ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 86, No. 4 (July, 2018), 1215–1255

LEARNING AND TYPE COMPATIBILITY IN SIGNALING GAMES

DREW FUDENBERG
Department of Economics, MIT

KEVIN HE
Department of Economics, Harvard University

Econometrica, Vol. 86, No. 4 (July, 2018), 1215–1255

LEARNING AND TYPE COMPATIBILITY IN SIGNALING GAMES

DREW FUDENBERG
Department of Economics, MIT

KEVIN HE
Department of Economics, Harvard University

Which equilibria will arise in signaling games depends on how the receiver interprets deviations from the path of play. We develop a micro-foundation for these off-path beliefs, and an associated equilibrium refinement, in a model where equilibrium arises through non-equilibrium learning by populations of patient and long-lived senders and receivers. In our model, young senders are uncertain about the prevailing distribution of play, so they rationally send out-of-equilibrium signals as experiments to learn about the behavior of the population of receivers. Differences in the payoff functions of the types of senders generate different incentives for these experiments. Using the Gittins index (Gittins (1979)), we characterize which sender types use each signal more often, leading to a constraint on the receiver's off-path beliefs based on "type compatibility" and hence a learning-based equilibrium selection.

KEYWORDS: Bandit problems, equilibrium refinements, learning in games, signaling games.

1. INTRODUCTION

IN A SIGNALING GAME, a privately informed *sender* (for instance, a student) observes their type (e.g., ability) and chooses a signal (e.g., education level) that is observed by a *receiver* (such as an employer), who then picks an action without observing the sender's type. These signaling games can have many perfect Bayesian equilibria, which are supported by different specifications of how the receiver would update his beliefs about the sender's type following the observation of off-path signals that the equilibrium predicts will never occur. These off-path beliefs are not pinned down by Bayes's rule, and solution concepts such as perfect Bayesian equilibrium and sequential equilibrium place no restrictions on them. This has led to the development of equilibrium refinements like Cho and Kreps's (1987) Intuitive Criterion and Banks and Sobel's (1987) divine equilibrium that reduce the set of equilibria by imposing restrictions on off-path beliefs, using arguments about how players should infer the equilibrium meaning of observations that the equilibrium says should never occur.

This paper uses a learning model to provide a micro-foundation for restrictions on the off-path beliefs in signaling games, and thus derive restrictions on which Nash equilibria can emerge from learning. Our learning model has a continuum of agents who are randomly matched each period, with a constant inflow of new agents who do not know the

Drew Fudenberg: drew.fudenberg@gmail.com
Kevin He: hesichao@gmail.com

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

 https://doi.org/10.3982/ECTA15085

prevailing distribution of strategies and a constant outflow of equal size. The large population makes it rational for the agents to ignore repeated-game effects and ensures the aggregate system is deterministic, while turnover in the population lets us analyze learning in a stationary model where social steady states exist, even though individual agents learn.[1] To give agents adequate learning opportunities, we assume that their expected lifetimes are long, so that most agents in the population live a long time. And to ensure that agents have sufficiently strong incentives to experiment, we suppose that they are very patient. This leads us to analyze what we call the "*patiently stable*" steady states of our learning model.

Our agents are Bayesians who believe they face a time-invariant distribution of opponents' play. As in much of the learning-in-games literature and most laboratory experiments, these agents only learn from their personal observations and not from sources such as newspapers, parents, or friends.[2] Therefore, patient young senders will rationally try out different signals to see how receivers react. This implies some "off-path" signals that have probability zero in a given equilibrium will occur with small but positive probabilities in the steady states that approximate it, so we can use Bayes's rule to derive restrictions on the receivers' typical posterior beliefs following these rare but positive-probability observations. Moreover, differences in the payoff functions of the sender types lead them to experiment in different ways. As a consequence, we can prove that patiently stable steady states must be a subset of Nash equilibria where the receiver responds to beliefs about the sender's type that respect a *type compatibility* condition. This provides a learning-based justification for eliminating certain "unintuitive" equilibria in signaling games. These results also suggest that learning theory could be used to control the rates of off-path play and hence generate equilibrium refinements in other games.

1.1. *A Toy Example*

To give some of the intuition for our general results, we study a particular stage game embedded in an artificially simple learning model, and explain why optimal experimentation rules out a seemingly unappealing equilibrium outcome. Consider the following signaling game: the sender is either the high type θ_H or the low type θ_L, both equally likely. The sender chooses between two signals, $s \in \{\textbf{In}, \textbf{Out}\}$. If the sender plays **Out**, the game ends and both parties get 0 payoff. If the sender plays **In**, the receiver then chooses an action $a \in \{\textbf{Up}, \textbf{Down}\}$. Payoffs following the signal **In** depend on the sender's type and receiver's action, as in the following matrix:

Signal: **In**	Action: **Up**	Action: **Down**
Type: θ_H	$2, 2$	$-2, 0$
Type: θ_L	$1, -1$	$-3, 0$

Both sender types prefer (**In**, **Up**) to **Out** and prefer **Out** to (**In**, **Down**), while the receiver prefers **Up** over **Down** after signal **In** if he believes there is greater than $\frac{1}{3}$ chance that the sender has type θ_H.

[1] It is interesting to note that Spence (1973) also interpreted equilibria as steady states (or "nontransitory configurations") of a learning process, though he did not explicitly specify what sort of process he had in mind.

[2] As we explain in Corollary 1, our main result extends to environments where some fraction of the population has access to data about the play of others.

This game has a perfect Bayesian equilibrium (PBE) where both types choose **Out** and the receiver plays **Down** after **In**, sustained by the belief that anyone who sends **In** has probability $p \leq \frac{1}{3}$ of being θ_H. This updating requires the receiver to interpret the off-path **In** as a signal that the sender is more likely to be θ_L, even though θ_H gets 1 more utility than θ_L does from **In** regardless of the receiver's strategy. So, "both **Out**" is eliminated by the D1 criterion.[3]

Now suppose there are three infinitely lived agents: θ_H, θ_L, and R (for receiver). Suppose that in each period $t \in \{1, 2, 3, \ldots\}$, the three agents play a simultaneous-move game, where each sender type θ_i chooses a signal s_t^i, and R chooses a single action a_t to use against both of the senders. (This is a deterministic analog of the receiver randomly matching with each type with probability 1/2 without knowing the sender's type.) At the end of period t, R observes the signal choices of both types, while θ_i observes a_t if and only if $s_t^i = \mathbf{In}$. That is, each agent only learns from his/her personal experience; by choosing the "outside option" **Out**, the sender does not learn how the receiver would have responded to signal **In** that period.

Agents think that each opponent is committed to some mixed strategy of the stage game and plays this strategy each period, regardless of their observations of past play: that is, all agents are strategically myopic in the sense of Fudenberg and Kreps (1993) and do not try to influence the distribution of strategies they will face in future rounds. At the beginning of $t = 1$, each type θ_i is endowed with a Beta(c_U, c_D) prior about the probability that R responds to **In** with **Up**, with $c_D > c_U > 0$, so they assign higher probability to **Down** than to **Up**. R starts with two independent priors Beta(c_I^H, c_O^H) and Beta(c_I^L, c_O^L) about the probabilities that θ_H and θ_L choose **In** each period, where we only assume $c_I^H, c_O^H, c_I^L, c_O^L > 0$. The independence assumption means that R does not learn about the behavior of one type from the play of the other.

Agents discount payoffs in future periods at rate $0 \leq \delta < 1$ and choose a signal or action each period so as to maximize expected discounted payoffs. Because expected utility maximizing agents never strictly prefer to randomize, each of them has a deterministic optimal policy, so that each discount factor δ induces a deterministic infinite history of play $(s_t^H, s_t^L, a_t)_{t=1}^{\infty} =: Y(\delta)$. When $\delta = 0$, the agents play myopically every period, and because of our assumption that $c_D > c_U$, both types choose **Out** in $t = 1$. They thus gain no information about R's play, do not update their beliefs, and continue playing **Out** in every future period. So, the unintuitive "both **Out**" PBE is the learning outcome when agents are sufficiently impatient. However, we can show, for all large enough δ, that eventually behavior converges to R playing **Up** and θ_H playing **In** each period.[4]

We give a sketch of the argument, beginning with characterizing agents' optimal behavior each period. R observes the same information regardless of his play, so he plays myopically under any δ. Let $p(h_t)$ be R's Bayesian posterior belief about the probability that an **In** sender has type θ_H, given history h_t. Then $a_{t+1} = \mathbf{Up}$ if $p(h_t) > \frac{1}{3}$ and $a_{t+1} = \mathbf{Down}$ if $p(h_t) < \frac{1}{3}$.

Now we turn to θ_i, whose problem involves active experimentation. Formally, the dynamic optimization problem facing θ_i is a one-armed Bernoulli bandit. Choosing $s_t^i = \mathbf{Out}$ is equivalent to taking the safe outside option, while choosing $s_t^i = \mathbf{In}$ is equivalent to

[3]Any receiver play at the off-path signal **In** that makes it weakly optimal for θ_L to deviate to **In** would also make it strictly optimal for θ_H to deviate. Cho and Kreps's (1987) D1 criterion therefore requires the receiver to put 0 probability on $\theta = \theta_L$ after **In**. However, the PBE passes their Intuitive Criterion.

[4]In practice, the required patience level is not unreasonably high. When $c_D = 1.1$ $c_U = 1$, $c_I^H = c_O^L = 1$, and $c_O^H = c_I^L = 3$, for example, $\delta = 0$ yields the pathological PBE as the long-run outcome, but when $\delta \geq 0.92$, the long-run outcome involves $s_t^H = \mathbf{In}$ and $a_t = \mathbf{Up}$.

pulling the risky arm and getting a payoff depending on whether the pull results in a success ($a_t = \textbf{Up}$) or a failure ($a_t = \textbf{Down}$). The optimal policy for θ_i involves the Gittins index (defined later in Equation (2)). Type θ_i plays **In** at those histories where **In** has a positive Gittins index.

Once a type chooses to play **Out** in some period, she receives no further information and will continue to play **Out** in all subsequent periods. Denote the period in $Y(\delta)$ that θ_i first switches from **In** to **Out** as $T(i, \delta) \in \mathbb{N} \cup \{\infty\}$, where $T(i, \delta) = \infty$ means θ_i plays **In** forever. The argument that learning eliminates pooling on **Out** follows from three observations:

OBSERVATION 1: *The high type switches to **Out** later than the low type does*, that is, $T(H, \delta) \geq T(L, \delta)$. To see why, suppose by way of contradiction that $T(H, \delta) < T(L, \delta)$. Then, in period $t = T(H, \delta)$, both θ_H and θ_L have played **In** until now and have seen the same history, so they hold the same belief about R's play. Yet θ_H chooses **Out** at this history while θ_L chooses **In**, meaning θ_H has a negative Gittins index for **In** while θ_L has a positive one. This is impossible, since θ_H's payoff from **In** is always 1 higher than that of θ_L, so θ_H's index for **In** is also always 1 higher than that of θ_L when the two types have the same belief about R's play.

OBSERVATION 2: *As the high type becomes patient, she experiments with **In** arbitrarily many times*, that is, $\lim_{\delta \to 1} T(H, \delta) = \infty$. This follows because for any fixed full-support prior belief of θ_H about R's mixed strategy, the Gittins index for **In** stays close to the "success payoff" of 2 for a length of time that grows to infinity as $\delta \to 1$, even in the worst case where R plays **Down** in every period.

OBSERVATION 3: *If the high type plays **In** sufficiently many times and more often than the low type does, then eventually R will believe that **In** senders have greater than $\frac{1}{3}$ chance of being θ_H*, that is, there exists $\bar{N} \in \mathbb{N}$ so that $p(h_T) > \frac{1}{3}$ for any history h_T where (i) θ_H played **In** at least $\bar{N}$ times and (ii) θ_L played **In** no more than θ_H did. This follows from the fact that R's belief about θ_i's play after n_I^i instances of **In** and n_O^i instances of **Out** is $\text{Beta}(c_I^i + n_I^i, c_O^i + n_O^i)$.

From Observation 2, we see that $T(H, \delta)$ is larger than the $\bar{N}$ of Observation 3 when δ is sufficiently large. The history up to period t for any $t \geq \bar{N}$ will therefore contain at least $\bar{N}$ periods of θ_H playing **In** (namely, the very first $\bar{N}$ periods of the game), and by Observation 1, θ_L will have played **In** no more than θ_H did in this history. So by Observation 3, $p(h_t) > \frac{1}{3}$ for $t \geq \bar{N}$, meaning $a_t = \textbf{Up}$ for $t \geq \bar{N}$. Since $s_t^H = \textbf{In}$ for all $t \leq \bar{N}$ and observing **Up** increases the Gittins index of **In**, the high type must always play **In**. This means $\lim_{t \to \infty} s_t^H = \textbf{In}$ and $\lim_{t \to \infty} a_t = \textbf{Up}$ for large $\delta < 1$.

In this simple learning model, agents are patient and face the same opponents many times but do not try to influence their future play. Furthermore, agents believe that opponents' play is stationary but it changes markedly over time. Finally, the analysis was greatly simplified because there are only two signals, one of which gives a certain payoff to the senders; this makes playing **Out** an absorbing state and, together with the assumption of Beta priors, lets us explicitly calculate how the system evolves. This paper's focus is on general signaling games embedded in a learning model with large populations and anonymous random matching, eliminating repeated-game effects. We focus on steady states of the model, where the stationary assumption is satisfied. Also, we relax the Beta prior assumption and allow learners to have fairly general non-doctrinaire priors. Many results about the steady-state model, however, have analogs in the simple model above.

Intuitively, θ_H is "more compatible" with signal **In** than θ_L. Definition 2 formalizes this relation in general signaling games. Observation 1 corresponds to Lemma 2, which shows that whenever one type is more compatible than another with a signal, the more compatible type sends the signal more often. Observation 2 corresponds to Lemma 4, which says a sufficiently patient and long-lived sender type will experiment many times with all signals that have the potential to strictly improve that type's equilibrium payoff. Observation 3 corresponds to Lemma 3, which says receivers can eventually learn the compatibility relation associated with each signal, provided senders' play respects the relation and the more compatible type experiments enough with the signal. Lemmas 2, 3, and 4 are combined to prove the main result of the paper (Theorem 2), a learning-based refinement in general signaling games.

1.2. *Outline and Overview of Results*

Section 2 lays out the notation we will use for signaling games and introduces our learning model. Section 3 introduces the Gittins index, which we use to analyze the senders' learning problem. It also defines type compatibility, which is a partial order that drives our results. We say that type θ' is *more type-compatible with signal* s' than type θ'' if, whenever s' is a weak best response for θ'' against some receiver behavior strategy, it is a strict best response for θ' against the same strategy. To relate this static definition to the senders' optimal dynamic learning behavior, we show that, under our assumptions, the senders' learning problem is formally a multi-armed bandit, so the optimal policy of each type is characterized by the Gittins index. Theorem 1 shows that the compatibility order on types is equivalent to an order on their Gittins indices: θ' is more type-compatible with signal s' than type θ'' if and only if, whenever s' has the (weakly) highest Gittins index for θ'', it has the strictly highest index for θ', provided the two types hold the same beliefs and have the same discount factor.

Section 4 studies the aggregate behavior of the sender and receiver populations. There we define and characterize the *aggregate responses* of the senders and of the receivers, which are the analogs of the best-response functions in the one-shot signaling game. First, we use a coupling argument to extend Theorem 1 to the aggregate sender behavior, proving that types who are more compatible with a signal send it more often in aggregate (Lemma 2). Then we turn to the receivers. Intuitively, we would expect that when receivers are long-lived, most of them will have beliefs that respect type compatibility, and we show that this is the case. More precisely, we show that most receivers best respond to a posterior belief whose likelihood ratio of θ' to θ'' dominates the prior likelihood ratio of these two types whenever they observe a signal s which is more type-compatible with θ' than θ''. Lemma 3 shows this is true for any signal that is sent "frequently enough" relative to the receivers' expected lifespan, using a result of Fudenberg, He, and Imhof (2017) on updating posteriors after rare events.

Finally, Section 5 combines the earlier results to characterize the steady states of the learning model, which can be viewed as pairs of mutual aggregate responses, analogous to the definition of Nash equilibrium. We start by proving Lemma 4, which shows that any signal that is not *weakly equilibrium dominated* (see Definition 11) gets sent "frequently enough" in steady state when senders are sufficiently patient and long lived. Combining the three lemmas discussed above, we establish our main result: any patiently stable steady state must be a Nash equilibrium satisfying the additional restriction that the receivers best respond to certain *admissible beliefs* after every off-path signal (Theorem 2).

As an example, consider Cho and Kreps's (1987) beer-quiche game, where it is easy to verify that the strong type is more compatible with **Beer** than the weak type. Our results

imply that the strong types will in aggregate send this signal at least as often as the weak types do, and that a very patient strong type will experiment with it "many times." As a consequence, when senders are patient, long-lived receivers are unlikely to revise the probability of the strong type downwards following an observation of **Beer**. Thus, the "both types eat quiche" equilibrium is not a patiently stable steady state of the learning model, as it would require receivers to interpret **Beer** as a signal that the sender is weak.

Finally, Theorem 3 provides a stronger implication of patient stability in generic pure-strategy equilibria, showing that off-path beliefs must assign probability zero to types that are equilibrium dominated in the sense of Cho and Kreps (1987).

1.3. *Related Work*

Fudenberg and Kreps (1988, 1994, 1995) pointed out that experimentation plays an important role in determining learning outcomes in extensive-form games. As in Fudenberg and Kreps (1993), they studied a model with a single infinitely-lived and strategically myopic agent in each player role who acts as if the opponent's play is stationary. Because these models involved accumulating information over time, they did not have steady states. Our work is closer to that of Fudenberg and Levine (1993) and Fudenberg and Levine (2006) which also studied learning by Bayesian agents in a large population who believe that society is in a steady state. A key issue in this work, and more generally in studying learning in extensive-form games, is characterizing how much agents will experiment with myopically suboptimal actions. If agents do not experiment at all, then non-Nash equilibria can persist, because players can maintain incorrect but self-confirming beliefs about off-path play. Fudenberg and Levine (1993) showed that patient long-lived agents will experiment enough at their on-path information sets to learn if they have any profitable deviations, thus ruling out steady states that are not Nash equilibria. However, more experimentation than that is needed for learning to generate the sharper predictions associated with backward induction and sequential equilibrium. Fudenberg and Levine (2006) showed that patient rational agents need not do enough experimentation to imply backwards induction in games of perfect information. Later on, we say how the models and proofs of those papers differ from ours.

This paper is also related to the Bayesian learning models of Kalai and Lehrer (1993), which studied two-player games with one agent on each side, so that every self-confirming equilibrium is path-equivalent to a Nash equilibrium, and Esponda and Pouzo (2016), which allowed agents to experiment but did not characterize when and how this occurs. It is also related to the literature on boundedly rational experimentation in extensive-form games (e.g., Jehiel and Samet (2005), Laslier and Walliser (2015)), where the experimentation rules of the agents are exogenously specified. We assume that each sender's type is fixed at birth, as opposed to being i.i.d. over time. Dekel, Fudenberg, and Levine (2004) showed some of the differences this can make using various equilibrium concepts, but they did not develop an explicit model of non-equilibrium learning.

For simplicity, we assume here that agents do not know the payoffs of other players and have full support priors over the opposing side's behavior strategies. Our companion paper Fudenberg and He (2017) supposed that players assign zero probability to dominated strategies of their opponents, as in the Intuitive Criterion (Cho and Kreps (1987)), divine equilibrium (Banks and Sobel (1987)), and rationalizable self-confirming equilibrium (Dekel, Fudenberg, and Levine (1999)). There, we analyzed how the resulting micro-founded equilibrium refinement compares to those in past work.

2. MODEL

2.1. *Signaling Game Notation*

A *signaling game* has two players, a sender (player 1, "she") and a receiver (player 2, "he"). 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.[5] There is a finite set S of signals for the sender and a finite set A of actions for the receiver.[6] The utility functions of the sender and receiver are $u_1 : \Theta \times S \times A \to \mathbb{R}$ and $u_2 : \Theta \times S \times A \to \mathbb{R}$, respectively.

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

A *behavior strategy for the sender* $\pi_1 = (\pi_1(\cdot|\theta))_{\theta\in\Theta}$ is a type-contingent mixture over signals S. Write Π_1 for the set of all sender behavior strategies.

A *behavior strategy for the receiver* $\pi_2 = (\pi_2(\cdot|s))_{s\in S}$ is a signal-contingent mixture over actions A. Write Π_2 for the set of all receiver behavior strategies.

2.2. *Learning by Individual Agents*

We now build a learning model with a given signaling game as the stage game. In this subsection, we explain an individual agent's learning problem. In the next subsection, we complete the learning model by describing a society of learning agents who are randomly matched to play the signaling game every period.

Time is discrete and all agents are rational Bayesians with geometrically distributed lifetimes. They survive between periods with probability $0 \leq \gamma < 1$ and further discount future utility flows by $0 \leq \delta < 1$, so their objective is to maximize the expected value of $\sum_{t=0}^{\infty}(\gamma\delta)^t \cdot u_t$. Here, $0 \leq \gamma\delta < 1$ is the effective discount factor, and u_t is the payoff t periods from today.

At birth, each agent is assigned a role in the signaling game: either as a sender with type θ or as a receiver. Agents know their role, which is fixed for life. Every period, each agent is randomly and anonymously matched with an opponent to play the signaling game, and the game's outcome determines the agent's payoff that period. At the end of each period, agents observe the outcomes of their own matches, that is, the signal sent, the action played in response, and the sender's type. They do not observe the identity, age, or past experiences of their opponents, nor does the sender observe how the receiver would have reacted to a different signal.[7] Agents update their beliefs and play the signaling game again with new random opponents next period, provided they are still alive.

Agents believe they face a fixed but unknown distribution of opponents' aggregate play, so they believe that their observations will be exchangeable. We feel that this is a plausible first hypothesis in many situations, so we expect that agents will maintain their belief in stationarity when it is approximately correct, but will reject it given clear evidence to the

[5]Here and subsequently, $\Delta(X)$ denotes the collection of probability distributions on the set X.

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

[7]The receiver's payoff reveals the sender's type for generic assignments of payoffs to terminal nodes. If the receiver's payoff function is independent of the sender's type, his beliefs about it are irrelevant. If the receiver does care about the sender's type but observes neither the sender's type nor his own realized payoff, a great many outcomes can persist, as in Dekel, Fudenberg, and Levine (2004).

contrary, as when there is a strong time trend or a high-frequency cycle. The environment will indeed be constant in the steady states that we analyze.

Formally, each sender is born with a prior density function over the aggregate behavior strategy of the receivers, $g_1 : \Pi_2 \to \mathbb{R}_+$, which integrates to 1. Similarly, each receiver is born with a prior density over the sender's behavior strategies,[8] $g_2 : \Pi_1 \to \mathbb{R}_+$. We denote the marginal distribution of g_1 on signal s as $g_1^{(s)}$, so that $g_1^{(s)}(\pi_2(\cdot|s))$ is the density of the new senders' prior over how receivers respond to signal s. Similarly, we denote the θ marginal of g_2 as $g_2^{(\theta)}$, so that $g_2^{(\theta)}(\pi_1(\cdot|\theta))$ is the new receivers' prior density over $\pi_1(\cdot|\theta) \in \Delta(S)$.

It is important to remember that g_1 and g_2 are beliefs over opponents' strategies, but not strategies themselves. A new sender expects the response to s to be $\int \pi_2(\cdot|s) \cdot g_1(\pi_2)\, d\pi_2$, while a new receiver expects type θ to play $\int \pi_1(\cdot|\theta) \cdot g_2(\pi_1)\, d\pi_1$.

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

DEFINITION 1: A prior $g = (g_1, g_2)$ is **regular** if:

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

(ii) [g_1 *non-doctrinaire*] g_1 is continuous and strictly positive on the interior of Π_2.

(iii) [g_2 *nice*] for each type θ, there are positive constants $(\alpha_s^{(\theta)})_{s\in S}$ such that

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

is uniformly continuous and bounded away from zero on the relative interior of $\Pi_1^{(\theta)}$, the set of behavior strategies of type θ.

Independence ensures that a receiver does not learn how type θ plays by observing the behavior of some other type $\theta' \neq \theta$, and that a sender does not learn how receivers react to signal s by experimenting with some other signal $s' \neq s$. For example, this means in Cho and Kreps's (1987) beer-quiche game that the sender does not learn how receivers respond to beer by eating quiche.[9] The non-doctrinaire nature of g_1 and g_2 implies that the agents never see an observation that they assigned zero prior probability, so that they have a well-defined optimization problem after any history. Non-doctrinaire priors also imply that a large enough data set can outweigh prior beliefs (Diaconis and Freedman (1990)). The niceness assumption in (iii) ensures that g_2 behaves like a power function near the boundary of Π_1. Any density that is strictly positive on Π_1 satisfies this condition, as does the Dirichlet distribution, which is the prior associated with fictitious play (Fudenberg and Kreps (1993)).

[8]Note that the agent's prior belief is over opponents' *aggregate* play (i.e., Π_1 or Π_2) and not over the prevailing distribution of behavior strategies in the opponent population (i.e., $\Delta(\Pi_2)$ or $\Delta(\Pi_1)$), since under our assumption of anonymous random matching, these are observationally equivalent for our agents. For instance, a receiver cannot distinguish between a society where all type θ randomize 50–50 between signals s_1 and s_2 each period, and another society where half of the type θ always plays s_1 while the other half always plays s_2. Note also that because agents believe the system is in a steady state, they do not care about calendar time and do not have beliefs about it. Fudenberg and Kreps (1994) supposed that agents append a non-Bayesian statistical test of whether their observations are exchangeable to a Bayesian model that presumes exchangeability.

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

The set of histories for an age t sender of type θ is $Y_\theta[t] := (S \times A)^t$, where, each period, the history records the signal sent and the action that her receiver opponent took in response. The set of all histories for a type θ is the union $Y_\theta := \bigcup_{t=0}^{\infty} Y_\theta[t]$. The dynamic optimization problem of type θ has an optimal policy function $\sigma_\theta : Y_\theta \to S$, where $\sigma_\theta(y_\theta)$ is the signal that a type θ with history y_θ would send the next time she plays the signaling game. Analogously, the set of histories for an age t receiver is $Y_2[t] := (\Theta \times S)^t$, where, each period, the history records the type of his sender opponent and the signal that she sent. The set of all receiver histories is the union $Y_2 := \bigcup_{t=0}^{\infty} Y_2[t]$. The receiver's learning problem admits an optimal policy function $\sigma_2 : Y_2 \to A^S$, where $\sigma_2(y_2)$ is the pure strategy that a receiver with history y_2 would commit to next time he plays the game.[10]

2.3. *Random Matching and Aggregate Play*

We analyze learning in a deterministic stationary model with a continuum of agents, as in Fudenberg and Levine (1993, 2006). One innovation is that we let lifetimes follow a geometric distribution instead of the finite and deterministic lifetimes assumed in those earlier papers, so that we can use the Gittins index.

The society contains a unit mass of agents in the role of receivers and mass $\lambda(\theta)$ in the role of type θ for each $\theta \in \Theta$. As described in Section 2.2, each agent has $0 \leq \gamma < 1$ chance of surviving at the end of each period and complementary chance $1 - \gamma$ of dying. To preserve population sizes, $(1 - \gamma)$ new receivers and $\lambda(\theta)(1 - \gamma)$ new type θ are born into the society every period.

Each period, agents in the society are matched uniformly at random to play the signaling game. In the spirit of the law of large numbers, each sender has probability $(1 - \gamma)\gamma^t$ of matching with a receiver of age t, while each receiver has probability $\lambda(\theta)(1 - \gamma)\gamma^t$ of matching with a type θ of age t.

A *state* ψ of the learning model is described by the mass of agents with each possible history. We write it as

$$\psi \in \Big(\underset{\theta \in \Theta}{\times} \Delta(Y_\theta) \Big) \times \Delta(Y_2).$$

We refer to the components of a state ψ by $\psi_\theta \in \Delta(Y_\theta)$ and $\psi_2 \in \Delta(Y_2)$.

Given the agents' optimal policies, each possible history for an agent completely determines how that agent will play in their next match. The sender policy functions σ_θ are maps from sender histories to signals,[11] so they naturally extend to maps from distributions over sender histories to distributions over signals. That is, given the policy function σ_θ, each state ψ induces an aggregate behavior strategy $\sigma_\theta(\psi_\theta) \in \Delta(S)$ for each type θ population, where we extend the domain of σ_θ from Y_θ to $\Delta(Y_\theta)$ in the natural way:

$$\sigma_\theta(\psi_\theta)(s) := \psi_\theta\big\{y_\theta \in Y_\theta : \sigma_\theta(y_\theta) = s\big\}. \tag{1}$$

Similarly, state ψ and the optimal receiver policy σ_2 together induce an aggregate behavior strategy $\sigma_2(\psi_2)$ for the receiver population, where

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

[10] Because our agents are expected-utility maximizers, it is without loss of generality to assume each agent uses a deterministic policy rule. If more than one such rule exists, we fix one arbitrarily. Of course, the optimal policies σ_θ and σ_2 depend on the prior g as well as the effective discount factor $\delta\gamma$. Where no confusion arises, we suppress these dependencies.

[11] Remember that we have fixed deterministic policy functions.

We will study the steady states of this learning model, to be defined more precisely in Section 5. Loosely speaking, a steady state is a state ψ that reproduces itself indefinitely when agents use their optimal policies. Put another way, a steady state induces a time-invariant distribution over how the signaling game is played in the society. Suppose society is at steady state today and we measure what fraction of type θ sent a certain signal s in today's matches. After all agents modify their strategies based on their updated beliefs and all births and deaths take place, the fraction of type θ playing s in the matches tomorrow will be the same as today.

3. SENDERS' OPTIMAL POLICIES AND TYPE COMPATIBILITY

This section studies the senders' learning problem. We will prove that differences in the payoff structures of the various sender types generate certain restrictions on their behavior in the learning model. Section 3.1 notes that the senders face a multi-armed bandit, so the Gittins index characterizes their optimal policies, and shows how to relate the Gittins index of a signal to the expected sender payoff versus a particular mixed strategy of the receiver. In Section 3.2, we define *type compatibility*, which formalizes what it means for type θ' to be more "compatible" with a given signal s than type θ'' is. The definition of type compatibility is static, in the sense that it depends only on the two types' payoff functions in the one-shot signaling game. Section 3.3 relates type compatibility to the Gittins index, which applies to the dynamic learning model. Lemma 2 in Section 4 uses this relationship to show that if type θ' is more compatible with signal s than type θ'', then, faced with any fixed distribution of receiver play, the type θ' population sends s more often in the aggregate than the type θ'' population does.

3.1. *Optimal Policies and Multi-Armed Bandits*

Each type θ sender thinks she is facing a fixed but unknown aggregate receiver behavior strategy π_2, so each period when she sends signal s, she believes that the response is drawn from some $\pi_2(\cdot|s) \in \Delta(A)$, i.i.d. across periods. Because her beliefs about the responses to the various signals are independent, her problem is equivalent to a discounted multi-armed bandit, with signals $s \in S$ as the arms, where the rewards of arm s are distributed according to $u_1(\theta, s, \pi_2(\cdot|s))$.

Let $\nu_s \in \Delta(\Delta(A))$ be a belief over the space of mixed replies to signal s, and let $\nu = (\nu_s)_{s\in S}$ be a profile of such beliefs. Write $I(\theta, s, \nu, \beta)$ for the Gittins index of signal s for type θ, with beliefs ν over receiver's play after various signals and with effective discount factor $\beta = \delta\gamma$, so that

$$I(\theta, s, \nu, \beta) := \sup_{\tau>0} \frac{\mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau-1} \beta^t \cdot u_1\big(\theta, s, a_s(t)\big)\right\}}{\mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau-1} \beta^t\right\}}. \tag{2}$$

Here $a_s(t)$ is the receiver's response that the sender observes the tth time she sends signal s, τ is a stopping time,[12] and the expectation $\mathbb{E}_{\nu_s}$ over the sequence of responses $\{a_s(t)\}_{t\geq 0}$ depends on the sender's belief ν_s about responses to signal s.[13]

[12]That is, whether or not $\tau = t$ depends only on the realizations of $a_s(0), a_s(1), \ldots, a_s(t-1)$.

[13]The Gittins index can be interpreted as the value of an auxiliary optimization problem, where type θ chooses each period to either send signal s and obtain a payoff according to a random receiver action drawn

The Gittins index theorem (Gittins (1979)) implies that after every positive-probability history y_θ, the optimal policy σ_θ for a sender of type θ sends the signal that has the highest Gittins index for that type under the profile of posterior beliefs $(\nu_s)_{s\in S}$ that is induced by y_θ.

Importantly, we can reformulate the objective function defining the Gittins index in Equation (2), linking it to the one-shot signaling game payoff structure.

LEMMA 1: *For every signal s, stopping time τ, belief ν_s, and discount factor β, there exists $\pi_{2,s}(\tau, \nu_s, \beta) \in \Delta(A)$ so that for every θ,*

$$\frac{\mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau-1}\beta^t \cdot u_1\big(\theta, s, a_s(t)\big)\right\}}{\mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau-1}\beta^t\right\}} = u_1\big(\theta, s, \pi_{2,s}(\tau, \nu_s, \beta)\big).$$

That is to say, when the stopping problem in Equation (2) is evaluated at an arbitrary stopping time τ, the payoff is equal to sender's expected utility from playing s against the receiver strategy $\pi_{2,s}(\tau, \nu_s, \beta)$ in the one-shot signaling game.

The proof of Lemma 1 is in Appendix A.2 and shows how to construct $\pi_{2,s}(\tau, \nu_s, \beta)$, which can be interpreted as a discounted time average over the receiver actions that are observed before stopping. To illustrate the construction, suppose ν_s is supported on two pure receiver strategies after s: either $\pi_2(a'|s) = 1$ or $\pi_2(a''|s) = 1$, with both strategies equally likely. Suppose also $u_1(\theta, s, a') > u_1(\theta, s, a'')$. Consider the stopping time τ that specifies stopping after the first time the receiver plays a''. Then the discounted time average frequency of a'' is

$$\frac{\sum_{t=0}^{\infty}\beta^t \cdot \mathbb{P}_{\nu_s}\big[\tau \geq t \text{ and receiver plays } a'' \text{ in period } t\big]}{\sum_{t=0}^{\infty}\beta^t \cdot \mathbb{P}_{\nu_s}[\tau \geq t]} = \frac{0.5}{1+\sum_{t=1}^{\infty}\beta^t \cdot 0.5} = \frac{1-\beta}{2-\beta}.$$

So $\pi_{2,s}(\tau, \nu_s, \beta)(a'') = \frac{1-\beta}{2-\beta}$ and similarly, we can calculate that $\pi_{2,s}(\tau, \nu_s, \beta)(a') = \frac{1}{2-\beta}$, which shows that $\pi_{2,s}$ indeed corresponds to a mixture over receiver actions for each β. As $\beta \to 1$, this mixture converges to the pure strategy of always playing a', so $u_1(\theta, s, \pi_{2,s}(\tau, \nu_s, \beta))$ converges to $u_1(\theta, s, a')$, the highest possible payoff for type θ after s; this parallels the fact that as β tends to 1, the Gittins index for θ after s converges to the highest payoff in the support of the belief ν_s.

according to $\pi_2(\cdot|s)$, or to stop forever. The objective of the auxiliary problem is to maximize the per-period expected discounted payoff until stopping, as the numerator of Equation (2) describes the expected discounted sum of payoffs until stopping while the denominator shows the expected discounted number of periods until stopping.

3.2. *Type Compatibility in Signaling Games*

We now introduce a notion of the comparative compatibility of two types with a given signal in the one-shot signaling game.

DEFINITION 2: Signal s' is *more type-compatible* with θ' than θ'', written as $\theta' \succsim_{s'} \theta''$, if for every $\pi_2 \in \Pi_2$ such that

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

we have

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

In words, $\theta' \succsim_{s'} \theta''$ means that whenever s' is a weak best response for θ'' against some receiver behavior strategy π_2, it is also a strict best response for θ' against π_2.

The following proposition says the compatibility order is transitive and essentially asymmetric. Its proof is in Appendix A.1.

PROPOSITION 1:

(i) *$\succsim_{s'}$ is transitive.*

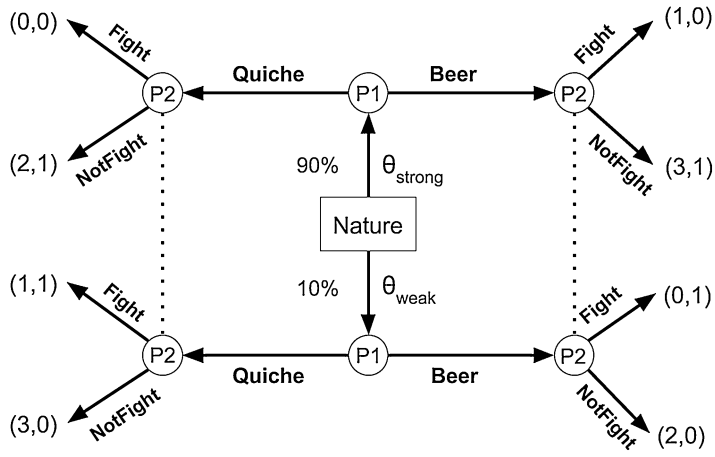
(ii) *Except when s' is either strictly dominant for both θ' and θ'' or strictly dominated for both θ' and θ'', $\theta' \succsim_{s'} \theta''$ implies $\theta'' \not\succsim_{s'} \theta'$.*

To check the compatibility condition, one must consider all strategies in Π_2, 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 $u_1(\theta, s, a) = v(\theta, s) + z(a)$, as in Spence's (1973) job-market signaling game and in Cho and Kreps's (1987) beer-quiche game (given below), a sufficient condition for $\theta' \succsim_{s'} \theta''$ is

$$v\big(\theta', s'\big) - v\big(\theta'', s'\big) > \max_{s'' \neq s'} v\big(\theta', s''\big) - v\big(\theta'', s''\big).$$

This can be interpreted as saying s' is the least costly signal for θ' relative to θ''. In the Supplemental Material (Fudenberg and He (2018)), we present a general sufficient condition for $\theta' \succsim_{s'} \theta''$ under general payoff functions.

EXAMPLE 1—Cho and Kreps's (1987) Beer-Quiche Game: The sender (P1) 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 (P2), observing this breakfast choice but not the sender's type, chooses whether to **Fight** the sender. If the sender is θ_{weak}, the receiver prefers to **Fight**. If the sender is θ_{strong}, the receiver prefers to **NotFight**. 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.



This game has separable sender utility with $v(\theta_{\text{strong}}, \textbf{Beer}) = v(\theta_{\text{weak}}, \textbf{Quiche}) = 1$, $z(\textbf{Fight}) = 0$, and $z(\textbf{NotFight}) = 2$. So, we have $\theta_{\text{strong}} \succsim_{\textbf{Beer}} \theta_{\text{weak}}$.

It is easy to see that in every Nash equilibrium π^*, if $\theta' \succsim_{s'} \theta''$, then $\pi_1^*(s'|\theta'') > 0$ implies $\pi_1^*(s'|\theta') = 1$. By Bayes's rule, this implies that the receiver's equilibrium belief p after every *on-path* signal s' satisfies the restriction $\frac{p(\theta''|s')}{p(\theta'|s')} \leq \frac{\lambda(\theta'')}{\lambda(\theta')}$ if $\theta' \succsim_{s'} \theta''$. Thus, in every Nash equilibrium of the beer-quiche game, if the sender chooses **Beer** with positive ex ante probability, then the receiver's odds ratio that the sender is tough after seeing this signal cannot be less than the prior odds ratio. Our main result, Theorem 2, essentially shows, for any strategy profile that can be approximated by steady-state outcomes with patient and long-lived agents, that the same compatibility-based restriction is satisfied even for *off-path* signals. In particular, this allows us to place restrictions on the receiver's belief after seeing **Beer** in equilibria where no type of sender ever plays this signal.

3.3. *Type Compatibility and the Gittins Index*

We now connect the type compatibility order for a given signal with the associated Gittins indices.

THEOREM 1: *$\theta' \succsim_{s'} \theta''$ if and only if, for every $\beta \in [0, 1)$ and every profile of beliefs ν, $I(\theta'', s', \nu, \beta) \geq \max_{s'' \neq s'} I(\theta'', s'', \nu, \beta)$ implies $I(\theta', s', \nu, \beta) > \max_{s'' \neq s'} I(\theta', s'', \nu, \beta)$.*

That is, $\theta' \succsim_{s'} \theta''$ if and only if whenever s' has the (weakly) highest Gittins index for θ'', it has the highest index for θ', provided the two types hold the same beliefs and have the same discount factor. The proof involves reformulating the Gittins index as in Lemma 1, then applying the compatibility definition.

PROOF OF THEOREM 1: *Step 1: Only If.*

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

On any arm $s'' \neq s'$, type θ'' could use the (suboptimal) stopping time $\tau_{s''}^{\theta'}$, which by Lemma 1 yields an expected per-period payoff of $u_1(\theta'', s'', \pi_{s''}(\nu_{s''}, \tau_{s''}^{\theta'}, \beta))$. This is a lower bound for the Gittins index of arm s'' for type θ'', so combined with the hypothesis that

$I(\theta'', s', \nu, \beta) \geq \max_{s'' \neq s'} I(\theta'', s'', \nu, \beta)$, we get

$$I\big(\theta'', s', \nu, \beta\big) \geq \max_{s'' \neq s'} u_1\big(\theta'', s'', \pi_{s''}\big(\nu_{s''}, \tau_{s''}^{\theta'}, \beta\big)\big). \tag{3}$$

Now define the receiver strategy $\pi_2 \in \Pi_2$ by $\pi_2(\cdot|s') := \pi_{s'}(\nu_{s'}, \tau_{s'}^{\theta''}, \beta)$, $\pi_2(\cdot|s'') := \pi_{s''}(\nu_{s''}, \tau_{s''}^{\theta'}, \beta)$ for all $s'' \neq s'$. Then Equation (3) can be rewritten as

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

that is, s' is weakly optimal for θ'' against π_2. By the definition of $\theta' \succsim_{s'} \theta''$, this implies s' is strictly optimal for θ' against π_2.

From the definition of π_2 and Lemma 1, the expected utility of θ' playing any $s'' \neq s'$ against π_2 is equal to the Gittins index of that arm for θ', namely, $I(\theta', s'', \nu, \beta)$. On the other hand, $u_1(\theta', s', \pi_2(\cdot|s'))$ is only a lower bound for $I(\theta', s', \nu, \beta)$. This shows $I(\theta', s', \nu, \beta) > \max_{s'' \neq s'} I(\theta', s'', \nu, \beta)$ as desired.

Step 2: If.

Suppose $\theta' \not\succsim_{s'} \theta''$. Then there is some receiver strategy $\pi_2^* \in \Pi_2$ such that

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

and

$$u_1\big(\theta', s', \pi_2^*\big(\cdot|s'\big)\big) \leq \max_{s'' \neq s'} u_1\big(\theta', s'', \pi_2^*\big(\cdot|s''\big)\big).$$

Let ν^* be any belief that induces π_2^* on average, that is to say, for each s,

$$\pi_2^*(\cdot|s) = \int_{\pi_{2,s} \in \Delta(A)} \pi_{2,s}\, d\nu_s^*(\pi_{2,s}).$$

Let $\beta = 0$. Then $I(\theta, s, \nu^*, 0) = u_1(\theta, s, \pi_2^*(\cdot|s))$ for every θ, s, since the Gittins index is equal to the myopic payoff when the decision-maker is perfectly impatient. This shows $I(\theta'', s', \nu^*, 0) \geq \max_{s'' \neq s'} I(\theta'', s'', \nu^*, 0)$ and $I(\theta', s', \nu^*, 0) \leq \max_{s'' \neq s'} I(\theta', s'', \nu^*, 0)$. *Q.E.D.*

4. THE AGGREGATE SENDER AND RECEIVER RESPONSES

In this section, we will define and analyze the aggregate sender response $\mathcal{R}_1 : \Pi_2 \to \Pi_1$ and the aggregate receiver response $\mathcal{R}_2 : \Pi_1 \to \Pi_2$. Loosely speaking, these are the large-populations learning analogs of the best-response functions in the static signaling game. If we fix the aggregate play of $-i$ population at π_{-i} and run the learning model period after period from an arbitrary initial state, the distribution of play in i population will approach $\mathcal{R}_i[\pi_{-i}]$. Later, in Section 5, the fixed points of the pair $(\mathcal{R}_1, \mathcal{R}_2)$ will characterize the steady states of the learning system.

4.1. *The Aggregate Sender Response*

To formally define the aggregate sender response, we first introduce the one-period-forward map.

DEFINITION 3: The *one-period-forward map for type* θ, $f_\theta : \Delta(Y_\theta) \times \Pi_2 \to \Delta(Y_\theta)$, is

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

and $f_\theta[\psi_\theta, \pi_2](\emptyset) := 1 - \gamma$.

If the distribution over histories in the type θ population is ψ_θ and the receiver population's aggregate play is π_2, the resulting distribution over histories in the type θ population is $f_\theta[\psi_\theta, \pi_2]$. Specifically, there will be a $1 - \gamma$ mass of new type θ who will have no history. Also, if the optimal first signal of a new type θ is s', that is, if $\sigma_\theta(\emptyset) = s'$, then $f_\theta[\psi_\theta, \pi_2](s', a') = \gamma \cdot (1 - \gamma) \cdot \pi_2(a'|s')$ new senders send s' in their first match, observe action a' in response, and survive. In general, a type θ who has history y_θ and whose policy $\sigma_\theta(y_\theta)$ prescribes playing s has $\pi_2(a|s)$ chance of having subsequent history $(y_\theta, (s, a))$ provided she survives until next period; the survival probability corresponds to the factor γ.

Write f_θ^T for the T-fold application of f_θ on $\Delta(Y_\theta)$, holding fixed some π_2. Note that for arbitrary states ψ and ψ', if $(y_\theta, (s, a))$ is a length-1 history (i.e., $y_\theta = \emptyset$), then $\psi_\theta(y_\theta) = \psi'_\theta(y_\theta)$ because both states must assign mass $1 - \gamma$ to $\emptyset$, so $f_\theta^1[\psi_\theta, \pi_2]$ and $f_\theta^1[\psi'_\theta, \pi_2]$ agree on $Y_\theta[1]$. Iterating, for $T = 2$, $f_\theta^2[\psi_\theta, \pi_2]$ and $f_\theta^2[\psi'_\theta, \pi_2]$ agree on $Y_\theta[2]$, because each history in $Y_\theta[2]$ can be written as $(y_\theta, (s, a))$ for $y_\theta \in Y_\theta[1]$, and $f_\theta^1[\psi_\theta, \pi_2]$ and $f_\theta^1[\psi'_\theta, \pi_2]$ match on all $y_\theta \in Y_\theta[1]$. Proceeding inductively, we can conclude that $f_\theta^T(\psi_\theta, \pi_2)$ and $f_\theta^T(\psi'_\theta, \pi_2)$ agree on all $Y_\theta[t]$ for $t \leq T$ for any pair of type θ states ψ_θ and ψ'_θ. This means $\lim_{T\to\infty} f_\theta^T(\psi_\theta, \pi_2)$ exists and is independent of the initial state ψ_θ. Denote this limit as $\psi_\theta^{\pi_2}$. It is the long-run distribution over type θ histories induced by starting at an arbitrary state and fixing the receiver population's play at π_2, as stated formally in the next definition.

DEFINITION 4: The *aggregate sender response* $\mathcal{R}_1 : \Pi_2 \to \Pi_1$ is defined by

$$\mathcal{R}_1[\pi_2](s|\theta) := \psi_\theta^{\pi_2}\big(y_\theta : \sigma_\theta(y_\theta) = s\big),$$

where $\psi_\theta^{\pi_2} := \lim_{T\to\infty} f_\theta^T(\psi_\theta, \pi_2)$ with ψ_θ any arbitrary θ state.

That is, $\mathcal{R}_1[\pi_2](\cdot|\theta)$ is the long-run aggregate behavior in the type θ population when the receivers' aggregate play is fixed at π_2.

REMARK 1: Technically, $\mathcal{R}_1$ depends on g_1, δ, and γ, just like σ_θ does. When relevant, we will make these dependencies clear by adding the appropriate parameters as superscripts to $\mathcal{R}_1$, but we will mostly suppress them to lighten notation.

REMARK 2: Although the aggregate sender response is defined at the aggregate level, $\mathcal{R}_1[\pi_2](\cdot|\theta)$ also describes the probability distribution of the play of a single type θ sender over her lifetime when she faces receiver play drawn from π_2 every period.[14]

[14]Observe that $f_\theta[\psi_\theta, \pi_2]$ restricted to $Y_\theta[1]$ gives the probability distribution over histories for a type θ who uses σ_θ and faces play drawn from π_2 for one period: it puts weight $\pi_2(a'|s')$ on history (s', a') where $s' = \sigma_\theta(\emptyset)$. Similarly, $f_\theta^T[\psi_\theta, \pi_2]$ restricted to $Y_\theta[t]$ for any $t \leq T$ gives the probability distribution over histories for someone who uses σ_θ and faces play drawn from π_2 for t periods. Since $\psi_\theta^{\pi_2}$ assigns probability $(1 - \gamma)\gamma^t$ to the set of histories $Y_\theta[t]$, $\mathcal{R}_1[\pi_2](\cdot|\theta) = \sigma_\theta(\psi_\theta^{\pi_2})$ is a weighted average over the distributions of period t play ($t = 1, 2, 3, \ldots$) of someone using σ_θ and facing π_2, with weight $(1 - \gamma)\gamma^t$ given to the period t distribution.

4.2. *Type Compatibility and the Aggregate Sender Response*

The next lemma shows how type compatibility translates into restrictions on the aggregate sender response for different types.

LEMMA 2: *Suppose* $\theta' \succsim_{s'} \theta''$. *Then, for any regular prior* g_1, $0 \leq \delta, \gamma < 1$, *and any* $\pi_2 \in \Pi_2$, *we have* $\mathcal{R}_1[\pi_2](s'|\theta') \geq \mathcal{R}_1[\pi_2](s'|\theta'')$.

Theorem 1 showed that when $\theta' \succsim_{s'} \theta''$ and the two types share the same beliefs, if θ''plays s', then θ' must also play s'. But even though new agents of both types start with the same prior g_1, their beliefs may quickly diverge during the learning process due to $\sigma_{\theta'}$ and $\sigma_{\theta''}$ prescribing different experiments after the same history. This lemma shows that compatibility still imposes restrictions on the aggregate play of the sender population: Regardless of the aggregate play π_2 in the receiver population, the frequencies that s' appears in the aggregate responses of different types are always co-monotonic with the compatibility order $\succsim_{s'}$.

To gain intuition for Lemma 2, consider two new senders with types θ_{strong} and θ_{weak} who are learning to play the beer-quiche game from Example 1. Suppose they have uniform priors over the responses to each signal, and that they face a sequence of receivers programmed to play **Fight** after **Beer** and **NotFight** after **Quiche**. Since observing **Fight** is the worst possible news about a signal's payoff, the Gittins index of a signal decreases when **Fight** is observed. Conversely, the Gittins index of a signal increases after each observation of **NotFight**.[15] Thus, given the assumed play of the receivers, there are $n_1, n_2 \geq 0$ such that type θ_{strong} play **Beer** for n_1 periods (and observe n_1 instances of **Fight**) and then switch to **Quiche** forever after, while type θ_{weak} will play **Beer** for n_2 periods before switching to **Quiche** forever after. Now we claim that $n_1 \geq n_2$. To see why, suppose instead that $n_1 < n_2$, and let ν be the posterior belief about receivers' aggregate play induced from n_1 periods of observing **Fight** after **Beer**. After n_1 periods, both types would share the belief ν. Then, at belief ν, type θ_{weak} must play **Beer** while type θ_{strong} plays **Quiche**, so signal **Beer** must have the highest Gittins index for θ_{weak} but not for θ_{strong}. But this would contradict Theorem 1.

The proof of Lemma 2 relies on the similar idea of fixing a particular "programming" of receiver play and studying the induced paths of experimentation for different types. In the aggregate learning model, the sequence of responses that a given sender encounters in her life depends on the realization of the random matching process, because different receivers have different histories and respond differently to a given signal. We can index all possible sequences of random matching realizations using a device we call the "pre-programmed response path." To show that more compatible types play a given signal more often, it suffices to show this comparison holds on each pre-programmed response path, thus coupling the learning processes of types θ' and θ''. We will show that the intuition above extends to signaling games with any number of signals and to any pre-programmed response path.

DEFINITION 5: A *pre-programmed response path* $\mathfrak{a} = (a_{1,s}, a_{2,s}, \dots,)_{s \in S}$ is an element in $\times_{s \in S}(A^\infty)$.

A pre-programmed response path is an $|S|$-tuple of infinite sequences of receiver actions, one sequence for each signal. For a given pre-programmed response path $\mathfrak{a}$, we can

[15] This follows from Bellman's (1956) Theorem 2 on Bernoulli bandits.

imagine starting with a new type θ and generating receiver play each period in the following programmatic manner: when the sender plays s for the jth time, respond with receiver action $a_{j,s}$. (If the sender sends s'' five times and then sends $s' \neq s''$, the response she gets to s' is $a_{1,s'}$, not $a_{6,s'}$.) For a type θ who applies σ_θ each period, $\mathfrak{a}$ induces a deterministic history of experiments and responses, which we denote $y_\theta(\mathfrak{a})$. The induced history $y_\theta(\mathfrak{a})$ can be used to calculate $\overline{\mathcal{R}}_1[\mathfrak{a}](\cdot|\theta)$, the distribution of signals over the lifetime of a type θ induced by the pre-programmed response path $\mathfrak{a}$. Namely, $\overline{\mathcal{R}}_1[\mathfrak{a}](\cdot|\theta)$ is simply a mixture over all signals sent along the history $y_\theta(\mathfrak{a})$, with weight $(1-\gamma)\gamma^{t-1}$ given to the signal in period t.

Now consider a type θ facing actions generated i.i.d. from the receiver behavior strategy π_2 each period, as in the interpretation of $\mathcal{R}_1$ in Remark 2. This data-generating process is equivalent to drawing a random pre-programmed response path $\mathfrak{a}$ at time 0 according to a suitable distribution, then producing all receiver actions using $\mathfrak{a}$. That is, $\mathcal{R}_1[\pi_2](\cdot|\theta) = \int \overline{\mathcal{R}}_1[\mathfrak{a}](\cdot|\theta)\, d\pi_2(\mathfrak{a})$, where we abuse notation and use $d\pi_2(\mathfrak{a})$ to denote the distribution over pre-programmed response paths associated with π_2. Importantly, any two types θ' and θ'' face the same distribution over pre-programmed response paths, so to prove the proposition, it suffices to show $\overline{\mathcal{R}}_1[\mathfrak{a}](s'|\theta') \geq \overline{\mathcal{R}}_1[\mathfrak{a}](s'|\theta'')$ for all $\mathfrak{a}$.

PROOF OF LEMMA 2: For $t \geq 0$, write y_θ^t for the truncation of infinite history y_θ to the first t periods, with $y_\theta^\infty := y_\theta$. Given a finite or infinite history y_θ^t for type θ, the signal counting function $\#(s|y_\theta^t)$ returns how many times signal s has appeared in y_θ^t. (We need this counting function since the receiver play generated by a pre-programmed response path each period depends on how many times each signal has been sent so far.)

As discussed above, we need only show $\overline{\mathcal{R}}_1[\mathfrak{a}](s'|\theta') \geq \overline{\mathcal{R}}_1[\mathfrak{a}](s'|\theta'')$. Let $\mathfrak{a}$ be given and write T_j^θ for the period in which type θ sends signal s' for the jth time in the induced history $y_\theta(\mathfrak{a})$. If no such period exists, then set $T_j^\theta = \infty$. Since $\overline{\mathcal{R}}_1[\mathfrak{a}](\cdot|\theta)$ is a weighted average over signals in $y_\theta(\mathfrak{a})$ with decreasing weights given to later signals, to prove $\overline{\mathcal{R}}_1[\mathfrak{a}](s'|\theta') \geq \overline{\mathcal{R}}_1[\mathfrak{a}](s'|\theta'')$ it suffices to show that $T_j^{\theta'} \leq T_j^{\theta''}$ for every j. Towards this goal, we will prove a sequence of statements by induction:

`Statement` j: Provided $T_j^{\theta''}$ is finite, $\#(s'' \mid y_{\theta'}^{T_j^{\theta'}}(\mathfrak{a})) \leq \#(s'' \mid y_{\theta''}^{T_j^{\theta''}}(\mathfrak{a}))$ for all $s'' \neq s'$.

For every j where $T_j^{\theta''} < \infty$, `Statement` j implies that the number of periods type θ' spent sending each signal $s'' \neq s'$ before sending s' for the jth time is fewer than the number of periods θ'' spent doing the same. Therefore, it follows that θ' sent s' for the jth time sooner than θ'' did, that is, $T_j^{\theta'} \leq T_j^{\theta''}$. Finally, if $T_j^{\theta''} = \infty$, then evidently $T_j^{\theta'} \leq \infty = T_j^{\theta''}$.

It now remains to prove the sequence of statements by induction.

`Statement 1` is the base case. By way of contradiction, suppose $T_1^{\theta''} < \infty$ and

$$\#\big(s'' \mid y_{\theta'}^{T_1^{\theta'}}(\mathfrak{a})\big) > \#\big(s'' \mid y_{\theta''}^{T_1^{\theta''}}(\mathfrak{a})\big)$$

for some $s'' \neq s'$. Then there is some earliest period $t^* < T_1^{\theta'}$ where

$$\#\big(s'' \mid y_{\theta'}^{t^*}(\mathfrak{a})\big) > \#\big(s'' \mid y_{\theta''}^{T_1^{\theta''}}(\mathfrak{a})\big),$$

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

But by construction, by the end of period t^*-1, type θ' has sent s'' exactly as many times as type θ'' has sent it by period $T_1^{\theta''}-1$, so that

$$\#\big(s'' \mid y_{\theta'}^{t^*-1}(\mathfrak{a})\big) = \#\big(s'' \mid y_{\theta''}^{T_1^{\theta''}-1}(\mathfrak{a})\big).$$

Furthermore, neither type has sent s' yet, so also

$$\#\big(s' \mid y_{\theta'}^{t^*-1}(\mathfrak{a})\big) = \#\big(s' \mid y_{\theta''}^{T_1^{\theta''}-1}(\mathfrak{a})\big).$$

Therefore, type θ' holds the same posterior over the receiver's reaction to signals s' and s'' at period t^*-1 as type θ'' does at period $T_1^{\theta''}-1$. So[16] by Theorem 1,

$$s' \in \arg\max_{\hat{s}\in S} I\big(\theta'', \hat{s}, y_{\theta''}^{T_1^{\theta''}-1}(\mathfrak{a})\big) \implies I\big(\theta', s', y_{\theta'}^{t^*-1}(\mathfrak{a})\big) > I\big(\theta', s'', y_{\theta'}^{t^*-1}(\mathfrak{a})\big). \tag{4}$$

However, by construction of $T_1^{\theta''}$, we have $\sigma_{\theta''}(y_{\theta''}^{T_1^{\theta''}-1}(\mathfrak{a})) = s'$. By the optimality of the Gittins index policy, the left-hand side of Equation (4) is satisfied. But, again by the optimality of the Gittins index policy, the right-hand side of Equation (4) contradicts $\sigma_{\theta'}(y_{\theta'}^{t^*-1}(\mathfrak{a})) = s''$. Therefore, we have proven `Statement 1`.

Now suppose `Statement` j holds for all $j \le K$. 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

$$\#\big(s'' \mid y_{\theta'}^{T_K^{\theta'}}(\mathfrak{a})\big) \le \#\big(s'' \mid y_{\theta''}^{T_K^{\theta''}}(\mathfrak{a})\big)$$

for every $s'' \neq s'$. Suppose there is some $s'' \neq s'$ such that

$$\#\big(s'' \mid y_{\theta'}^{T_{K+1}^{\theta'}}(\mathfrak{a})\big) > \#\big(s'' \mid y_{\theta''}^{T_{K+1}^{\theta''}}(\mathfrak{a})\big).$$

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

$$t^* := \min\big\{t : \#\big(s'' \mid y_{\theta'}^{t}(\mathfrak{a})\big) > \#\big(s'' \mid y_{\theta''}^{T_{K+1}^{\theta''}}(\mathfrak{a})\big)\big\},$$

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

$$\#\big(s'' \mid y_{\theta'}^{t^*-1}(\mathfrak{a})\big) = \#\big(s'' \mid y_{\theta''}^{T_{K+1}^{\theta''}-1}(\mathfrak{a})\big),$$

and also

$$\#\big(s' \mid y_{\theta'}^{t^*-1}(\mathfrak{a})\big) = K = \#\big(s' \mid y_{\theta''}^{T_{K+1}^{\theta''}-1}(\mathfrak{a})\big).$$

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

[16] In the following equation and elsewhere in the proof, we abuse notation and write $I(\theta, s, y)$ to mean $I(\theta, s, g_1(\cdot|y), \delta\gamma)$, which is the Gittins index of type θ for signal s at the posterior obtained from updating the prior g_1 using history y, with effective discount factor $\delta\gamma$.

4.3. *The Aggregate Receiver Response*

We now turn to the receivers' problem. Each new receiver thinks he is facing a fixed but unknown aggregate sender behavior strategy π_1, with belief over π_1 given by his regular prior g_2. To maximize his expected utility, the receiver must learn to infer the type of the sender from the signal, using his personal experience.

Unlike the senders whose optimal policies may involve experimentation, the receivers' problem only involves passive learning. Since the receiver observes the same information in a match regardless of his action, the optimal policy $\sigma_2(y_2)$ simply best responds to the posterior belief induced by history y_2.

DEFINITION 6: The *one-period-forward map for receivers* $f_2 : \Delta(Y_2) \times \Pi_1 \to \Delta(Y_2)$ is

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

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

As with the one-period-forward maps f_θ for senders, $f_2[\psi_2, \pi_1]$ describes the new distribution over receiver histories tomorrow if the distribution over histories in the receiver population today is ψ_2 and the sender population's aggregate play is π_1. We write $\psi_2^{\pi_1} := \lim_{T\to\infty} f_2^T(\psi_2, \pi_1)$ for the long-run distribution over Y_2 induced by fixing sender population's play at π_1, which is independent of the particular choice of initial state ψ_2.

DEFINITION 7: The *aggregate receiver response* $\mathcal{R}_2 : \Pi_1 \to \Pi_2$ is

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

where $\psi_2^{\pi_1} := \lim_{T\to\infty} f_2^T(\psi_2, \pi_1)$ with ψ_2 any arbitrary receiver state.

We are interested in the extent to which $\mathcal{R}_2[\pi_1]$ responds to inequalities of the form $\pi_1(s'|\theta') \geq \pi_1(s'|\theta'')$ embedded in π_1, such as those generated when $\theta' \succsim_{s'} \theta''$ (Lemma 2). To this end, for any two types θ', θ'', we define $P_{\theta' \triangleright \theta''}$ as those beliefs where the odds ratio of θ' to θ'' exceeds their prior odds ratio, that is,

$$P_{\theta' \triangleright \theta''} := \left\{ p \in \Delta(\Theta) : \frac{p(\theta'')}{p(\theta')} \leq \frac{\lambda(\theta'')}{\lambda(\theta')} \right\}. \tag{5}$$

If $\pi_1(s'|\theta') \geq \pi_1(s'|\theta'')$, $\pi_1(s'|\theta') > 0$, and receiver knows π_1, then receiver's posterior belief about sender's type after observing s' falls in the set $P_{\theta' \triangleright \theta''}$. The next lemma shows that under the additional provisions that $\pi_1(s'|\theta')$ is "large enough" and receivers are sufficiently long-lived, $\mathcal{R}_2[\pi_1]$ will best respond to $P_{\theta' \triangleright \theta''}$ with high probability when s' is sent.

For $P \subseteq \Delta(\Theta)$, we let[17] $\mathrm{BR}(P, s) := \bigcup_{p\in P}(\arg\max_{a'\in A} u_2(p, s, a'))$; this is the set of best responses to s supported by some belief in P.

LEMMA 3: *Let regular prior g_2, types θ', θ'', and signal s' be fixed. For every $\varepsilon > 0$, there exist $C > 0$ and $\underline{\gamma} < 1$ so that for any $0 \leq \delta < 1$, $\underline{\gamma} \leq \gamma < 1$, and $n \geq 1$, if $\pi_1(s'|\theta') \geq \pi_1(s'|\theta'')$ and $\pi_1(s'|\theta') \geq (1-\gamma)nC$, then*

$$\mathcal{R}_2[\pi_1]\big(\mathrm{BR}\big(P_{\theta' \triangleright \theta''}, s'\big)|s'\big) \geq 1 - \frac{1}{n} - \varepsilon.$$

[17] We abuse notation here and write $u_2(p, s, a')$ to mean $\sum_{\theta\in\Theta} u_2(\theta, s, a') \cdot p(\theta)$.

This lemma gives a lower bound on the probability that $\mathcal{R}_2[\pi_1]$ best responds to $P_{\theta' \triangleright \theta''}$ after signal s'. Note that the bound only applies for survival probabilities γ that are close enough to 1, because when receivers have short lifetimes, they need not get enough data to outweigh their prior. Note also that more of the receivers learn the compatibility condition when $\pi_1(s'|\theta')$ is large compared to $(1-\gamma)$ and almost all of them do in the limit of $n \to \infty$.The proof of Lemma 3 relies on Theorem 2 from Fudenberg, He, and Imhof (2017) about updating Bayesian posteriors after rare events, where the rare event corresponds to observing θ' play s'. The details are in Appendix A.3.

To interpret the condition $\pi_1(s'|\theta') \geq (1-\gamma)nC$, recall that an agent with survival chance γ has a typical lifespan of $\frac{1}{1-\gamma}$. If π_1 describes the aggregate play in the sender population, then on average a type θ' plays s' for $\frac{1}{1-\gamma} \cdot \pi_1(s'|\theta')$ periods in her life. So when a typical type θ' plays s' for nC periods, this lemma provides a bound of $1 - \frac{1}{n} - \varepsilon$ on the share of the receiver responses that lie in $\mathrm{BR}(P_{\theta' \triangleright \theta''}, s')$. Note that the hypothesis θ' plays s' for nC periods does not require that $\pi_1(s'|\theta')$ is bounded away from 0 as $\gamma \to 1$. To preview, Lemma 4 in the next section will establish that signals that are not weakly equilibrium dominated for a given type are played sufficiently often that Lemma 3 has bite when both δ and γ are close to 1.

5. STEADY-STATE IMPLICATIONS FOR AGGREGATE PLAY

Section 4 separately examined the senders' and receivers' learning problems. In this section, we turn to the two-sided learning problem. We will first define steady-state strategy profiles, which are signaling game strategy profiles π^* where π_1^* and π_2^* are mutual aggregate responses, and then characterize the steady states using our previous results.

5.1. *Steady States, δ-Stability, and Patient Stability*

We introduced the one-period-forward maps f_θ and f_2 in Section 4, which describe the deterministic transition between state ψ^t this period to state ψ^{t+1} next period through the learning dynamics and the birth–death process. More precisely, $\psi_\theta^{t+1} = f_\theta(\psi_\theta^t, \sigma_2(\psi_2^t))$ and $\psi_2^{t+1} = f_2(\psi_2^t, (\sigma_\theta(\psi_\theta^t))_{\theta \in \Theta})$. A steady state is a fixed point ψ^* of this transition map.

DEFINITION 8: A state ψ^* is a *steady state* if $\psi_\theta^* = f_\theta(\psi_\theta^*, \sigma_2(\psi_2^*))$ for every θ and $\psi_2^* = f_2(\psi_2^*, (\sigma_\theta(\psi_\theta^*))_{\theta \in \Theta})$. The set of all steady states for regular prior g and $0 \leq \delta, \gamma < 1$ is denoted $\Psi^*(g, \delta, \gamma)$, while the set of steady-state strategy profiles is $\Pi^*(g, \delta, \gamma) := \{\sigma(\psi^*) : \psi^* \in \Psi^*(g, \delta, \gamma)\}$.

The strategy profiles associated with steady states represent time-invariant distributions of play, as the information lost when agents die each period exactly balances out the information agents gain through learning that period. This means the exchangeability assumption of the learners will be satisfied in any steady state.

We now give an equivalent characterization $\Pi^*(g, \delta, \gamma)$ in terms of $\mathcal{R}_1$ and $\mathcal{R}_2$. The proof is in Appendix A.4.

PROPOSITION 2: *$\pi^* \in \Pi^*(g, \delta, \gamma)$ if and only if $\mathcal{R}_1^{g,\delta,\gamma}(\pi_2^*) = \pi_1^*$ and $\mathcal{R}_2^{g,\delta,\gamma}(\pi_1^*) = \pi_2^*$.*

(Note that here we make the dependence of $\mathcal{R}_1$ and $\mathcal{R}_2$ on parameters (g, δ, γ) explicit to avoid confusion.) That is, a steady-state strategy profile is a pair of mutual aggregate replies.

The next proposition guarantees that there always exists at least one steady-state strategy profile.

PROPOSITION 3: *$\Pi^*(g, \delta, \gamma)$ is nonempty and compact in the norm topology.*

The proof is in the Supplemental Material (Fudenberg and He (2018)). We establish that $\Psi^*(g, \delta, \gamma)$ is nonempty and compact in the ℓ_1 norm on the space of distributions, which immediately implies the same properties for $\Pi^*(g, \delta, \gamma)$. Intuitively, if lifetimes are finite, the set of histories is finite, so the set of states is of finite dimension. Here the one-period-forward map $f = ((f_\theta)_{\theta\in\Theta}, f_2)$ is continuous, so the usual version of Brouwer's fixed-point theorem applies. With geometric lifetimes, very old agents are rare, so truncating the agents' 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 that (because of the geometric lifetimes) the feasible states form a compact locally convex Hausdorff space. This lets us appeal to a fixed-point theorem for that domain.

We now focus on the iterated limit

$$\lim_{\delta\to 1}\lim_{\gamma\to 1}\Pi^*(g, \delta, \gamma),$$

that is, the set of steady-state strategy profiles for δ and γ near 1, where we first send γ to 1 holding δ fixed, and then send δ to 1.

DEFINITION 9: For each $0 \le \delta < 1$, 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 $\pi^{(k)} \in \Pi^*(g, \delta, \gamma_k)$, 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 under g and $\pi^{(k)} \to \pi^*$. Strategy profile π^* is *patiently stable* if it is patiently stable under some regular prior g.

Heuristically, patiently stable strategy profiles are the limits of learning outcomes when agents become infinitely patient (so that senders are willing to make many experiments) and long lived (so that agents on both sides can learn enough for their data to outweigh their prior). As in past work on steady-state learning (Fudenberg and Levine (1993, 2006)), the reason for this order of limits is to ensure that most agents have enough data that they stop experimenting and play myopic best responses.[18] We do not know whether our results extend to the other order of limits; we explain the issues involved below, after sketching the intuition for Proposition 5.

5.2. *Preliminary Results on δ-Stability and Patient Stability*

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's (1993) result that δ-stable strategy profiles are self-confirming equilibria.

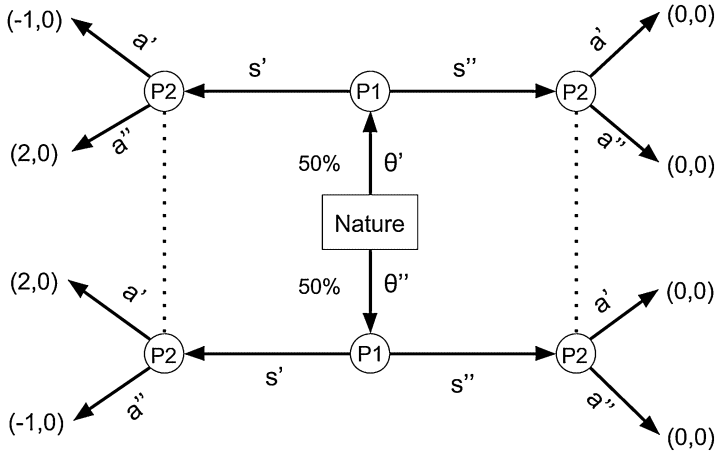
[18] If agents did not eventually stop experimenting as they age, then even if most agents have approximately correct beliefs, aggregate play need not be close to a Nash equilibrium because most agents would not be playing a (static) best response to their beliefs.

PROPOSITION 4: *Suppose strategy profile π^* is δ-stable under a regular prior. Then, for every type θ and signal s with $\pi_1^*(s|\theta) > 0$, s is a best response to some $\pi_2 \in \Pi_2$ for type θ, and furthermore, $\pi_2(\cdot|s) = \pi_2^*(\cdot|s)$. Also, for any signal s such that $\pi_1^*(s|\theta) > 0$ for at least one type θ, $\pi_2^*(\cdot|s)$ is supported on pure best responses to the Bayesian belief generated by π_1^* after s.*

We prove this result in the Supplemental Material (Fudenberg and He (2018)). The idea of the proof is the following: If signal s has positive probability in the limit, then it is played many times by the senders, so the receivers eventually learn the correct posterior distribution for θ given s. As the receivers have no incentive to experiment, their actions after s will be a best response to this correct posterior belief. For the senders, suppose $\pi_1^*(s|\theta) > 0$, but s is not a best response for type θ to any $\pi_2 \in \Pi_2$ that matches $\pi_2^*(\cdot|s)$. Yet if a sender has played s many times then with high probability her belief about $\pi_2(\cdot|s)$ is close to $\pi_2^*(\cdot|s)$, so playing s is not myopically optimal. This would imply that type θ has persistent option value for signal s, which contradicts the fact that this option value must converge to 0 with the sample size.

REMARK 3: This proposition says that each sender type is playing a best response to a belief about the receiver's play that is correct on the equilibrium path, and that the receivers are playing an aggregate best response to the aggregate play of the senders. Thus the δ-stable outcomes are a version of self-confirming equilibrium where different types of sender are allowed to have different beliefs. Moreover, as the next example shows, this sort of heterogeneity in the senders' beliefs about the aggregate strategy of the receivers can endogenously arise in a δ-stable strategy profile even when all types of new senders start with the same prior over how the receivers play.[19]

EXAMPLE 2: Consider the following game:



The receiver is indifferent between all responses. Fix any regular prior g_2 for the receiver and any regular prior $g_1^{(s'')}$ for the sender. Let $g_1^{(s')}$ be Beta$(1, 3)$ on a' and a'',

[19] Dekel, Fudenberg, and Levine (2004) defined type-heterogeneous self-confirming equilibrium in static Bayesian games. As they noted, this sort of heterogeneity is natural when the type of each agent is fixed, but not if each agent's type is drawn i.i.d. in each period. To extend their definition to signaling 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 (2018).

respectively. We claim that it is δ-stable when $\delta = 0$ for both types to send s'' and for the receiver to respond to every signal with a', which is a type-heterogeneous rationalizable self-confirming equilibrium. However, this pooling behavior cannot occur in a Nash equilibrium or in a unitary self-confirming equilibrium, where both sender types must hold the same belief about how the receiver responds to s'.

To establish this claim, note that since $\delta = 0$, each sender plays the myopically optimal signal after every history. For any γ, there is a steady state where the receivers' policy responds to every signal with a' after every history, type θ'' senders play s'' after every history and never update their prior belief about how receivers react to s', and type θ' senders with fewer than 6 periods of experience play s' but switch to playing s'' forever starting at age 7. The behavior of the θ' agents is optimal because after k periods of playing s' and seeing response a' every period, the sender's posterior belief about $\pi_2(\cdot|s')$ is Beta$(1+k, 3)$, so the expected payoff from playing s' 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' aged 6 and below approaches 0 as $\gamma \to 1$, hence we have constructed a sequence of steady-state strategy profiles converging to the s'' pooling equilibrium. So even though both types start with the same prior g_1, their beliefs about how the receivers react to s' eventually diverge.

In contrast to the plethora of δ-stable profiles, we now show that only Nash equilibrium profiles can be steady-state outcomes as δ tends to 1. Moreover, this limit also rules out strategy profiles in which the sender's strategy can only be supported by the belief that the receiver would play a dominated action in response to some of the unsent signals.

DEFINITION 10: In a signaling game, a *perfect Bayesian equilibrium with heterogeneous off-path beliefs* is a strategy profile (π_1^*, π_2^*) such that:

- For each $\theta \in \Theta$, $u_1(\theta; \pi^*) = \max_{s \in S} u_1(\theta, s, \pi_2^*(\cdot|s))$.
- For each on-path signal s, $u_2(p^*(\cdot|s), s, \pi_2^*(\cdot|s)) = \max_{\hat{a} \in A} u_2(p^*(\cdot|s), s, \hat{a})$.
- For each off-path signal s and each $a \in A$ with $\pi_2^*(a|s) > 0$, there exists a belief $p \in \Delta(\Theta)$ such that $u_2(p, s, a) = \max_{\hat{a} \in A} u_2(p, s, \hat{a})$.

Here $u_1(\theta; \pi^*)$ refers to type θ's payoff under π^*, and $p^*(\cdot|s)$ is the Bayesian posterior belief about sender's type after signal s, under strategy π_1^*.

The first two conditions imply that the profile is a Nash equilibrium. The third condition resembles that of perfect Bayesian equilibrium, but is somewhat weaker as it allows the receiver's play after an off-path signal s to be a mixture over several actions, each of which is a best response to a different belief about the sender's type. This means $\pi_2^*(\cdot|s) \in \Delta(\text{BR}(\Delta(\Theta), s))$, but $\pi_2^*(\cdot|s)$ itself may not be a best response to any unitary belief about the sender's type.

PROPOSITION 5: *If strategy profile π^* is patiently stable, then it is a perfect Bayesian equilibrium with heterogeneous off-path beliefs.*

PROOF: In the Supplemental Material (Fudenberg and He (2018)), we prove that patiently stable profiles must be Nash equilibria. This argument follows the proof strategy of Fudenberg and Levine (1993), which derived a contradiction via excess option values.

In outline, if π^* is patiently stable, each player's strategy is a best response to a belief that is correct about the opponent's on-path play. Thus, if π^* is not a Nash equilibrium, some type should perceive a persistent option value to experimenting with some signal that she plays with probability 0. But this would contradict the fact that the option values evaluated at sufficiently long histories must go to 0. We now explain why a patiently stable profile π^* must satisfy the third condition in Definition 10. After observing any history y_2, a receiver who started with a regular prior thinks every signal has positive probability in his next match. So, his optimal policy prescribes for each signal s a best response to that receiver's posterior belief about the sender's type upon seeing signal s after history y_2. For any regular prior g, $0 \leq \delta, \gamma < 1$, and any sender aggregate play π_1, we thus deduce $\mathcal{R}_2^{g,\delta,\gamma}[\pi_1](\cdot|s)$ is entirely supported on $\mathrm{BR}(\Delta(\Theta), s)$. This means the same is true about the aggregate receiver response in every steady state and hence in every patiently stable strategy profile. *Q.E.D.*

In Fudenberg and Levine (1993), this argument relies on the finite lifetime of the agents only to ensure that "almost all" histories are long enough, by picking a large enough lifetime. We can achieve the analogous effect in our geometric-lifetime model by picking γ close to 1. Our proof uses the fact that if δ is fixed and $\gamma \to 1$, then the number of experiments that a sender needs to exhaust her option value is negligible relative to her expected lifespan, so that most senders play approximate best responses to their current beliefs. The same conclusion does not hold if we fix γ and let $\delta \to 1$, even though the optimal sender policy only depends on the product $\delta\gamma$, because for a fixed sender policy the induced distribution on sender play depends on γ but not on δ.

5.3. *Patient Stability Implies the Compatibility Criterion*

Proposition 5 allows the receiver to sustain his off-path actions using any belief $p \in \Delta(\Theta)$. We now turn to our main result, which focuses on refining off-path beliefs. We prove that patient stability selects a strict subset of the Nash equilibria, namely, those that satisfy the *compatibility criterion*.

DEFINITION 11: For a fixed strategy profile π^*, let $u_1(\theta; \pi^*)$ denote the payoff to type θ under π^*, and let

$$J(s, \pi^*) := \left\{\theta \in \Theta : \max_{a \in A} u_1(\theta, s, a) > u_1(\theta; \pi^*)\right\}$$

be the set of types for which *some* response to signal s is strictly better than their payoff under π^*. Signal s is *weakly equilibrium dominated* for types in the complement of $J(s, \pi^*)$.

The *admissible beliefs at signal s under profile π^** are

$$P(s, \pi^*) := \bigcap \left\{P_{\theta' \triangleright \theta''} : \theta' \succsim_s \theta'' \text{ and } \theta' \in J(s, \pi^*)\right\},$$

where $P_{\theta' \triangleright \theta''}$ is defined in Equation (5).

That is, $P(s, \pi^*)$ is the joint belief restriction imposed by a family of $P_{\theta' \triangleright \theta''}$ for (θ', θ'') satisfying two conditions: θ' is more type-compatible with s than θ'', and furthermore, the more compatible type θ' belongs to $J(s, \pi^*)$. If there are no pairs (θ', θ'') satisfying these two conditions, then (by convention of intersection over no elements) $P(s, \pi^*)$ is defined as $\Delta(\Theta)$. In any signaling game and for any π^*, the set $P(s, \pi^*)$ is always nonempty because it always contains the prior λ.

DEFINITION 12: Strategy profile π^* *satisfies the compatibility criterion* if $\pi_2(\cdot|s) \in \Delta(\text{BR}(P(s, \pi^*), s))$ for every s.

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

One might imagine a version of the compatibility criterion where the belief restriction $P_{\theta' \rhd \theta''}$ applies whenever $\theta' \succsim_s \theta''$. To understand why we require the additional condition that $\theta' \in J(s, \pi^*)$ in the definition of admissible beliefs, recall that Lemma 3 only gives a learning guarantee in the receiver's problem when $\pi_1(s|\theta')$ is "large enough" for the more type-compatible θ'. In the extreme case where s is a strictly dominated signal for θ', she will never play it during learning. It turns out that if s is weakly equilibrium dominated for θ', then θ' may still not experiment very much with it. On the other hand, the next lemma provides a lower bound on the frequency that θ' experiments with s' when $\theta' \in J(s', \pi^*)$ and δ and γ are close to 1.

LEMMA 4: *Fix a regular prior g and a strategy profile π^* where, for some type θ' and signal s', $\theta' \in J(s', \pi^*)$. There exist a number $\varepsilon \in (0, 1)$ and threshold functions $\bar{\delta} : \mathbb{N} \to (0, 1)$ and $\bar{\gamma} : \mathbb{N} \times (0, 1) \to (0, 1)$ such that whenever $\pi \in \Pi^*(g, \delta, \gamma)$ with $\delta \geq \bar{\delta}(N)$ and $\gamma \geq \bar{\gamma}(N, \delta)$ and π is no more than ε away from π^* in L_1 distance,*[20] *we have $\pi_1(s'|\theta') \geq (1 - \gamma) \cdot N$.*

Note that since $\pi_1(s|\theta')$ is between 0 and 1, we know that $(1 - \bar{\gamma}(N, \delta)) \cdot N < 1$ for each N.

The proof of this lemma is in the Supplemental Material (Fudenberg and He (2018)). To gain an intuition for it, suppose that not only is s' equilibrium undominated in π^*, but furthermore, s' can lead to the highest signaling game payoff for type θ' under some receiver response a'. Because the prior is non-doctrinaire, the Gittins index of each signal in the learning problem approaches its highest possible payoff in the stage game as the sender becomes infinitely patient. Therefore, for every $N \in \mathbb{N}$, when γ and δ are close enough to 1, a new type θ' will play s' 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, proving the lemma in this special case. It turns out that even if s' does not lead to the highest potential payoff in the signaling game, long-lived players will have a good estimate of their steady-state payoff. So, type θ' will still play any s' 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.

THEOREM 2: *Every patiently stable strategy profile π^* satisfies the compatibility criterion.*

The proof combines Lemma 2, Lemma 3, and Lemma 4. Lemma 2 shows that types that are more compatible with s' play it more often. Lemma 4 says that types for whom s' is not weakly equilibrium dominated will play it "many times." Finally, Lemma 3 shows

[20] The L_1 distance is

$$d(\pi, \pi^*) = \sum_{\theta \in \Theta} \sum_{s \in S} \left| \pi_1(s|\theta) - \pi_1^*(s|\theta) \right| + \sum_{s \in S} \sum_{a \in A} \left| \pi_2(a|s) - \pi_2^*(a|s) \right|.$$

that the "many times" here is sufficiently large that most receivers correctly believe that more compatible types play s' more than less compatible types do, so their posterior odds ratio for more versus less compatible types exceeds the prior ratio.

PROOF OF THEOREM 2: Suppose π^* is patiently stable under regular prior g. Fix an s' and an action $\hat{a} \notin \mathrm{BR}(P(s', \pi^*), s')$. Let $h > 0$ be given. We will show that $\pi_2^*(\hat{a}|s') < h$. Since the choices of s', $\hat{a}$, and $h > 0$ are arbitrary, we will have proven the theorem.

Step 1: Setting some constants.

In the statement of Lemma 3, for each pair θ', θ'' such that $\theta' \succsim_{s'} \theta''$ and $\theta' \in J(s', \pi^*)$, put $\varepsilon = \frac{h}{2|\Theta|^2}$ and find $C_{\theta',\theta''}$ and $\underline{\gamma}_{\theta',\theta''}$ so that the result holds. Let C be the maximum of all such $C_{\theta',\theta''}$ and $\underline{\gamma}$ be the maximum of all such $\underline{\gamma}_{\theta',\theta''}$. Also find $\underline{n} \geq 1$ so that

$$1 - \frac{1}{\underline{n}} > 1 - \frac{h}{2|\Theta|^2}. \tag{6}$$

In the statement of Lemma 4, for each θ' such that $\theta' \succsim_{s'} \theta''$ for at least one θ'', find $\varepsilon_{\theta'}, \bar{\delta}_{\theta'}(\underline{n}C), \bar{\gamma}_{\theta'}(\underline{n}C, \delta)$ so that the lemma holds. Write $\varepsilon^* > 0$ as the minimum of all such $\varepsilon_{\theta'}$ and let $\bar{\delta}^*(\underline{n}C)$ and $\bar{\gamma}^*(\underline{n}C, \delta)$ represent the maximum of $\delta_{\theta'}$ and $\gamma_{\theta'}$ across such θ'.

Step 2: Finding a steady-state profile with large δ, γ that approximates π^*.

Since π^* is patiently stable under g, there exists a sequence of strategy profiles $\pi^{(j)} \to \pi^*$ where $\pi^{(j)}$ is δ_j-stable under g with $\delta_j \to 1$. Each $\pi^{(j)}$ can be written as the limit of steady-state strategy profiles. That is, for each j, there exist $\gamma_{j,k} \to 1$ and a sequence of steady-state profiles $\pi^{(j,k)} \in \Pi^*(g, \delta_j, \gamma_{j,k})$ such that $\lim_{k\to\infty} \pi^{(j,k)} = \pi^{(j)}$.

The convergence of the array $\pi^{(j,k)}$ to π^* means we may find $\underline{j} \in \mathbb{N}$ and function $k(j)$ so that whenever $j \geq \underline{j}$ and $k \geq k(j)$, $\pi^{(j,k)}$ is no more than $\min(\varepsilon^*, \frac{h}{2|\Theta|^2})$ away from π^*. Find $j^\circ \geq \underline{j}$ large enough so $\delta^\circ := \delta_{j^\circ} > \bar{\delta}^*(\underline{n}C)$, and then find a large enough $k^\circ > k(j^\circ)$ so that $\gamma^\circ := \gamma_{j^\circ,k^\circ} > \max(\bar{\gamma}^*(\underline{n}C, \delta^\circ), \underline{\gamma})$. So we have identified a steady-state profile $\pi^\circ := \pi^{(j^\circ,k^\circ)} \in \Pi^*(g, \delta^\circ, \gamma^\circ)$ which approximates π^* to within $\min(\varepsilon^*, \frac{h}{2|\Theta|^2})$.

Step 3: Applying properties of $\mathcal{R}_1$ and $\mathcal{R}_2$.

For each pair θ', θ'' such that $\theta' \succsim_{s'} \theta''$ and $\theta' \in J(s', \pi^*)$, we will bound the probability that $\pi_2^\circ(\cdot|s')$ does not best respond to $P_{\theta' \triangleright \theta''}$ by $\frac{h}{|\Theta|^2}$. Since there are at most $|\Theta| \cdot (|\Theta| - 1)$ such pairs in the intersection defining $P(s', \pi^*)$, this would imply that $\pi_2^\circ(\hat{a}|s') < [|\Theta| \cdot (|\Theta| - 1)] \cdot \frac{h}{|\Theta|^2}$ since $\hat{a} \notin \mathrm{BR}(P(s', \pi^*), s')$. And since π_2° is no more than $\frac{h}{2|\Theta|^2}$ away from π_2, this would show $\pi_2(\hat{a}|s') < h$.

By construction, π° is closer than $\varepsilon_{\theta'}$ to π^*, and furthermore, $\delta^\circ \geq \bar{\delta}_{\theta'}(\underline{n}C)$ and $\gamma^\circ \geq \bar{\gamma}_{\theta'}(\underline{n}C, \delta^\circ)$. By Lemma 4, $\pi_1^\circ(s'|\theta') \geq \underline{n}C(1 - \gamma^\circ)$. At the same time, $\pi_1^\circ = \mathcal{R}_1[\pi_2^\circ]$ and $\theta' \succsim_{s'} \theta''$, so Lemma 2 implies that $\pi_1^\circ(s'|\theta') \geq \pi_1^\circ(s'|\theta'')$. Turning to the receiver side, $\pi_2^\circ = \mathcal{R}_2[\pi_1^\circ]$ with $\pi_{\circ 1}$ satisfying the conditions of Lemma 3 associated with $\varepsilon = \frac{h}{2|\Theta|^2}$ and $\gamma^\circ \geq \underline{\gamma}$. Therefore, we conclude

$$\pi_2^\circ\big(\mathrm{BR}\big(P_{\theta' \triangleright \theta''}, s'\big)|s'\big) \geq 1 - \frac{1}{\underline{n}} - \frac{h}{2|\Theta|^2}.$$

But by construction of $\underline{n}$ in Equation (6), $1 - \frac{1}{\underline{n}} > 1 - \frac{h}{2|\Theta|^2}$. So the LHS is at least $1 - \frac{h}{|\Theta|^2}$, as desired. *Q.E.D.*

REMARK 4: More generally, consider *any* model for our populations of agents with geometrically distributed lifetimes that generates aggregate response functions $\mathcal{R}_1$ and $\mathcal{R}_2$.

Defining the steady states under (g, δ, γ) as the strategy profiles π^* such that $\mathcal{R}_1^{g,\delta,\gamma}(\pi_2^*) = \pi_1^*$ and $\mathcal{R}_2^{g,\delta,\gamma}(\pi_1^*) = \pi_2^*$, the proof of Theorem 2 applies to the patiently stable profiles of the new learning model provided that $\mathcal{R}_1$ satisfies the conclusion of Lemma 2, $\mathcal{R}_2$ satisfies the conclusion of Lemma 3, and Lemma 4 is valid for (θ', s') pairs such that $\theta' \succsim_{s'} \theta''$ for at least one type $\theta''$and $\theta' \in J(s', \pi^*)$.

We outline two such more general learning models below. (The proof is in the Supplemental Material (Fudenberg and He (2018)).)

COROLLARY 1: *With either of the following modifications of the steady-state learning model from Section 2, every patiently stable strategy profile still satisfies the compatibility criterion.*

(i) Heterogeneous priors. *There is a finite collection of regular sender priors $\{g_{1,k}\}_{k=1}^n$ and a finite collection of regular receiver priors $\{g_{2,k}\}_{k=1}^n$. Upon birth, an agent is endowed with a random prior, where the distributions over priors are μ_1 and μ_2 for senders and receivers. An agent's prior is independent of her payoff type, and furthermore, no one ever observes another person's prior.*

(ii) Social learning. *Suppose $1 - \alpha$ fraction of the senders are "normal learners" as described in Section 2, but the remaining $0 < \alpha < 1$ fraction are "social learners." At the end of each period, a social learner can observe the extensive-form strategies of her matched receiver and of $c > 0$ other matches sampled uniformly at random. Each sender knows whether she is a normal learner or a social learner upon birth, which is uncorrelated with her payoff type. Receivers cannot distinguish between the two kinds of senders.*

EXAMPLE 1—Continued: The beer-quiche game of Example 1 has two components of Nash equilibria: "beer-pooling equilibria" where both types play **Beer** with probability 1, and "quiche-pooling equilibria" where both types play **Quiche** with probability 1. In a quiche-pooling equilibrium π^*, type θ_{strong}'s equilibrium payoff is 2, so $\theta_{\text{strong}} \in J(\textbf{Beer}, \pi^*)$ since θ_{strong}'s highest possible payoff under **Beer** is 3, and we have already shown that $\theta_{\text{strong}} \succsim_{\textbf{Beer}} \theta_{\text{weak}}$. So,

$$P(\textbf{Beer}, \pi^*) = \left\{ p \in \Delta(\Theta) : \frac{p(\theta_{\text{weak}})}{p(\theta_{\text{strong}})} \leq \frac{\lambda(\theta_{\text{weak}})}{\lambda(\theta_{\text{strong}})} = 1/9 \right\}.$$

Fight is not a best response after **Beer** to any such belief, so equilibria in which **Fight** occurs with positive probability after **Beer** do not satisfy the compatibility criterion, and thus no quiche-pooling equilibrium is patiently stable. Since the set of patiently stable outcomes is a nonempty subset of the set of Nash equilibria, pooling on beer is the unique patiently stable outcome.

By Corollary 1, quiche-pooling equilibria are still not patiently stable in more general learning models involving either heterogeneous priors or social learners.

5.4. *Patient Stability and Equilibrium Dominance*

In generic signaling games, equilibria where the receiver plays a pure strategy must satisfy a stronger condition than the compatibility criterion to be patiently stable.

DEFINITION 13: Let

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

If $\widetilde{J}(s', \pi^*)$ is nonempty, define the *strongly admissible beliefs at signal s' under profile π^** to be

$$\tilde{P}(s', \pi^*) := \Delta\big(\widetilde{J}(s', \pi^*)\big) \cap \big\{P_{\theta' \triangleright \theta''} : \theta' \succsim_{s'} \theta''\big\},$$

where $P_{\theta' \triangleright \theta''}$ is defined in Equation (5). Otherwise, define $\tilde{P}(s', \pi^*) := \Delta(\Theta)$.

Here, $\widetilde{J}(s, \pi^*)$ is the set of types for which *some* response to signal s is at least as good as their equilibrium payoff under π^*—that is, the set of types for whom s is not equilibrium dominated in the sense of Cho and Kreps (1987). Note that $\widetilde{P}$, unlike P, assigns probability 0 to equilibrium-dominated types, which is the belief restriction of the Intuitive Criterion.

DEFINITION 14: A Nash equilibrium π^* is *on-path strict for the receiver* if, for every on-path signal s^*, $\pi_2(a^*|s^*) = 1$ for some $a^* \in A$ and $u_2(s^*, a^*, \pi_1) > \max_{a \neq a^*} u_2(s^*, a, \pi_1)$.

Of course, the receiver cannot have strict ex ante preferences over play at unreached information sets; this condition is called "on-path strict" because it places no restrictions on the receiver's incentives after off-path signals. In generic signaling games, all pure-strategy equilibria are on-path strict for the receiver, but the same is not true for mixed-strategy equilibria.

DEFINITION 15: A strategy profile π^* satisfies the *strong compatibility criterion* if, at every signal s', we have

$$\pi_2^*(\cdot|s') \in \Delta\big(\mathrm{BR}\big(\widetilde{P}(s', \pi^*), s'\big)\big).$$

It is immediate that the strong compatibility criterion implies the compatibility criterion, since it places more stringent restrictions on the receiver's behavior. It is also immediate that the strong compatibility criterion implies the Intuitive Criterion.

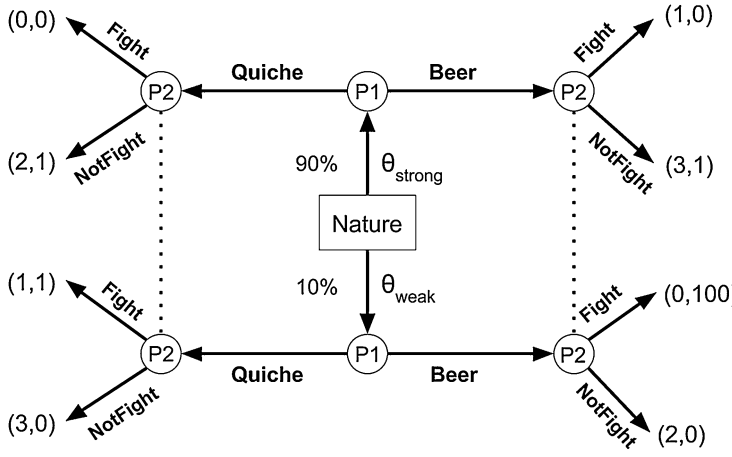
THEOREM 3: *Suppose π^* is on-path strict for the receiver and patiently stable. Then it satisfies the strong compatibility criterion.*

The proof of this theorem appears in Appendix A.5. The main idea is that when off-path signal s' is equilibrium dominated in π^* for type θ^{D} but not even weakly equilibrium dominated for type θ^{U}, type θ^{U} will experiment "infinitely more often" with s' than θ^{D} does. Indeed, we can provide an upper bound on the steady-state probability that θ^{D} ever switches away from its equilibrium signal s^* after trying it for the first time,[21] which is also an upper bound on how often θ^{D} experiments with s', while Lemma 4 provides a lower bound for how often θ^{U} plays s'. We show there is a sequence of steady-state profiles $\pi^{(k)} \in \Pi^*(g, \delta_k, \gamma_k)$ with $\gamma_k \to 1$ and $\pi^{(k)} \to \pi^*$ where the ratio of the lower bound to the upper bound goes to infinity. Applying Theorem 2 of Fudenberg, He, and Imhof (2017), we can then prove receivers will infer that an s'-sender is "infinitely more likely" to be θ^{U} than θ^{D}, which means receivers must assign probability 0 to θ^{D} after s' in equilibrium π^*.

[21] This upper bound does not apply when π^* is not on-path strict for the receiver. When π^* involves the receiver strictly mixing between several responses after s^*, some of these responses might make θ^{D} strictly worse off than her worst payoff after s', so there is non-vanishing probability that θ^{D} observes a large number of these bad responses in a row and then stops playing s^*.

REMARK 5: 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 signal 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 signals. When the equilibrium involves on-path receiver randomization, a nontrivial fraction of receivers could play an action after a type's equilibrium signal that the type finds strictly worse than her worst payoff under an off-path signal. In this case, we do not see how to show that the probability she ever switches away from her equilibrium signal 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 3: In the following modified beer-quiche game, the payoffs of fighting a type θ_{weak} who drinks beer have been substantially increased relative to Example 1, so that **Fight** is now a best response to the prior belief λ after **Beer**.



Since the prior λ is always an admissible belief in any signaling game after any signal, the Nash equilibrium π^* where both types play **Quiche** (supported by the receiver playing **Fight** after **Beer**) is not ruled out by the compatibility criterion, unlike in Example 1. However, this equilibrium is ruled out by the strong compatibility criterion. To see why, note that this pooling equilibrium is on-path strict for the receiver, because the receiver has a strict preference for **NotFight** at the only on-path signal, **Quiche**. Moreover, π^* does not satisfy the strong compatibility criterion, because $\widetilde{J}(\textbf{Beer}, \pi^*) = \{\theta_{\text{strong}}\}$ implies the only strongly admissible belief after **Beer** assigns probability 1 to the sender being θ_{strong}. Thus, Theorem 3 implies that this equilibrium is not patiently stable.

6. DISCUSSION

Our learning model supposes that the agents have geometrically distributed lifetimes, which is one of the reasons that the senders' optimization problems can be solved using

the Gittins index. If agents were to have fixed finite lifetimes, as in Fudenberg and Levine (1993, 2006), their optimization problem would not be stationary, and the finite-horizon analog of the Gittins index is only approximately optimal for the finite-horizon multi-armed bandit problem (Niño-Mora (2011)). Applying the geometric-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 behavior of various players or player types, and in studies of other sorts of dynamic decisions.

Theorem 1 provides a comparison between the dynamic behavior of two agents in a geometric-lifetime bandit problem based on their static preferences over the prizes. As an immediate application, consider a principal-agent setting where the agent faces a multi-armed bandit with arms $s \in S$, where s leads a prize drawn from Z_s according to some distribution. The principal knows the agent's per-period utility function $u : \bigcup_s Z_s \to \mathbb{R}$, but not the agent's beliefs over the prize distributions of different arms or agent's discount factor. Suppose the principal observes the agent choosing arm 1 in the first period. The principal can impose taxes and subsidies on the different prizes and arms, changing the agent's utility function to $\tilde{u}$. For what taxes and subsidies would the agent still have chosen arm 1 in the first period, irrespective of her initial beliefs and discount factor? According to Theorem 1, the answer is precisely those taxes and subsidies such that arm 1 is more type-compatible with $\tilde{u}$ than u.

Our results provide an upper bound on the set of patiently stable strategy profiles in a signaling game. In Fudenberg and He (2017), we provided a lower bound for the same set, as well as a sharper upper bound under additional restrictions on the priors. But, together, these results will not give an exact characterization of patiently stable outcomes. Nevertheless, our results do show how the theory of learning in games provides a foundation for refining the set of equilibria in signaling games.

In future work, we hope to investigate a learning model featuring 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 becomes a restless bandit.

APPENDIX: RELEGATED PROOFS

A.1. *Proof of Proposition 1*

PROPOSITION 1:

(i) $\succsim_{s'}$ *is transitive*.

(ii) *Except when s' is either strictly dominant for both θ' and θ'' or strictly dominated for both θ' and θ'', $\theta' \succsim_{s'} \theta''$ implies $\theta'' \not\succsim_{s'} \theta'$.*

PROOF: To show (i), suppose $\theta' \succsim_{s'} \theta''$ and $\theta'' \succsim_{s'} \theta'''$. For any $\pi_2 \in \Pi_2$ where s' is weakly optimal for θ''', it must be strictly optimal for θ'', hence also strictly optimal for θ'. This shows $\theta' \succsim_{s'} \theta'''$.

To establish (ii), partition the set of receiver strategies as $\Pi_2 = \Pi_2^+ \cup \Pi_2^0 \cup \Pi_2^-$, where the three subsets refer to receiver strategies that make s' strictly better, indifferent, or strictly worse than the best alternative signal for θ''. If the set Π_2^0 is nonempty, then $\theta' \succsim_{s'} \theta''$ implies $\theta'' \not\succsim_{s'} \theta'$. This is because against any $\pi_2 \in \Pi_2^0$, signal s' is strictly optimal for θ' but only weakly optimal for θ''. At the same time, if both Π_2^+ and Π_2^- are nonempty, then Π_2^0 is nonempty. This is because both $\pi_2 \mapsto u_1(\theta'', s', \pi_2(\cdot|s'))$ and $\pi_2 \mapsto$

$\max_{s''\neq s'} u_1(\theta'', s'', \pi_2(\cdot|s''))$ are continuous functions, so for any $\pi_2^+ \in \Pi_2^+$ and $\pi_2^- \in \Pi_2^-$, there exists $\alpha \in (0,1)$ so that $\alpha\pi_2^+ + (1-\alpha)\pi_2^- \in \Pi_2^0$. If only Π_2^+ is nonempty and $\theta' \succsim_{s'} \theta''$, then s' is strictly dominant for both θ' and θ''. If only Π_2^- is nonempty, then we can have $\theta'' \succsim_{s'} \theta'$ only when s' is never a weak best response for θ' against any $\pi_2 \in \Pi_2$. *Q.E.D.*

A.2. *Proof of Lemma 1*

LEMMA 1: *For every signal s, stopping time τ, belief ν_s, and discount factor β, there exists $\pi_{2,s}(\tau, \nu_s, \beta) \in \Delta(A)$ so that for every θ,*

$$\frac{\mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau-1} \beta^t \cdot u_1\big(\theta, s, a_s(t)\big)\right\}}{\mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau-1} \beta^t\right\}} = u_1\big(\theta, s, \pi_{2,s}(\tau, \nu_s, \beta)\big).$$

PROOF: *Step 1: Induced mixed actions.*

A belief ν_s and a stopping time τ_s together define a stochastic process $(A_t)_{t\geq 0}$ over the space $A \cup \{\emptyset\}$, where $A_t \in A$ corresponds to the receiver action seen in period t if τ_s has not yet stopped ($\tau_s > t$), and $A_t := \emptyset$ if τ_s has stopped ($\tau_s \leq t$). Enumerating $A = \{a_1, \ldots, a_n\}$, we write $p_{t,i} := \mathbb{P}_{\nu_s}[A_t = a_i]$ for $1 \leq i \leq n$ to record the probability of seeing receiver action a_i in period t and $p_{t,0} := \mathbb{P}_{\nu_s}[A_t = \emptyset] = \mathbb{P}_{\nu_s}[\tau_s \leq t]$ for the probability of seeing no receiver action in period t due to τ_s having stopped.

Given ν_s and τ_s, we define the *induced mixed actions after signal s*, $\pi_{2,s}(\nu_s, \tau_s, \beta) \in \Delta(A)$, by

$$\pi_{2,s}(\nu_s, \tau_s, \beta)(a) := \frac{\sum_{t=0}^{\infty} \beta^t p_{t,i}}{\sum_{t=0}^{\infty} \beta^t (1 - p_{t,0})} \quad \text{for } i \text{ such that } a = a_i.$$

As $\sum_{i=1}^n p_{t,i} = 1 - p_{t,0}$ for each $t \geq 0$, it is clear that $\pi_{2,s}(\nu_s, \tau_s, \beta)$ puts nonnegative weights on actions in A that sum to 1, so $\pi_{2,s}(\nu_s, \tau_s, \beta) \in \Delta(A)$ may indeed be viewed as a mixture over receiver actions.

Step 2: Induced mixed actions and per-period payoff.

We now show that for any β and any stopping time τ_s for signal s, the normalized payoff in the stopping problem is equal to the utility of playing s against $\pi_{2,s}(\nu_s, \tau_s, \beta)$ for one period, that is, that

$$u_1\big(\theta, s, \pi_{2,s}(\nu_s, \tau_s, \beta)\big) = \mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau_s-1} \beta^t \cdot u_1\big(\theta, s, a_s(t)\big)\right\} \Big/ \mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau_s-1} \beta^t\right\}.$$

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

$$\mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau_s-1} \beta^t\right\} = \mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\infty} [1_{\tau_s > t}] \cdot \beta^t\right\} = \sum_{t=0}^{\infty} \beta^t \cdot \mathbb{P}_{\nu_s}[\tau_s > t] = \sum_{t=0}^{\infty} \beta^t (1 - p_{t,0}),$$

and rewrite the numerator as

$$\mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau_s-1}\beta^t\cdot u_1\big(\theta,s,a_s(t)\big)\right\}=\sum_{t=0}^{\infty}\beta^t\cdot\Bigg(\underbrace{p_{t,0}\cdot 0}_{\text{get 0 if already stopped}}+\underbrace{\sum_{i=1}^{n}p_{t,i}\cdot u_1(\theta,s,a_i)}_{\text{else, } a_s(t)\text{ distributed as }(p_{t,i})}\Bigg)$$

$$=\sum_{i=1}^{n}\left(\sum_{t=0}^{\infty}\beta^t\cdot p_{t,i}\right)\cdot u_1(\theta,s,a_i).$$

So overall, we get, as desired:

$$\mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau_s-1}\beta^t\cdot u_1\big(\theta,s,a_s(t)\big)\right\}\Big/\mathbb{E}_{\nu_s}\left\{\sum_{t=0}^{\tau_s-1}\beta^t\right\}=\sum_{i=1}^{n}\left[\frac{\left(\sum_{t=0}^{\infty}\beta^t\cdot p_{t,i}\right)}{\sum_{t=0}^{\infty}\beta^t(1-p_{t,0})}\right]\cdot u_1(\theta,s,a_i)$$

$$=u_1\big(\theta,s,\pi_{2,s}(\nu_s,\tau_s,\beta)\big).\qquad \textit{Q.E.D.}$$

A.3. *Proof of Lemma 3*

LEMMA 3: *Let regular prior g_2, types θ', θ'', and signal s' be fixed. For every $\varepsilon>0$, there exists $C>0$ and $\underline{\gamma}<1$ so that for any $0\le\delta<1$, $\underline{\gamma}\le\gamma<1$, and $n\ge 1$, if $\pi_1(s'|\theta')\ge\pi_1(s'|\theta'')$ and $\pi_1(s'|\theta')\ge(1-\gamma)nC$, then*

$$\mathcal{R}_2[\pi_1]\big(\mathrm{BR}\big(P_{\theta'\triangleright\theta''},s'\big)|s'\big)\ge 1-\frac{1}{n}-\varepsilon.$$

We invoke Theorem 2 of Fudenberg, He, and Imhof (2017), which in our setting says:

Let regular prior g_2 and signal s' be fixed. Let $0<\varepsilon,h<1$. There exists C such that whenever $\pi_1(s'|\theta')\ge\pi_1(s'|\theta'')$ and $t\cdot\pi_1(s'|\theta')\ge C$, we get

$$\psi_2^{\pi_1}\left(y_2\in Y_2[t]:\frac{p\big(\theta''|s';y_2\big)}{p\big(\theta'|s';y_2\big)}\le\frac{1}{1-h}\cdot\frac{\lambda\big(\theta''\big)}{\lambda\big(\theta'\big)}\right)\Big/\psi_2^{\pi_1}\big(Y_2[t]\big)\ge 1-\varepsilon,$$

where $p(\theta|s;y_2)$ refers to the conditional probability that a sender of s is type θ according to the posterior belief induced by history y_2.

That is, if at age t a receiver would have observed in expectation C instances of type θ' sending s', then the belief of at least $1-\varepsilon$ fraction of age t receivers (essentially) falls in $P_{\theta'\triangleright\theta''}$ after seeing the signal s'. The proof of Lemma 3 calculates what fraction of receivers meets this "age requirement."

PROOF: We will show the following stronger result:

Let regular prior g_2, types θ', θ'', and signal s' be fixed. For every $\varepsilon>0$, there exists $C>0$ so that for any $0\le\delta$, $\gamma<1$ and $n\ge 1$, if $\pi_1(s'|\theta')\ge\pi_1(s'|\theta'')$ and $\pi_1(s'|\theta')\ge(1-\gamma)nC$, then

$$\mathcal{R}_2[\pi_1]\big(\mathrm{BR}\big(P_{\theta'\triangleright\theta''},s'\big)|s'\big)\ge\gamma^{\lceil\frac{1}{n(1-\gamma)}\rceil}-\varepsilon.$$

The lemma follows because we may pick a large enough $\underline{\gamma} < 1$ so that $\gamma^{\lceil \frac{1}{n(1-\gamma)} \rceil} > 1 - \frac{1}{n}$ for all $n \geq 1$ and $\gamma \geq \underline{\gamma}$.

For each $0 < h < 1$, define $P^h_{\theta' \triangleright \theta''} := \{p \in \Delta(\Theta) : \frac{p(\theta'')}{p(\theta')} \leq \frac{1}{1-h} \cdot \frac{\lambda(\theta'')}{\lambda(\theta')}\}$, with the convention that $\frac{0}{0} = 0$. Then it is clear that each $P^h_{\theta' \triangleright \theta''}$, as well as $P_{\theta' \triangleright \theta''}$ itself, is a closed subset of $\Delta(\Theta)$. Also, $P^h_{\theta' \triangleright \theta''} \to P_{\theta' \triangleright \theta''}$ as $h \to 0$.

Fix action $a \in A$. If, for all $\bar{h} > 0$, there exists some $0 < h \leq \bar{h}$ so that $a \in \mathrm{BR}(P^h_{\theta' \triangleright \theta''}, s')$, then $a \in \mathrm{BR}(P_{\theta' \triangleright \theta''}, s')$ also due to best-response correspondence having a closed graph. This means that, for each $a \notin \mathrm{BR}(P_{\theta' \triangleright \theta''}, s')$, there exists $\bar{h}_a > 0$ so that $a \notin \mathrm{BR}(P^h_{\theta' \triangleright \theta''}, s')$ whenever $0 < h \leq \bar{h}_a$. Let $\bar{h} := \min_{a \notin \mathrm{BR}(P_{\theta' \triangleright \theta''}, s')} \bar{h}_a$. Let $\varepsilon > 0$ be given and apply Theorem 2 of Fudenberg, He, and Imhof (2017) with ε and $\bar{h}$ to find constant C.

When $\pi_1(s'|\theta') \geq \pi_1(s'|\theta'')$ and $\pi_1(s'|\theta') \geq (1-\gamma)nC$, consider an age t receiver for $t \geq \lceil \frac{1}{n(1-\gamma)} \rceil$. Since $t \cdot \pi_1(s'|\theta') \geq C$, Theorem 2 of Fudenberg, He, and Imhof (2017) implies there is probability at least $1 - \varepsilon$ this receiver's belief about the types who send s' falls in $P^{\bar{h}}_{\theta' \triangleright \theta''}$. By construction of $\bar{h}$, $\mathrm{BR}(P^{\bar{h}}_{\theta' \triangleright \theta''}, s') = \mathrm{BR}(P_{\theta' \triangleright \theta''}, s')$, so $1 - \varepsilon$ of age t receivers have a history y_2 where $\sigma_2(y_2)(s') \in \mathrm{BR}(P_{\theta' \triangleright \theta''}, s')$.

Since agents survive between periods with probability γ, the mass of the receiver population aged $\lceil \frac{1}{n(1-\gamma)} \rceil$ or older is $(1-\gamma) \cdot \sum_{t=\lceil \frac{1}{n(1-\gamma)} \rceil}^{\infty} \gamma^t = \gamma^{\lceil \frac{1}{n(1-\gamma)} \rceil}$. This shows

$$\mathcal{R}_2[\pi_1]\big(\mathrm{BR}\big(P_{\theta' \triangleright \theta''}, s'\big)|s'\big) \geq \gamma^{\frac{1}{n(1-\gamma)}} \cdot (1-\varepsilon) \geq \gamma^{\lceil \frac{1}{n(1-\gamma)} \rceil} - \varepsilon,$$

as desired. *Q.E.D.*

A.4. *Proof of Proposition 2*

PROPOSITION 2: *$\pi^* \in \Pi^*(g, \delta, \gamma)$ if and only if $\mathcal{R}_1^{g,\delta,\gamma}[\pi_2^*] = \pi_1^*$ and $\mathcal{R}_2^{g,\delta,\gamma}[\pi_1^*] = \pi_2^*$.*

PROOF: *If*: Suppose π^* is such that $\mathcal{R}_1[\pi_2^*] = \pi_1^*$ and $\mathcal{R}_2[\pi_1^*] = \pi_2^*$. Consider the state ψ^* defined as $\psi^*_\theta := \psi_\theta^{\pi_2^*}$ for each θ and $\psi^*_2 := \psi_2^{\pi_1^*}$. Then, by construction, $\sigma_\theta(\psi_\theta^{\pi_2^*}) = \pi^*_\theta$ and $\sigma_2(\psi_2^{\pi_1^*}) = \pi_2^*$, so the state ψ^* gives rise to π^*. To verify that ψ^* is a steady state, we can expand by the definition of $\psi_\theta^{\pi_2^*}$,

$$f_\theta\big(\psi_\theta^{\pi_2^*}, \pi_2^*\big) = f_\theta\Big(\lim_{T\to\infty} f_\theta^T\big(\tilde{\psi}_\theta, \pi_2^*\big), \pi_2^*\Big),$$

where $\tilde{\psi}_\theta$ is any arbitrary initial state.

Since f_θ is continuous[22] at $\psi_\theta^{\pi_2^*}$ in L_1 distance defined in Footnote 20,

$$\lim_{T\to\infty} f_\theta^T\big(\tilde{\psi}_\theta, \pi_2^*\big) = \psi_\theta^{\pi_2^*}$$

is a fixed point of $f_\theta(\cdot, \pi_2^*)$. To see this, write $\psi_\theta^{(T)} := f_\theta^T(\tilde{\psi}_\theta, \pi_2^*)$ for each $T \geq 1$ and let $\varepsilon > 0$ be given. Continuity of f_θ implies there is $\zeta > 0$ so that $d(f_\theta(\psi_\theta^{\pi_2^*}, \pi_2^*), f_\theta(\psi_\theta^{(T)}, \pi_2^*)) <$

[22]This is implied by Step 1 of the proof of Proposition 3 in the Supplemental Material (Fudenberg and He (2018)), which shows f_θ is continuous at all states that assign $(1-\gamma)\gamma^t$ mass to the set of length-t histories.

$\varepsilon/2$ whenever $d(\psi_\theta^{\pi_2^*}, \psi_\theta^{(T)}) < \zeta$. So pick a large enough T so that $d(\psi_\theta^{\pi_2^*}, \psi_\theta^{(T)}) < \zeta$ and also $d(\psi_\theta^{\pi_2^*}, \psi_\theta^{(T+1)}) < \varepsilon/2$. Then

$$d\big(f_\theta\big(\psi_\theta^{\pi_2^*}, \pi_2^*\big), \psi_\theta^{\pi_2^*}\big) \le d\big(f_\theta\big(\psi_\theta^{\pi_2^*}, \pi_2^*\big), f_\theta\big(\psi_\theta^{(T)}, \pi_2^*\big)\big) + d\big(\psi_\theta^{(T+1)}, \psi_\theta^{\pi_2^*}\big) < \varepsilon/2 + \varepsilon/2.$$

Since $\varepsilon > 0$ was arbitrary, we have shown that $f_\theta(\psi_\theta^{\pi_2^*}, \pi_2^*) = \psi_\theta^{\pi_2^*}$ and a similar argument shows $f_2(\psi_2^{\pi_1^*}, \pi_1^*) = \psi_2^{\pi_1^*}$. This tells us $\psi^* = ((\psi_\theta^{\pi_2^*})_{\theta\in\Theta}, \psi_2^{\pi_1^*})$ is a steady state.

Only if: Conversely, suppose $\pi^* \in \Pi^*(g, \delta, \gamma)$. Then there exists a steady state $\psi^* \in \Psi^*(g, \delta, \gamma)$ such that $\pi^* = \sigma(\psi^*)$. This means $f_\theta(\psi_\theta^*, \pi_2^*) = \psi_\theta^*$, so iterating shows

$$\psi_\theta^{\pi_2^*} := \lim_{T\to\infty} f_\theta^T\big(\psi_\theta^*, \pi_2^*\big) = \psi_\theta^*.$$

Since $\mathcal{R}_1[\pi_2^*](\cdot|\theta) := \sigma_\theta(\psi_\theta^{\pi_2^*})$, the above implies $\mathcal{R}_1[\pi_2^*](\cdot|\theta) = \sigma_\theta(\psi_\theta^*) = \pi_1^*(\cdot|\theta)$ by the choice of ψ^*. We can similarly show $\mathcal{R}_2[\pi_1^*] = \pi_2^*$. *Q.E.D.*

A.5. *Proof of Theorem 3*

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

FACT—Hoeffding's Inequality: *Suppose $X_1, \dots, X_n$ 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,*

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

LEMMA A.1: *In strategy profile π^*, suppose s^* is on-path and $\pi_2^*(a^*|s^*) = 1$, where a^* is a strict best response to s^* given π_1^*. Then there exists $N \in \mathbb{R}$ so that, for any regular prior and any sequence of steady-state strategy profiles $\pi^{(k)} \in \Pi^*(g, \delta_k, \gamma_k)$ where $\gamma_k \to 1, \pi^{(k)} \to \pi^*$, there exists $K \in \mathbb{N}$ such that whenever $k \ge K$, we have $\pi_2^{(k)}(a^*|s^*) \ge 1 - (1 - \gamma_k) \cdot N$.*

PROOF: Since a^* is a strict best response after s^* for π_1^*, there exists $\varepsilon > 0$ so that a^* will continue to be a strict best response after s^* for any $\pi_1' \in \Pi_1$ where, for every $\theta \in \Theta$, $|\pi_1'(s^*|\theta) - \pi_1^*(s^*|\theta)| < 3\varepsilon$.

Since $\pi^{(k)} \to \pi^*$, find large enough K such that $k \ge K$ implies, for every $\theta \in \Theta$, $|\pi_1^{(k)}(s^*|\theta) - \pi_1^*(s^*|\theta)| < \varepsilon$.

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 $e^{\text{obs}}_{n,\theta}$:

$$e^{\text{obs}}_{n,\theta} \leq \mathbb{P}\left[\left|S_n - \mathbb{E}[S_n]\right| \geq \frac{1}{2}n\lambda(\theta)\right] \leq 2\exp\left(-\frac{2\cdot\left[\frac{1}{2}n\lambda(\theta)\right]^2}{n}\right).$$

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

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

So we set $N^{\text{obs}} := \sum_{\theta\in\Theta} N^{\text{obs}}_{\theta}$.

Next, write $e^{\text{bias},k}_{n,\theta}$ for the probability that, after observing $\lfloor\frac{1}{2}n\lambda(\theta)\rfloor$ i.i.d. draws from $\pi_1^{(k)}(\cdot|\theta)$, the empirical frequency of signal s^* differs from $\pi_1^{(k)}(s^*|\theta)$ by more than 2ε. So again, write $Z_t^{\theta,k} \in \{0,1\}$ to indicate if the tth draw resulted in signal s^*, with $\mathbb{E}[Z_t^{\theta,k}] = \pi_1^{(k)}(s^*|\theta)$, and put $S_{n,k} := \sum_{t=1}^{\lfloor\frac{1}{2}n\lambda(\theta)\rfloor} Z_t^{\theta,k}$ for total number of s^* out of $\lfloor\frac{1}{2}n\lambda(\theta)\rfloor$ draws. We have $\mathbb{E}[S_{n,k}] = \lfloor\frac{1}{2}n\lambda(\theta)\rfloor\cdot\pi_1^{(k)}(s^*|\theta)$, but $|\pi_1^{(k)}(s^*|\theta) - \pi_1^*(s^*|\theta)| < \varepsilon$ whenever $k \geq K$. That means

$$\begin{aligned}
e^{\text{bias},k}_{n,\theta} &:= \mathbb{P}\left[\left|\frac{S_{n,k}}{\left\lfloor\frac{1}{2}n\lambda(\theta)\right\rfloor} - \pi_1^*\big(s^*|\theta\big)\right| \geq 2\varepsilon\right] \\
&\leq \mathbb{P}\left[\left|\frac{S_{n,k}}{\left\lfloor\frac{1}{2}n\lambda(\theta)\right\rfloor} - \pi_1^{(k)}\big(s^*|\theta\big)\right| \geq \varepsilon\right] \quad \text{if } k \geq K \\
&= \mathbb{P}\left[\left|S_{n,k} - \mathbb{E}[S_{n,k}]\right| \geq \left\lfloor\frac{1}{2}n\lambda(\theta)\right\rfloor\cdot\varepsilon\right] \\
&\leq 2\exp\left(-\frac{2\cdot\left(\left\lfloor\frac{1}{2}n\lambda(\theta)\right\rfloor\cdot\varepsilon\right)^2}{\left\lfloor\frac{1}{2}n\lambda(\theta)\right\rfloor}\right) \quad \text{by Hoeffding's inequality.}
\end{aligned}$$

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

Finally, since g is regular, we appeal to Proposition 1 of Fudenberg, He, and Imhof (2017) 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 s^* differs from the empirical distribution by no more than ε. Put $N^{\text{age}} := \frac{2\underline{N}}{\min_{\theta \in \Theta} \lambda(\theta)}$.

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 s^* differs from $\pi_1^*(s^*|\theta)$ 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 s^* is closer than 3ε to $\pi_1^*(s^*|\theta)$. By the construction of ε, playing a^* after s^* is the unique best response to such a posterior.

Therefore, for $k \geq K$, the probability that the sender population plays some action other than a^* after s^* 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 $\sum_{\theta \in \Theta} e_{n,\theta}^{\text{obs}}$ fraction has a sample size smaller than $\frac{1}{2} n\lambda(\theta)$ for any type θ, while $\sum_{\theta \in \Theta} e_{n,\theta}^{\text{bias},k}$ 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 s^* differs by more than 2ε from $\pi_1^*(s^*|\theta)$.

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

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

and

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

We conclude that whenever $k \geq K$,

$$\pi_2^{(k)}\left(a^*|s^*\right) \geq 1 - (1 - \gamma_k) \cdot \left(N^{\text{age}} + N^{\text{obs}} + N^{\text{bias}} \right).$$

Finally, observe that none of $N^{\text{age}}, N^{\text{obs}}, N^{\text{bias}}$ depends on the sequence $\pi^{(k)}$, so N is chosen independent of the sequence $\pi^{(k)}$. *Q.E.D.*

LEMMA A.2: *Assume g is regular. Suppose there is some $a^* \in A$ and $v \in \mathbb{R}$ so that $u_1(\theta, s^*, a^*) > v$. Then, there exist $C_1 \in (0, 1)$, $C_2 > 0$ so that in every sender history y_θ, $\#(s^*, a^*|y_\theta) \geq C_1 \cdot \#(s^*|y_\theta) + C_2$ implies $\mathbb{E}[u_1(\theta, s^*, \pi_2(\cdot|s^*))|y_\theta] > v$.*

PROOF: Write $\underline{u} := \min_{a \in A} u_1(\theta, s^*, a)$. There exists $q \in (0, 1)$ so that

$$q \cdot u_1\left(\theta, s^*, a^*\right) + (1 - q) \cdot \underline{u} > v.$$

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

Since g is regular, Proposition 1 of Fudenberg, He, and Imhof (2017) tells us there exists some C_0 so that the posterior mean belief of sender with history y_θ is no less than

$$(1-\varepsilon)\cdot\frac{\#\big(s^*,a^*|y_\theta\big)}{\#\big(s^*|y_\theta\big)+C_0}.$$

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

$$(1-\varepsilon)\cdot\frac{\#\big(s^*,a^*|y_\theta\big)}{\#\big(s^*|y_\theta\big)+C_0}\geq q \quad\Longleftrightarrow\quad \#\big(s^*,a^*|y_\theta\big)\geq\frac{q}{1-\varepsilon}\#\big(s^*|y_\theta\big)+\frac{q}{1-\varepsilon}\cdot C_0.$$

Putting $C_1:=\frac{q}{1-\varepsilon}$ and $C_2:=\frac{q}{1-\varepsilon}\cdot C_0$ proves the lemma. *Q.E.D.*

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

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

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*

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

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

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

For every $G_1>0$ and every $0<\varepsilon<1$,

$$\begin{aligned}\mathbb{P}[S_n\geq C_1n+C_2\ \forall n\geq G_1]&=1-\mathbb{P}\left[(\exists n\geq G_1)\sum_{t=1}^n Z_t<C_1n+C_2\right]\\&=1-\mathbb{P}\left[(\exists n\geq G_1)\sum_{t=1}^n(X_t-\varepsilon)>(1-\varepsilon-C_1)n-C_2\right],\end{aligned}$$

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

$$\mathbb{P}[S_n\geq C_1n+C_2\ \forall n\geq G_1]\geq 1-\mathbb{P}\left[(\exists n\geq G_1)\sum_{t=1}^n(X_t-\varepsilon)>\frac{1}{2}\bar\varepsilon n\right]$$

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

We now prove Theorem 3.

THEOREM 3: *Suppose π^* is on-path strict for the receiver and patiently stable. Then it satisfies the strong compatibility criterion.*

PROOF: Let some $a' \notin \mathrm{BR}(\Delta(\tilde{J}(s', \pi^*)), s')$ and $h > 0$ be given. We will show that $\pi_2^*(a'|s') \leq 3h$.

Step 1: Defining the constants ξ, θ^J, a_θ, s_θ, C_1, C_2, G_1, G_2, and N^{recv}.

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

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

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

(ii) Since $\tilde{J}(s', \pi^*)$ is nonempty, we can fix some $\theta^J \in \tilde{J}(s', \pi^*)$.

(iii) For each equilibrium-dominated type $\theta \in \Theta \setminus \tilde{J}(s', \pi^*)$, identify some on-path signal s_θ so that $\pi_1^*(s_\theta|\theta) > 0$. By assumption of on-path strictness for the receiver, there is some $a_\theta \in A$ so that $\pi_2^*(a_\theta|s_\theta) = 1$, and furthermore, a_θ is the strict best response to s_θ in π^*. By the definition of equilibrium dominance,

$$u_1(\theta, s_\theta, a_\theta) > \max_{a \in A} u_1\big(\theta, s', a\big) =: v_\theta.$$

By applying Lemma A.2 to each $\theta \in \Theta \setminus \tilde{J}(s', \pi^*)$, we obtain some $C_1 \in (0, 1)$, $C_2 > 0$ so for every $\theta \in \Theta \setminus \tilde{J}(s', \pi^*)$ and in every sender history y_θ, $\#(s_\theta, a_\theta|y_\theta) \geq C_1 \cdot \#(s_\theta|y_\theta) + C_2$ implies $\mathbb{E}[u_1(\theta, s_\theta, \pi_2(\cdot|s_\theta))|y_\theta] > v_\theta$.

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

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

(v) Because at π^*, a_θ is a strict best response to s_θ for every $\theta \in \Theta \setminus \tilde{J}(s', \pi^*)$, from Lemma A.1 we may find a N^{recv} so that for each sequence $\pi^{(k)} \in \Pi^*(g, \delta_k, \gamma_k)$ where $\gamma_k \to 1, \pi^{(k)} \to \pi^*$, there corresponds $K^{\mathrm{recv}} \in \mathbb{N}$ so that $k \geq K^{\mathrm{recv}}$ implies $\pi_2^{(k)}(a_\theta|s_\theta) \geq 1 - (1 - \gamma_k) \cdot N^{\mathrm{recv}}$ for every $\theta \in \Theta \setminus \tilde{J}(s', \pi^*)$.

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

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 (2017), 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 $\pi_1(s'|\theta^J) \geq c \cdot \pi_1(s'|\theta^{\mathrm{D}})$ for every equilibrium-dominated type $\theta^{\mathrm{D}} \notin \tilde{J}(s', \pi^*)$ and

$$n \cdot \pi_1\big(s'|\theta^J\big) \geq N^{\mathrm{rare}}, \tag{7}$$

then an age-n receiver in steady state ψ where $\pi = \sigma(\psi)$ has probability at least $1 - h$ of holding a posterior belief $g_2(\cdot|y_2)$ such that θ^J is at least $\frac{1}{2}c$ times as likely to play s' as θ^{D}

is for every $\theta^D \notin \tilde{J}(s', \pi^*)$. Thus, history y_2 generates a posterior belief after s', $p(\cdot|s'; y_2)$ such that

$$\frac{p(\theta^D|s'; y_2)}{p(\theta^J|s'; y_2)} \leq \frac{\lambda(\theta^D)}{\lambda(\theta^J)} \cdot \xi \cdot \frac{\lambda(\theta^J)}{\max_{\theta \in \Theta} \lambda(\theta)} \leq \xi.$$

In particular, $p(\cdot|s'; y_2)$ must assign weight no greater than ξ to each type not in $\tilde{J}(s', \pi^*)$; therefore, the belief belongs to $\Delta_\xi(\tilde{J}(s', \pi^*))$. By construction of ξ, a' is then not a best response to s' after history y_2.

A receiver whose age n satisfies Equation (7) plays a' with probability less than h, provided $\pi_1(s'|\theta^J) \geq c \cdot \pi_1(s'|\theta^D)$ for every $\theta^D \notin \tilde{J}(s', \pi^*)$. However, to bound the overall probability of a' in the entire receiver population in steady state ψ, we ensure that Equation (7) is satisfied for all except $2h$ fraction of receivers in ψ. We claim that when γ is large enough, a sufficient condition is for $\pi = \sigma(\psi)$ to satisfy $\pi_1(s'|\theta^J) \geq (1-\gamma)N^*$ 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 (7), while the fraction of receivers younger than $\frac{h}{1-\gamma}$ is $1 - (\gamma^{\frac{h}{1-\gamma}}) \leq 2h$ 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 $\pi = \sigma(\psi)$ has $\pi_2(a'|s') \leq 3h$ if the following two conditions are satisfied:

(*C1*) $\pi_1(s'|\theta^J) \geq c \cdot \pi_1(s'|\theta^D)$ for every equilibrium-dominated type $\theta^D \notin \tilde{J}(s', \pi^*)$;

(*C2*) $\pi_1(s'|\theta^J) \geq (1-\gamma)N^*$ 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 $\sigma(\psi^{(k)}) = \pi^{(k)} \to \pi^*$ such that, in every $\pi^{(k)}$, the above two conditions are satisfied. Using the fact that $\gamma_k \to 1$, we conclude that, for large enough k, we get $\pi_2^{(k)}(a'|s') \leq 3h$, which in turn shows $\pi^*(a'|s') \leq 3h$ due to the convergence $\pi^{(k)} \to \pi^*$.

Step 3: Extracting a suitable subsequence of steady states.

In the statement of Lemma 4, put $\theta' := \theta^J$. We obtain some number ε and functions $\bar{\delta}(N)$, $\bar{\gamma}(N, \delta)$. 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 $\pi^{(k)} = \lim_{j \to \infty} \pi^{(k,j)}$ with $\pi^{(k,j)} \in \Pi^*(g, \delta_k, \gamma_{k,j})$ with $\lim_{j \to \infty} \gamma_{k,j} = 1$. It is without loss to assume that for every $k \geq 1$, $\delta_k \geq \bar{\delta}(N^*)$, and that the L_1 distance between $\pi^{(k)}$ and π^* is less than $\varepsilon/2$. 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 $\pi^{(k,j)}$ and $\pi^{(k)}$ is less than $\min(\frac{\varepsilon}{2}, \frac{1}{k})$, 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) $\delta_k \geq \bar{\delta}(N^*)$ for each k; (3) $\gamma_k \geq \bar{\gamma}(N^*, \delta_k)$ for each k; (4) $\pi^{(k)} \to \pi^*$; (5) the L_1 distance between $\psi^{(k)}$ and π^* is no larger than ε. Lemma 4 implies that, for every k, $\pi_1^{(k)}(s'|\theta^J) \geq (1-\gamma_k)N^*$. 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 the aggregate receiver strategy $\pi_2^{(k)}$ converges to π_2^*. By Lemma A.1, there exists some K^{recv} so that $k \geq K^{\text{recv}}$ implies

$\pi_2^{(k)}(a_\theta|s_\theta) \geq 1-(1-\gamma_k)\cdot N^{\text{recv}}$ for every $\theta \in \Theta \setminus \tilde{J}(s', \pi^*)$. Find next a large enough K^{error} so that $k \geq K^{\text{error}}$ implies $(1-\gamma_k)\cdot N^{\text{recv}} < \bar{\varepsilon}$ (where $\bar{\varepsilon}$ was defined in *Step 1*).

We claim that when $k \geq \max(K^{\text{recv}}, K^{\text{error}})$, a type $\theta \notin \tilde{J}(s', \pi^*)$ sender who always sends signal s_θ against a receiver population that plays $\pi_2^{(k)}(\cdot|s_\theta)$ 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 s_θ is no greater than v_θ in some period $n \geq G_1$. This is because by Lemma A.3,

$$\mathbb{P}[S_n \geq C_1 n + C_2 \ \forall n \geq G_1] \geq 1 - G_2 \cdot \pi_2^{(k)}\big(\{a \neq a_\theta\}|s_\theta\big) \geq 1 - G_2 \cdot (1-\gamma_k)\cdot N^{\text{recv}},$$

where S_n refers to the number of times that the receiver population responded to s_θ with a_θ in the first n times that s_θ was sent. But Lemma A.2 guarantees that, provided $S_n \geq C_1 n + C_2$, sender's expected payoff for s_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 $\theta \notin \tilde{J}(s', \pi^*)$, the Gittins index for signal s_θ cannot fall below v_θ if s_θ has been used no more than G_1 times. (This is possible since the prior is non-doctrinaire.) 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 $\theta \notin \tilde{J}(s', \pi^*)$ will play s' even once. To see this, we observe that according to the prior, the Gittins index for s_θ is higher than that of s', whose index is no higher than its highest possible payoff v_θ. This means the sender will not play s' until her Gittins index for s_θ has fallen below v_θ. Since $k \geq K^{\text{recv}}$, this will not happen before the sender has played s_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 s_θ (and, a fortiori, the Gittins index for s_θ) ever falling below v_θ sometime after playing s_θ for the G_1th time is no larger than $G_2\cdot(1-\gamma_k)\cdot N^{\text{recv}}$.

This shows that, for $k \geq \max(K^{\text{recv}}, K^{\text{error}}, K^{\text{Gittins}})$, $\pi_1^{(k)}(s'|\theta) \leq G_2 N^{\text{recv}}\cdot(1-\gamma_k)$ for every $\theta \notin \tilde{J}(s', \pi^*)$. But since $\pi_1^{(k)}(s'|\theta^J) \geq N^*\cdot(1-\gamma_k)$ 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 $k \geq \max(K^{\text{recv}}, K^{\text{error}}, K^{\text{Gittins}})$. *Q.E.D.*

REFERENCES

BANKS, J. S., AND J. SOBEL (1987): "Equilibrium Selection in Signaling Games," *Econometrica*, 55, 647–661. [1215,1220]

BELLMAN, R. (1956): "A Problem in the Sequential Design of Experiments," *Sankhya. The Indian Journal of Statistics*, 16, 221–229. [1230]

BILLINGSLEY, P. (1995): *Probability and Measure*. New York: Wiley. [1251]

CHO, I.-K., AND D. M. KREPS (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179–221. [1215,1217,1219,1220,1222,1226,1239,1242]

DEKEL, E., D. FUDENBERG, AND D. K. LEVINE (1999): "Payoff Information and Self-Confirming Equilibrium," *Journal of Economic Theory*, 89, 165–185. [1220]

——— (2004): "Learning to Play Bayesian Games," *Games and Economic Behavior*, 46, 282–303. [1220,1221,1236,1244]

DIACONIS, P., AND D. FREEDMAN (1990): "On the Uniform Consistency of Bayes Estimates for Multinomial Probabilities," *The Annals of Statistics*, 18, 1317–1327. [1222]

ESPONDA, I., AND D. POUZO (2016): "Berk–Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models," *Econometrica*, 84, 1093–1130. [1220]

FUDENBERG, D., AND K. HE (2017): "Learning and Equilibrium Refinements in Signalling Games," Working Paper. Available at arXiv:1709.01024. [1220,1244]

——— (2018): "Supplement to 'Learning and Type Compatibility in Signaling Games'," *Econometrica Supplemental Material*, 86, https://doi.org/10.3982/ECTA15085. [1226,1235-1237,1239,1241,1247]

FUDENBERG, D., AND Y. KAMADA (2018): "Rationalizable Partition-Confirmed Equilibrium With Heterogeneous Beliefs," *Games and Economic Behavior*, 109, 364–381. [1236]

FUDENBERG, D., AND D. M. KREPS (1988): "A Theory of Learning, Experimentation, and Equilibrium in Games," Working Paper. [1220,1243]

——— (1993): "Learning Mixed Equilibria," *Games and Economic Behavior*, 5, 320–367. [1217,1220,1222]

——— (1994): "Learning in Extensive-Form Games, II: Experimentation and Nash Equilibrium," Working Paper. [1220,1222]

——— (1995): "Learning in Extensive-Form Games I. Self-Confirming Equilibria," *Games and Economic Behavior*, 8, 20–55. [1220]

FUDENBERG, D., AND D. K. LEVINE (1993): "Steady State Learning and Nash Equilibrium," *Econometrica*, 61, 547–573. [1220,1223,1235,1237,1238,1244]

——— (2006): "Superstition and Rational Learning," *American Economic Review*, 96, 630–651. [1220,1223, 1235,1244,1251]

FUDENBERG, D., K. HE, AND L. A. IMHOF (2017): "Bayesian Posteriors for Arbitrarily Rare Events," *Proceedings of the National Academy of Sciences of the United States of America*, 114, 4925–4929. [1219,1234,1242, 1246,1247,1249,1251,1252]

GITTINS, J. C. (1979): "Bandit Processes and Dynamic Allocation Indices," *Journal of the Royal Statistical Society, Series B*, 41, 148–177. [1215,1225]

JEHIEL, P., AND D. SAMET (2005): "Learning to Play Games in Extensive Form by Valuation," *Journal of Economic Theory*, 124, 129–148. [1220]

KALAI, E., AND E. LEHRER (1993): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61, 1019–1045. [1220]

LASLIER, J.-F., AND B. WALLISER (2015): "Stubborn Learning," *Theory and Decision*, 79, 51–93. [1220]

NIÑO-MORA, J. (2011): "Computing a Classic Index for Finite-Horizon Bandits," *INFORMS Journal on Computing*, 23, 254–267. [1244]

SOBEL, J., L. STOLE, AND I. ZAPATER (1990): "Fixed-Equilibrium Rationalizability in Signaling Games," *Journal of Economic Theory*, 52, 304–331. [1243]

SPENCE, M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355–374. [1216,1226]

Co-editor Joel Sobel handled this manuscript.

Manuscript received 6 February, 2017; final version accepted 6 February, 2018; available online 15 February, 2018.