zeta(m)/4$, then

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

with probability at least $1 - \frac{\epsilon_1}{2|M|^2}$. Moreover there is some $N(m, \epsilon_1)$ so that there is probability at least $1 - \frac{\epsilon_1}{2|M|}$ that a history y_R with length at least $N(m, \epsilon_1)$ satisfies above for all θ. But at such a history, for any θ such that $\pi_S^*(m|\theta) = 1$,

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

and

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

while for some θ such that $\pi_S^*(m|\theta) = 0$,

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

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

Step 3: Constructing g_S, $\gamma_S(\epsilon_1)$, and $\epsilon_S > 0$.

See Online Appendix. □