Observability, Dominance, and Induction in Learning Models*

Daniel Clark,[†] Drew Fudenberg,[‡] and Kevin He[§]

August 19, 2022

Abstract

Learning models do not in general imply that weakly dominated strategies are irrelevant or justify the related concept of "forward induction," because rational agents may use dominated strategies as experiments to learn how opponents play, and may not have enough data to rule out a strategy that opponents never use. Learning models also do not support the idea that the selected equilibria should only depend on a game's reduced normal form. However, playing the extensive form of a game is equivalent to playing the normal form augmented with the appropriate *terminal node partitions* so that two games are *information equivalent*, i.e., the players receive the same feedback about others' strategies.

Keywords: learning in games, equilibrium refinements, iterated dominance, forward induction

*We thank Adam Brandenburger, Amanda Friedenberg, Ying Gao, and George Mailath for helpful conversations, Jonathan Brownrigg for research assistance, and NSF grant 1951056 for financial support.

[†]Department of Economics, MIT. Email: dgclark@mit.edu

[‡]Department of Economics, MIT. Email: drew.fudenberg@gmail.com

[§]Department of Economics, University of Pennsylvania: hesichao@gmail.com

1 Introduction

Not all Nash equilibria seem equally plausible, which has led to an interest in various refinements of Nash equilibria, such as "forward induction" and Kohlberg and Mertens (1986)'s *strategic stability*.[1] The learning in games literature asks which equilibria are likely to persist in environments where new players are initially uncertain about the prevailing strategies and learn about the strategy distribution by repeatedly playing the game. This paper shows that learning models with patient players can lead to very different predictions than the axiomatically-justified refinements and invariance conditions proposed in the literature following Kohlberg and Mertens (1986).

We focus on patient players because learning by myopic agents need not lead to Nash equilibrium, as shown by Fudenberg and Levine (1993) and Fudenberg and Kreps (1995). Specifically, we use the model of learning by patient players with geometric lifetimes developed in Fudenberg and He (2018).[2] In the special class of signaling games, the long-run outcomes predicted by this model, and the related learning models of Fudenberg and He (2020) and Clark and Fudenberg (2021), resemble the equilibrium refinements of Banks and Sobel (1987) and Cho and Kreps (1987), which are implied by strategic stability. This paper explores the reasons that the predictions of learning models and classic refinements differ more substantially in other sorts of extensive forms.

There are two distinct reasons that the outcomes of learning models need not satisfy forward induction or iterated weak dominance. First, a dominated strategy may be used as an experiment to gain information about opponents' play at some off-path information sets, and the opponents may then correctly believe that the rare deviations from the equilibrium path use this dominated strategy. Second, even if a dominated

[1] There are many related definitions of forward induction in the literature, see the papers surveyed in Govindan and Wilson (2009).

[2] Fudenberg and He (2020) and Clark and Fudenberg (2021) also used learning models with geometrically distributed lifetimes. Fudenberg and Levine (1993) and Fudenberg and Levine (2006) assumed agents have fixed finite lifetimes, but this is not relevant for our results.

strategy is never used, agents in other player roles may not learn this if they start with a prior belief to the contrary and don't obtain enough data to learn the truth.

The Kohlberg and Mertens (1986) argument that a solution concept for games should only depend on the normal form is based on the claim that the differences between extensive forms with the same normal form are "irrelevant details" because they do not change the decision problem of a player who faces the *same fixed and known* strategies of the opponents. Because the normal form abstracts from many aspects of game play that are relevant for how people learn what strategies are used by others, there is no reason to expect learning to depend only on this very abstract representation of strategic interaction. Instead, the set of learning outcomes is only invariant to transformations that are both *decision invariant*, i.e., lead to the same best responses as a function of opponent strategies, and *information invariant* in the sense of providing the same feedback to the agents in their learning problems. Specifically, learning outcomes, unlike sequential equilibria, are invariant to the coalescing of consecutive moves by the same player. However, like sequential equilibria and unlike the various definitions of strategic stability, learning outcomes are not invariant to replacing an extensive-form game with the corresponding game in normal form: In the latter case there are no unreached information sets, and the terminal node that is reached reveals the strategy used by each player.

To capture what is essential for learning outcomes using the normal form, we augment it with *terminal node partitions* (Fudenberg and Kamada (2015), Fudenberg and Kamada (2018)) which describe the information players observe when the game is played. We show that from a learning perspective, the extensive-form game is equivalent to the simultaneous-move game corresponding to the normal form with terminal node partitions that give players the same information as would be revealed by the terminal nodes in the extensive form. We also show that if agents play the simultaneous-move version of a normal-form game and observe the realized pure strategies at the end of each play, the learning outcome is a refinement of backward induction and of $S^{\infty}W$ (Dekel and Fudenberg (1990)), but does not imply iterated weak dominance.

2 Informal Overview of the Learning Model

We begin with an informal overview of the learning model, deferring the full description of the learning model until Section 4. We consider an overlapping generations learning environment where time is discrete and doubly infinite, $t \in \{..., -2, -1, 0, 1, 2, ...\}$. There is a continuum of agents of mass 1 in each player role $i \in \{1, ..., I\}$. The agents have geometric lifespans, with i.i.d. survival probability γ per period. Each period newborn agents replace the departing agents so the sizes of the various populations are constant, and then agents are anonymously matched to play a fixed finite extensive-form stage game of perfect recall.

The game has information sets $\mathcal{H}_i$ for each player $i \in \{1, ..., I\}$, with available actions A_h at each $h \in \mathcal{H}_i$. A pure strategy $s_i \in S_i$ of i assigns an action $s_i(h) \in A_h$ to every information set h of i. Denote the (finite) set of terminal nodes of the game tree as Z, and let $\mathbf{z}(s)$ denote the terminal node reached by strategy profile s. Player i has a utility function defined on terminal nodes, $u_i : Z \to \mathbb{R}$ and a corresponding utility function on strategy profiles $u_i(s) = u_i(\mathbf{z}(s))$.

Each agent has a *terminal node partition* $\mathcal{P}_i$ (Fudenberg and Kamada (2015), Fudenberg and Kamada (2018)) over Z, and they observe which partition element contains the terminal node of their match at the end of each period.[3] In previous analyses of explicit learning models, this partition is discrete, i.e., every partition element is singleton and all agents observe the realized terminal node, and this will be our default assumption. However, in some settings it is natural to assume that agents observe less; for example, in a sealed-bid first price auction, agents might only observe the winning bid.

All agents are rational Bayesians who choose *policies* (maps from history of past

[3]Fudenberg and Kamada (2015) analyze settings where each player moves only once, and players who choose an **Out** action do not observe the choices made by others. The rationalizable conjectural equilibria of Rubinstein and Wolinsky (1994) and Esponda (2013) use *signal functions* to model what players observe when the game is played. These papers do not explicitly consider extensive-form games so their signal functions are more abstract.

observations to current play) that maximize their expected discounted payoff. They are born with priors over the prevailing steady-state distribution of play in the opponent populations, which they update using their observations. In every period t, the *state* of the system is the shares of agents in a given player role with the various possible histories. The state and the optimal policies induce an *aggregate strategy* that describes the distribution of strategies in each player-role population, and thus an *update rule* that maps states in period t to states in period $t+1$. We study this system's steady states, which are the fixed points of the update rule.

Agent's observations can depend on their play, so their optimal policies may incorporate a value for "experimenting" with various strategies that have the potential to improve payoff. The size of the experimentation incentive depends on their continuation probability, their discount factor $\delta \in [0,1)$, and how much they have already learned: inexperienced agents have more incentive to experiment, and they cease experimenting when they have enough data.

We focus on the limits of steady-state play when γ tends to 1, so agents can acquire enough observations to outweigh their prior. We also assume that δ goes to 1. Otherwise, agents may not experiment enough to rule out limits that are not Nash equilibria. We call the strategy profiles that emerge in this limit *patiently stable*, and say that any distribution over terminal nodes generated by a patiently stable profile is a *patiently stable outcome.*

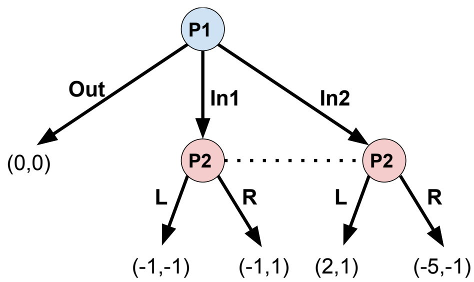
3 Examples

3.1 Failures of Forward Induction and Iterated Weak Dominance

We give simple examples to show that equilibria that violate minimal notions of forward induction or the related concept of iterated weak dominance can be patiently stable.

3.1.1 Information Value of Dominated Strategies

Consider the following game: P1 chooses from **Out**, **In1**, and **In2**. If P1 chooses **Out**, the game is over and each player gets 0. If P1 chooses **In1** or **In2**, P2 plays **L** or **R** without knowing P1's choice. The figure below shows the game in its extensive-form and normal-form representations.



	L	**R**
Out	0,0	0,0
In1	-1,-1	-1,1
In2	2,1	-5,-1

Figure 1: **In1** is strictly dominated by **Out** but provides the same information as **In2** and performs better than **In2** against some P2 strategies.

The strategy **In1** is strictly dominated by **Out** for P1, and the iterated-dominance criterion of Kohlberg and Mertens (1986) requires that "A solution of a game G contains a solution of any game G' obtained from G by deletion of a dominated strategy." In the game G' that results from the deletion of **In1**, $(\mathbf{In2}, \mathbf{L})$ is the only sequential equilibrium and so the only strategically stable equilibrium. Thus the Nash equilibrium $(\mathbf{Out}, \mathbf{R})$ is ruled out by forward induction.

In contrast, when an inexperienced P1 agent plays this game and observes the terminal node at the end of each match, the agent may find it optimal to play **In1**. This is because **In1** and **In2** are informationally equivalent experiments: they provide the same signal about how P2s play. But if P1's current belief puts much higher probability on P2s playing **R** than **L**, then P1's expected payoff from **In1** exceeds that of **In2**. A sufficiently patient P1 agent will choose to experiment and learn about P2's play in order to figure out whether **Out** or **In2** is a better response against the aggregate P2 play, but the cheapest such experiment may be **In1**.[4]

[4]Fudenberg and Levine (1993), footnote 10 pointed out the possibility that this might occur in their closely related learning model but did not provide a proof that it does.

Claim 1. ($\boldsymbol{Out}, \boldsymbol{R}$) *is a patiently stable strategy profile for the game in Figure 1.*

In Section 5.1 we establish more general sufficient conditions for patient stability in two-player games where each player moves at most once. These conditions give us a class of games where patiently stable profiles fail forward induction because of the informational value of dominated strategies. The idea of the proof is to choose "supportive" priors that lead the early mover to experiment in a way consistent with the desired equilibrium (such as choosing **In1** instead of **In2** in the example above) unless they have previously seen an out-of-equilibrium response from the second mover.

3.1.2 Insufficient Data to Eliminate Weakly Dominated Opponent Play

In the previous example, there is a dominated strategy that is still used by a rational agent as it provides information about their opponents' aggregate play. By contrast, for the game in Figure 2, the strategy **In2** is *doubly dominated* by **In1** for P2 agents: it provides the same information about opponent play but, whenever these actions can be played, **In2** always gives a strictly lower payoff than **In1**. A rational P2 agent will therefore never play **In2** even as an experiment, which makes it more surprising that the learning outcome for patient and long-lived agents can select a profile where P1 and P2 are deterred from entering by P3's **R**, which is strictly inferior to **L** against **In1**.

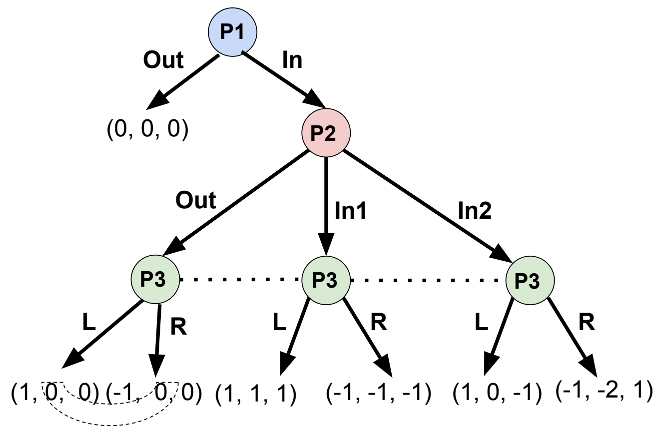


Figure 2: A three-player game where the strategy **In2** is doubly dominated by **In1** for P2. The dotted lines connecting two terminal nodes represent P2's terminal node partition.

Here we suppose that P1 and P3 always observe the terminal node, but P2's terminal node partition is such that they do not learn how P3 plays if they choose **Out**. Note that once the doubly dominated **In2** is deleted for P2, **L** is a strictly better strategy than **R** for P3 against any strictly mixed play of P1 and P2. But:

Claim 2. (***Out***, ***Out***, ***R***) *is a patiently stable strategy profile for the game in Figure 2.*

The presence of a third player is critical to this conclusion. The idea is that although aggregate P2 play puts zero probability on **In2** (as required by the elimination of weakly dominated strategies) and positive probability on **In1**, a P3 agent may not have enough data to learn this aggregate play, as P3s only observe a P2 agent entering when they encounter both a P1 and a P2 agent experimenting with some **In** action. The incentive for P2 to experiment is weak because they are located off the equilibrium path and do not expect to play often, as in Fudenberg and Levine (2006). As a result, most P3 agents will never obtain any data to correct a prior belief that says it is more likely for P2's to choose **In2** than **In1**, so they find it optimal to play **R**. Even though the aggregate steady-state play of the P2s puts zero probability on the weakly dominated strategy, most P3s fail to learn this. We formally analyze this example in Section 5.2.

3.2 Invariance

The refinements literature following Kohlberg and Mertens (1986) states that the selected set of equilibria “should only depend on the reduced normal form of the game,” so that any two extensive forms with the same reduced form will be played in the same way. According to Kohlberg and Mertens (1986), this follows from the fact that the reduced normal form “captures all the relevant information for decision purposes...” Implicitly, this argument for invariance holds each player’s beliefs about the play of their opponents fixed.

This is not true for the selections made by learning, since extensive forms with the same reduced normal form can give players different information about how their opponents play and so lead to different outcomes. As an example, compare the two games in Figure 3. In the game on the left, the unique backwards induction outcome is (**Pass**, **Pass**, **Pass**), but we know from Fudenberg and Levine (2006) that the outcome (**Drop**, **Drop**, **Pass**) is also patiently stable: in the steady state, the P2s play so rarely that they choose not to experiment with **Pass** and so never learn that the P3s **Pass**. But this outcome is ruled out when agents play the game on the right.

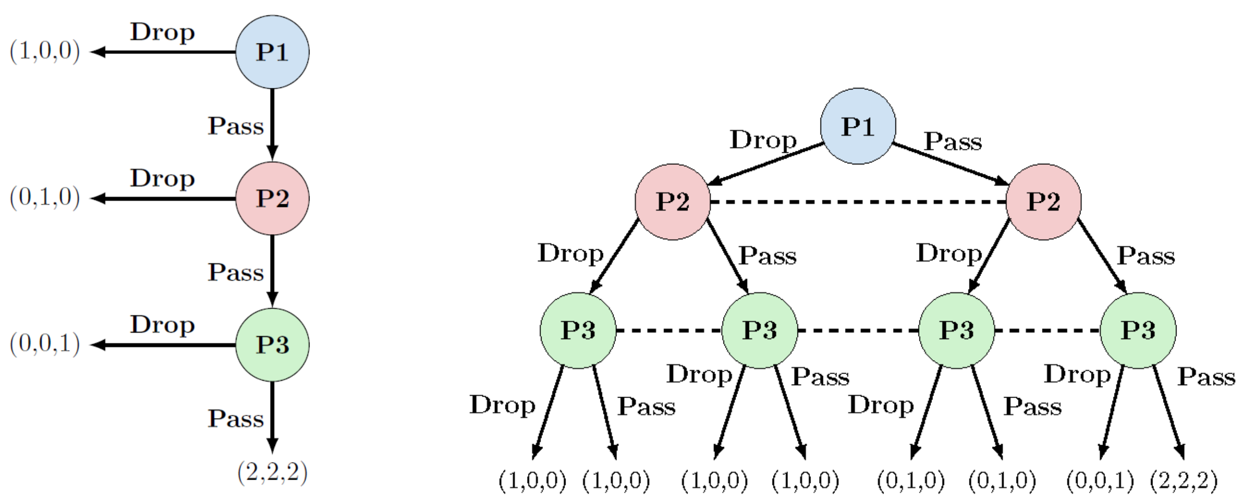


Figure 3: The two games have the same normal form, and (**Pass**, **Pass**, **Pass**) is the unique backward-induction profile for the extensive form on the left. (**Drop**, **Drop**, **Pass**) is patiently stable for the game on the left, but not for the game on the right.

Claim 3. *Suppose agents play the extensive form on the right of Figure 3. Then the only patiently stable profile is (**Pass**, **Pass**, **Pass**).*

This claims follows from Proposition 4 which we discuss later in Section 7.1. In the game on the right of Figure 3, P3s always **Pass** because they have full-support beliefs about what others do. Unlike for the game on the left, now P2s do not need to experiment to learn this. Once P2s learn that P3s play **Pass**, they themselves play **Pass**. This means that, when agents are long-lived, the vast majority of P2s in the population play **Pass**, so P1s learn to play **Pass** over **Drop** as well.

While the predictions derived from learning are not invariant to all transformations that preserve the reduced normal form, some of these transformations do leave the predictions of learning models unchanged, including transformations that do not preserve the selections made by sequential equilibrium. These are the transformations are both *decision invariant* in that they lead to the same best responses as a function of opponent strategies, and *information invariant* in the sense of providing the same feedback to the agents in their learning problems. Figure 4 depicts two games related by such a transformation. The game on the right is obtained from the game on the left by splitting P1's decision node so that, rather than immediately choosing between *Out*, In_1, and In_2, P1 chooses between *Out* and *In* and then chooses between subsequent actions 1 and 2. When we identify all P1 strategies in the right-hand game that use *Out* with the strategy *Out* in the left-hand game and identify $(In, 1)$ with In_1 and $(In, 2)$ with In_2, then both games have the same set of patiently stable profiles, as shown in Claim 4 below. However, the outcome **Out** is only a sequential equilibrium outcome in the extensive form on the left.[5] Section 6 generalizes this example and relates the findings to the fundamental transformations that Elmes and Reny (1994) showed have no effect on the reduced normal form.

[5] In any sequential equilibrium of the game on the right, P1 must play action 2 so P2 must play **L**, so P1 must play **In**.

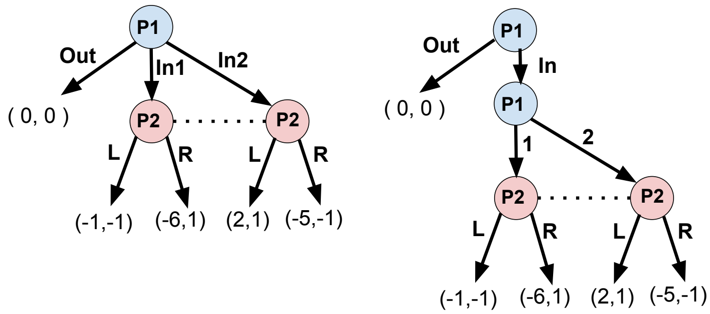


Figure 4: Both games have the same set of patiently stable profiles. But **Out** is only a sequential equilibrium outcome for the game on the left.

3.3 Information Equivalent Terminal Node Partitions

From a learning perspective, we can identify a given extensive form with its normal form when we augment the normal form with the appropriate terminal node partitions.[6] For example, the game in the left of Figure 3 has the same set of patiently stable profiles as the simultaneous-move game below, with the depicted terminal node partitions.

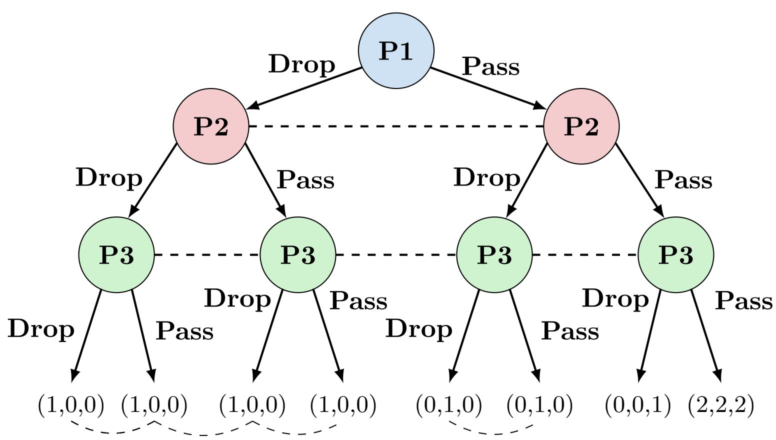


Figure 5: The game on the right of Figure 3 equipped with terminal node partitions. This game provides the same feedback to players as the game on the left of Figure 3.

The partition, which is common to all three players, says that if P1 plays **Drop**, players do not observe the choices of P2 and P3, and that if P1 plays **Pass** and P2

[6]Throughout the paper, when we refer to learning in a normal-form game, we mean learning in the equivalent simultaneous-move version of the normal-form game, where the players simultaneously choose actions, and the action space of each player is isomorphic to their strategy space in the original game.

plays **Drop** then they do not observe the choice of P3. Under this partition, (**Drop**, **Drop**, **Pass**) again becomes patiently stable. Section 7 discusses how the terminal node partitions influence which profiles are patiently stable.

4 The Learning Model

There is a unit mass population of agents who play each role $1 \leq i \leq I$ in a finite game of perfect recall. In every period, each agent is anonymously matched with opponents from the other populations uniformly at random to play the stage game. At the end of each play of the game, each agent observes the element of their *terminal node partition* $\mathcal{P}_i$ that contains the realized terminal node of the game, where we require that $u_i(z) = u_i(z')$ if z and z' are in the same cell of i's terminal node partition. The agent uses this information to update their beliefs about the distribution of play in opponent populations.

As in Fudenberg and He (2018) and Clark and Fudenberg (2021), we assume that the agents have geometrically distributed lifetimes: At the end of every period, each agent exits the system with probability $0 < 1 - \gamma \leq 1$, and a mass of newcomers is added to each population to replace the departing agents.[7] Agents maximize expected discounted utility, discounting future payoffs with a psychological discount factor $0 \leq \delta < 1$.

Denote the set of pure strategies of i in the game as $\mathbb{S}_i$ and the set of behavior strategies of i as Π_i. Agents believe that the aggregate distribution of play in the opponent population is constant, but they do not know what that distribution is. Each agent in population i starts with a prior belief $g_i \in \Delta(\times_{h \in \mathcal{H}_{-i}} \Delta(A_h))$ about the aggregate behavior strategy profile that describes play in opponent populations $j \neq i$ at different information sets. We assume that, for each i, the prior g_i is *non-doctrinaire*, meaning that it has a density which is strictly positive on the interior of $\times_{h \in \mathcal{H}_{-i}} \Delta(A_h)$.[8]

[7] Previous work by Fudenberg and Levine (1993) and Fudenberg and Levine (2006) assumed agents have fixed finite lifetimes. All of our results extend to this alternate lifetime specification.

[8] The strict positivity assumption lets us appeal to the classic Diaconis and Freedman (1990) result

As agents play the game and accumulate histories of past play and observations, they update their beliefs using Bayes' rule (which is always applicable because the priors assign positive probability to any finite sequence of observations) and modify their behavior. Let $Y_{i,t} = (\mathbb{S}_i \times \mathcal{P}_i)^t$ be the set of possible histories that can be observed by an i agent of age t. (By convention, $\Omega^0 = \emptyset$ for any set Ω.) Let $Y_i = \cup_{t \in \mathbb{N}} Y_{i,t}$ be the collection of all possible histories of agents from population i. We assume that all agents in each population i use the same optimal dynamic policy $\mathbf{s}_i^{g_i,\delta,\gamma} : Y_i \to \mathbb{S}_i$ that depends on their prior g_i, their discount factor δ, and their lifetime parameter γ.[9] (When the prior is held fixed and there is no risk of confusion, we sometimes omit the prior from our notation.)

In every period t, the *state* of the system, denoted $\mu_t = (\mu_{1,t}, ..., \mu_{I,t}) \in \times_i \Delta(Y_i)$, gives the shares of agents in the different player roles with the various possible histories. Given μ_t, the player i policy $\mathbf{s}_i^{g_i,\delta,\gamma}$ induces a player i behavior strategy $\sigma_i^{g_i,\delta,\gamma}(\mu_{i,t}) \in \Pi_i$ that we call the *aggregate strategy* of population i. We call $\sigma^{g,\delta,\gamma}(\mu_t) = (\sigma_i^{g_i,\delta,\gamma}(\mu_{i,t}))_i \in \times_i \Pi_i$ the *aggregate strategy profile.*[10]

A policy profile generates an *update rule* $\mathbf{f}^{g,\delta,\gamma} : \times_i \Delta(Y_i) \to \times_i \Delta(Y_i)$, taking the state in period t to the state in period $t+1$. It also generates, for each i, a mapping $\mathscr{R}_i^{g,\delta,\gamma} : \Pi_{-i} \to \Pi_i$ from aggregate play of all populations but i to the aggregate play of population i that gives the limit of the aggregate population i strategy when their opponent aggregate play is fixed at π_{-i} as the learning system of population i proceeds forward period after period.[11] We refer to the mapping $\mathscr{R}^{g,\delta,\gamma}(\pi) \equiv (\mathscr{R}_1^{g,\delta,\gamma}(\pi_{-1}), ..., \mathscr{R}_I^{g,\delta,\gamma}(\pi_{-I}))$ as the *aggregate response mapping.* Similar arguments to those in Clark and Fudenberg (2021) show that this mapping is contin-

on the rate of convergence of Bayesian posteriors to the empirical distribution. Note that if agents believe that they know their opponents' payoff functions, strict positivity requires that they assign positive probability to opponent strategies they believe are dominated. We discuss this issue more in the conclusion.

[9] This does not mean that they all play in the same way, as agents with the same policy may meet different opponents, and so have different histories and play different strategies.

[10] Formally, $\sigma^{g,\delta,\gamma}(\mu_t)[s_i] = \sum_{y_i \in Y_i \text{ s.t. } \mathbf{s}_i^{g_i,\delta,\gamma}(y_i) = s_i} \mu_{i,t}[y_i]$.

[11] Because the play of all populations but i is held fixed, and we have specified a policy for population i, this limit exists and does not depend on the initial distribution of population i histories.

uous.

To define $\mathscr{R}_i^{g,\delta,\gamma}$ more precisely, we first iteratively define a distribution $\mu_i \in \Delta(Y_i)$. First, assign $\mu_i(\emptyset) := 1-\gamma$. Once we have assigned a probability to each length t history in $Y_{i,t}$, write each $y_{i,t+1} \in Y_{i,t+1}$ as the concatenation of a one-period history with a t-period history, $y_{i,t+1} = (y_{i,t}, (s_{i,t+1}, P_{i,t+1}))$, where $s_{i,t+1} \in \mathbb{S}_i$ is the strategy that the agent used in the $(t+1)$-th match they played, and $P_{i,t+1} \in \mathcal{P}_i$ is the terminal node partition element observed in that match. Let $\mu_i(y_{i,t+1})$ be the survival probability times the mass of the agents with the history $y_{i,t}$, multiplied by the probability of the one-period history $(s_{i,t+1}, P_{i,t+1})$ when i plays according to the policy $\mathbf{s}_i^{g_i,\delta,\gamma}$ and opponents' play is drawn from π_{-i}.[12] This allows us to define μ_i on every element of $Y_{i,t+1}$, and iteratively we can define μ_i on all of Y_i. It is straightforward to verify that μ_i is a distribution on Y_i. The aggregate response $\mathscr{R}_i^{g,\delta,\gamma}(\pi_{-i})$ is defined to be $\sigma_i^{g_i,\delta,\gamma}(\mu_i)$, the aggregate strategy associated with μ_i.

The *steady states* of the learning model are the fixed points of $\mathbf{f}^{g,\delta,\gamma}$. We focus on the corresponding aggregate strategy profiles — that is, the $\sigma^{g,\delta,\gamma}(\mu)$ where μ is a steady state — the *steady state profiles*, and denote them by $\Pi^*(g,\delta,\gamma) \subseteq \times_{1\leq i \leq I}\Pi_i$. Again, similar arguments to those in Clark and Fudenberg (2021) show that these are the fixed points of the aggregate response mapping. Continuity of the aggregate response mapping, along with Brouwer's fixed point theorem, then implies that steady state profiles always exist.

Proposition 1. $\Pi^*(g,\delta,\gamma)$ *consists of the strategy profiles that are fixed points of the aggregate response mapping, and it is non-empty for all g, δ, and γ.*

When the agents are short-lived they have little chance to learn, and simply play a best response to their priors. When agents are long-lived but impatient, they do learn the steady state path of play, but need not learn how opponents respond to deviations, so any self-confirming equilibrium in strategies that are not weakly dominated could

[12] If $s_{i,t+1} \neq \mathbf{s}_i^{g_i,\delta,\gamma}(y_{i,t})$, set $\mu_i(y_{i,t+1}) := 0$. Otherwise, $\mu_i(y_{i,t+1}) := \mu_i(y_{i,t}) \cdot \gamma \cdot \mathbb{P}[P_{i,t+1} \mid s_{i,t+1}, \pi_{-i}]$, where $\mathbb{P}[P_{i,t+1} \mid s_{i,t+1}, \pi_{-i}]$ refers to the probability of reaching the terminal node partition element $P_{i,t+1}$ when i uses the strategy $s_{i,t+1}$ and $-i$'s strategy is drawn from the distribution π_{-i}.

arise. We will focus on steady states where agents are both long-lived and patient. More specifically, we focus on steady state profiles in the limit where agents become long lived ($\gamma \to 1$) and patient ($\delta \to 1$). Moreover, following Fudenberg and Levine (1993) and Fudenberg and Levine (2006), Fudenberg and He (2018), and Clark and Fudenberg (2021), we assume that the continuation probability γ converges to 1 faster than δ. We call the strategy profiles that can emerge in this limit the *patiently stable strategy profiles* and the distributions over terminal nodes generated by patiently stable strategy profiles *patiently stable outcomes*. The order of limits corresponds to an environment where agents are long-lived relative to their effective discount factors. This implies that people spend most of their lives myopically responding to their current beliefs.

Definition 1. *Strategy profile* π *is* **patiently stable** *if there are sequences* $\{\delta_j\}_{j\in\mathbb{N}}$, $\{\gamma_{j,k}\}_{j,k\in\mathbb{N}}$ *and associated steady-state profiles* $\{\pi_{j,k} \in \Pi^*(g, \delta_j, \gamma_{j,k})\}_{j,k\in\mathbb{N}}$ *such that* $\lim_{j\to\infty} \delta_j = 1$, $\lim_{k\to\infty} \gamma_{j,k} = 1$ *for each* j *and* $\lim_{j\to\infty} \lim_{k\to\infty} \pi_{j,k} = \pi$.

The literature has previously shown that patiently stable strategy profiles must be Nash equilibria when agents observe the realized terminal nodes in the games they play.[13]

Appendix A.2 shows that this is also true for the game and terminal node partition given in Figure 2, which is the only example in the paper that uses a non-discrete terminal node partition to exhibit a patiently stable profile that is ruled out by classic refinements. We conjecture that patiently stable profiles must be Nash equilibria in any game provided each agent's payoff is measurable with respect to their terminal node partition, but we have not shown this. Instead, Appendix A.3 gives a number of other examples from the literature where this conclusion does hold.

[13]Fudenberg and Levine (1993) established this in a learning model where players had fixed finite lifetimes rather than geometric lifespans. The adaptations of these arguments given in the supplementary information of Fudenberg and He (2018) show that this extends to geometric lifespans in general games, although the main text of Fudenberg and He (2018) only states this result for signaling games.

5 Patient Stability, Forward Induction, and Dominance

5.1 Dominated Actions in a Family of Two-Player Games

This section provides a sufficient condition for patient stability that generalizes the example from Section 3.1.1. We consider a family of two-player games where P1 first chooses an action $a_1 \in A_1$, which may end the game or give the play to P2. For each P2 information set h_2, P2 chooses among the actions $\mathcal{A}_2(h_2)$, and we let $\rho(h_2)$ denote the P1 actions that lead to h_2. Write $u_i(a_1, a_2)$ for i's utility at the terminal node reached by P1 playing a_1 and P2 playing a_2. We also write $u_i(\pi_1, \pi_2)$ for i's expected utility when players use behavior strategies π_1 and π_2.

We will show that equilibria of the following form are patiently stable under some non-doctrinaire prior that we construct.

1. P1 plays a single action $a_1^* \in A_1$ that uniquely maximizes their payoff given P2's strategy. (Formally, $\pi_1^*(a_1^*) = 1$ for the a_1^* that satisfies $u_1(a_1^*, \pi_2^*) > u_1(a_1, \pi_2^*)$ for all $a_1 \neq a_1^*$.)
2. For each P2 information set h_2, P2 plays some response $a_2^*(h_2)$ that is optimal given some $a_1^*(h_2) \in \rho(h_2)$. Moreover, out of $\rho(h_2)$, $a_1^*(h_2)$ is optimal for P1 given that P2 plays $a_2^*(h_2)$.
3. If a_1^* leads to P2 information set h_2^*, then $a_2^*(h_2^*)$ uniquely maximizes P2's payoff against a_1^*.

The $(\mathbf{Out}, \mathbf{R})$ equilibrium from Section 3.1.1 is of this form: **Out** serves the role of a_1^*, and for P2's only information set, P2's prescribed response of **R** is the unique best response to **In1**. In turn, **In1** is the best action out of $\{\mathbf{In1}, \mathbf{In2}\}$ for P1 when P2 chooses **R**.

In the equilibria we construct, P2 may best reply to dominated P1 actions $a_1^*(h_2)$ at some information sets h_2. Nevertheless, we show in Proposition 2 below that every such equilibrium is patiently stable, which implies Claim 1.

The key is to choose priors that are "supportive" of the equilibrium.

Definition 2. *Priors g_1 and g_2 are* **supportive** *priors for π^* if, for every off-path P2 information set h_2, (1) $\mathbb{E}_{g_1}[u_1(a_1^*(h_2), a_2(h_2))|y_1] \geq \mathbb{E}_{g_1}[u_1(a_1, a_2(h_2))|y_1]$ for all $a_1 \in \rho(h_2)$ and P1 histories y_1 that have never recorded a P2 agent play some action other than $a_2^*(h_2)$ at h_2, and (2) $\mathbb{E}_{g_2}[u_2(a_1, a_2^*(h_2))|y_2, h_2] \geq \mathbb{E}_{g_2}[u_2(a_1, a_2))|y_2, h_2]$ for all $a_2 \in \mathcal{A}_2(h_2)$ and histories y_2 that have never recorded a P1 agent play any $a_1 \in \rho(h_2)\backslash\{a_1^*(h_2)\}$.*

A supportive P1 prior is such that, for every off-path α_1, a P1 agent prefers to experiment with $a_1^*(\alpha_1)$ over any other action in α_1 unless they have previously experienced a P2 response to α_1 for which $a_1^*(\alpha_1)$ is not conditionally optimal. Similarly, a supportive P2 prior leads P2 agents to want to respond to α_1 with $a_2^*(\alpha_1)$ unless they have previously witnessed a P1 agent play some action in α_1 other than $a_1^*(\alpha_1)$. These properties facilitate the proof of the following proposition, which is given in Appendix A.4.

Proposition 2. *Suppose that π^* is an equilibrium of the form given above. Then π^* is patiently stable for any pair of non-doctrinaire P1 and P2 priors that are supportive of π^*.*

5.2 Stability and Doubly Dominated Actions: An Example

The example from Section 3.1.2 does not fit with the sufficient conditions for stability we gave in Section 5.1: it involves P3 best replying to the action **In2** by P2, a doubly dominated action that is not optimal among the P2 actions {**In1**,**In2**} that reach the same P3 information set. We use a different argument to show that the $(\mathbf{Out}, \mathbf{Out}, \mathbf{R})$ outcome is patiently stable.

Proposition 3. *For the game in Figure 2, (**Out**, **Out**, **R**) is a patiently stable profile for any non-doctrinaire P1 prior g_1, non-doctrinaire P2 prior g_2 under which the expected probability of **L** is strictly less than $1/2$, and non-doctrinaire P3 prior g_3 that leads a P3 agent to play **L** only when they have previously observed a P2 agent play **In1**.*

This proposition specifies the prior beliefs that make patient stability hold in Claim 2. The proof of this result in Appendix A.5 first notes that P1 observes P3's play if and only if they experiment with **In**. This lets us bound the number of periods that P1s will typically experiment with **In** before becoming pessimistic and switching to **Out** forever in a steady state where P3s play **R** with high enough probability, so most P2 agents will learn that their information set is rarely reached. Thus they will choose **Out** instead of experimenting with **In1**, since they do not value information they will rarely get to use. This lets us construct a steady state where most P3 agents have never observed any instance of matched P2 agents choosing any action other than **Out**, and therefore choose **R** based on their prior belief.

6 Invariance Under Patient Stability

As discussed in Section 3.2, predictions derived from learning are invariant under transformations that are both decision invariant and information invariant. Here we study which of the transformations that Elmes and Reny (1994) identified as preserving the reduced normal form of a game respect these conditions. Elmes and Reny (1994) showed that two finite games with perfect recall have the same reduced normal form if they can be transformed into the same extensive form under finite sequences of these three transformations: *Add* (ADD), which corresponds to the addition of decision nodes to an existing information set while preserving the overall structure of the game, *Coalesce* (COA), which reduces two consecutive moves by a single player into one move, and *Interchange* (INT), which changes the order of play between players when they do not know the choices made by each other.

ADD does not satisfy information invariance. For the example depicted in Figure 3, the extensive form on the right can be obtained from the extensive form on the left through repeated use of the ADD transformation. However, playing Pass in the game on the left prevents P1 from observing the play of their opponents, whereas this is not the case for the game on the right.

However, both COA and INT satisfy decision invariance and information invariance, and as the next result shows they lead to the same predictions under learning models.

Claim 4. *If $\hat{\mathcal{G}}$ can be obtained by applying either COA or INT to $\mathcal{G}$, then $\mathcal{G}$ and $\hat{\mathcal{G}}$ have the same set of patiently stable outcomes.*

Appendix A.1 gives the proof of the claim for COA. (Similar ideas handle the proof for INT.) Handling the coalescence of two information sets of i are coalesced requires modifying the domain of $-i$'s prior beliefs about i's play so that they are about i's single action at the combined information set. The proof establishes a bijection between non-doctrinaire prior densities g in the game $\mathcal{G}$ and $\hat{g}$ in the game $\hat{\mathcal{G}}$, so that the set of steady states are the same under g in $\mathcal{G}$ and $\hat{g}$ in the $\hat{\mathcal{G}}$ for any $0 \leq \delta, \gamma < 1$ (up to identifying histories of play at the two consecutive information sets in $\mathcal{G}$ with the corresponding histories of play at the single coalesced information set in $\hat{\mathcal{G}}$).

7 Observability and Patiently Stable Profiles

In this section, we study the effect of what agents observe at the end of each play of the game on the patiently stable profiles. Sections 7.1 and 7.2 show that, when agents play the normal form derived from a simple game and the terminal node partition is discrete, patiently stable profiles must select the same outcome as the backward induction outcome of the original game. Section 7.3 says that if the (extensive-form representation of the) normal form of an extensive form is equipped with the right terminal node partitions, it leads to the same patiently stable profiles as the extensive form. Section 7.4 provides an example where patiently stable profiles satisfy the iterated deletion of weakly dominated strategies with coarser observations but not finer ones.

7.1 Backward Induction in Simple Games when Agents Observe Strategies

A *simple game* is an extensive-form game of perfect information where no one moves more than once along any path and no player is indifferent between any two terminal nodes, so there exists a unique backward induction strategy profile.

The next result shows that the only patiently stable profile of the normal form of a simple game with discrete terminal node partitions is the backward induction strategy profile. In fact, we show something stronger: this is the only profile that is δ-*stable*.

Definition 3. *For* $0 \le \delta < 1$ *and non-doctrinaire prior* g*, strategy profile* π *is* δ**-stable** *under* g *if there is a collection of parameter sequences* $\{\gamma_k\}_{k\in\mathbb{N}}$ *and associated steady-state profiles* $\{\pi_k \in \Pi^*(g,\delta,\gamma_k)\}_{k\in\mathbb{N}}$ *such that* $\lim_{k\to\infty}\gamma_k = 1$ *and* $\lim_{k\to\infty}\pi_k = \pi$*.*

Proposition 4. *Suppose agents play the normal-form representation of a simple game. Then, every* δ*-stable profile puts probability* 1 *on a backward-induction outcome.*

In particular, this implies that for the right-hand game of Figure 3, (**Pass**, **Pass**, **Pass**) is the only learning outcome when agents are sufficiently long lived. As we show in Appendix A.7, Proposition 4 follows from a more general result in the next section about patient stability for normal-form games.

7.2 An Iterative Deletion Refinement in Normal Forms

The next proposition discusses the implications of patient stability in environments of "maximal observability": that is, agents play the normal form derived from an extensive-form game with discrete terminal node partitions. This result gives us a benchmark of what long-lived agents will learn in games if they do not need to experiment. The result takes the form of an iterative procedure that eliminates at each step some of the remaining strategies that do not best respond to strictly mixed conjectures that put arbitrarily low conditional probabilities on eliminated opponent strategies. Let $\mathcal{S} = \{\times_i \widetilde{S}_i : \forall i,\ \widetilde{S}_i \subseteq S_i\}$ be the set of product spaces generated by the subsets of the player strategy spaces.

Definition 4. *A sequence* $(S^{(0)}, D^{(0)}), (S^{(1)}, D^{(1)})... \in \mathcal{S}^2$ *is a* **valid elimination sequence** *if*

1. *For each* i*,* $S_i^{(0)} = S_i \setminus D_i^{(0)}$*, and* $D_i^{(0)}$ *is any subset of* i*'s weakly dominated strategies,*

2. *For each i and $m > 0$, $D_i^{(m)}$ is a subset of $S_i^{(m-1)}$ such that, for every $s_i \in D_i^{(m)}$, there exists some $\epsilon > 0$ where $\mathbb{E}_{\sigma_{-i}}[u_i(s_i, s_{-i})] < \max_{s_i' \in S_i} \mathbb{E}_{\sigma_{-i}}[u_i(s_i', s_{-i})]$ for all correlated opponent strategy profiles $\sigma_{-i} \in \Delta(S_{-i})$ satisfying $\sigma_{-i}(S_j^{(m-1)}|s_{-ij}) \geq 1 - \epsilon$ for every $j \neq i$ and $s_{-ij} \in S_{-ij}$, and*
3. *For each i and $m > 0$, $S_i^{(m)} = S_i^{(m-1)} \backslash D_i^{(m)}$.*

In a valid elimination sequence, at every stage of the iteration, the only player i strategies that can be eliminated are those for which the following condition holds: There is an $\epsilon > 0$ such that the strategy is suboptimal under any conjecture that, for each opponent j, puts probability at least $1 - \epsilon$ on j strategies that have not yet been eliminated conditional on any strategy profile of the opponents other than j.

Proposition 5. *For a valid elimination sequence $(S^{(0)}, D^{(0)}), (S^{(1)}, D^{(1)})... \in \mathcal{S}^2$, let $S_i^* = \cap_{m=0}^{\infty} S_i^{(m)}$. If agents observe matched opponents' strategy choices at the end of each game, then every δ-stable strategy profile is supported on the non-empty set $\times_i S_i^*$.*

The idea behind the proof is that agents never use weakly dominated strategies in $D_i^{(0)}$ because they have full-support beliefs about others' play, and experienced agents learn that these strategies are rarely used by an extension of Diaconis and Freedman (1990)'s result in Fudenberg, Lanzani, and Strack (2021). This implies strategies in $D_i^{(1)}$ only get used with very low probabilities in the steady state, as they are only played by the very young agents. Iterating this argument lets us eliminate the strategies in $D_i^{(2)}, D_i^{(3)}$, and so forth.

Different valid elimination sequences may lead to different strategy sets S_i^* in the end. Proposition 5, which we prove in Appendix A.6 gives a family of necessary conditions of patient stability, corresponding to different valid elimination sequences.

Some of the valid elimination sequences correspond to well-known solution concepts. One example is backward induction in simple games: Proposition 4 follows from Proposition 5 by letting $D_i^{(m)}$ be those extensive-form strategies of i that are inconsistent with backward induction at some decision node $m + 1$ steps away from the terminal nodes, but agree with it at all decision nodes m or fewer steps away from the terminal nodes. The proof of Proposition 4 verifies that these $D_i^{(m)}$ form a valid elimination sequence.

A second example is the solution concept $S^\infty W$ (Dekel and Fudenberg (1990)), which Börgers (1994) shows is equivalent to players having full support beliefs about the play of others and that this and the rationality of the players are "almost common knowledge." This solution concept corresponds to choosing $D_i^{(0)}$ to be all weakly dominated strategies of i in the original game, and, at each step m, choosing $D_i^{(m+1)} \subseteq S_i^{(m)}$ to be the strictly dominated strategies of i in the reduced game where i has the strategy set $S_i^{(m)}$. To see that this is a valid elimination sequence, note that if s_i is strictly dominated, then there is some $\sigma_i \in \Delta(S_i^{(m)})$ and $\eta > 0$ so that $u_i(\sigma_i, s_{-i}) > u_i(s_i, s_{-i}) + \eta$ for all $s_{-i} \in S_{-i}^{(m)}$. By continuity, there exists some $\epsilon > 0$ such that for any full-support correlated opponent strategy σ_{-i} of the original game where $\sigma_{-i}(S_{-i}^{(m)}) \geq 1-\epsilon$, we have $u_i(\sigma_i, \sigma_{-i}) > u_i(s_i, \sigma_{-i}) + \eta/2$, so in particular s_i is not a best response to any such σ_{-i}.

While the refinement in Proposition 5 is stronger than $S^\infty W$, it is weaker than iterated elimination of weakly dominated strategies. This is because in defining $D_i^{(m)}$ in the iterative procedure, we consider conjectures where the probabilities assigned to deleted strategies can be arbitrarily small, but need not be zero. Provided there are at least two remaining strategies, this does not imply that the highest probability assigned to a deleted strategy must be lower than the lowest probability assigned to a remaining strategy. This distinguishes the Proposition 5 refinement from other refinement concepts like the iterated admissibility of Brandenburger, Friedenberg, and Keisler (2008) and the consistent pairs of Börgers and Samuelson (1992).[14] For instance, for the game in Figure 6, there is no valid elimination sequence that uniquely selects the (**A**, **X**) strategy profile, even though (**A**, **X**) is the unique iteratively admissible profile. From a learning perspective, the idea is that although **C** is strictly dominated for P1, if P1 always play **B** then P2 can still maintain a belief that **C** is relatively more

[14] Consistent pairs capture the implications of assuming that players maximize expected utility, and that players form "cautious expectations." Such pairs are only defined for two-player games, and do not always exist.

	X	**Y**
A	2, 2	0, 0
B	1, 1	1, 1
C	-10, 0	-10, 1

Figure 6: In this game, (**A**, **X**) is the unique iteratively admissible outcome (Brandenburger, Friedenberg, and Keisler, 2008), but (**B**, **Y**) is also patiently stable.

likely than **A** and thus choose **Y**. Indeed, it is easy to see that (**B**, **Y**) is a steady-state profile for any $0 \leq \delta, \gamma < 1$ (and therefore, patiently stable) if P1 starts with a strong prior belief that P2s play **Y** and P2s start with a Dirichlet prior with weights (1, 1, 10) on the P1 actions (**A**, **B**, **C**).

7.3 Information-Equivalent Normal Forms

For an extensive form $\mathcal{G}$, consider an extensive-form representation $\mathcal{N}(\mathcal{G})$ of the normal form of $\mathcal{G}$, that is a game where the players simultaneously choose actions, and the action space of each player is isomorphic to their strategy space in $\mathcal{G}$.[15] Given a terminal node partition $\mathcal{P}$ of the game $\mathcal{G}$, learning in $\mathcal{G}$ with the terminal node partition $\mathcal{P}$ and learning in $\mathcal{N}(\mathcal{G})$ with the usual discrete terminal node partitions (so that players observe one another's choice of pure strategy) can lead to different patiently stable profiles, as shown above. However, when $\mathcal{N}(\mathcal{G})$ is equipped with the appropriate terminal node partitions, it will have the same set of patiently stable profiles as $\mathcal{G}$.

The $\mathcal{P}-$*information equivalent* terminal node partitions are $\hat{\mathcal{P}}_i$ for i in $\mathcal{N}$ are such that $\hat{\mathcal{P}}_i(s) = \hat{\mathcal{P}}_i(s')$ if and only if $\mathcal{P}_i(\mathbf{z}(s)) = \mathcal{P}_i(\mathbf{z}(s'))$. Players hold beliefs over opponents' behavior strategies in $\mathcal{G}$ and mixed strategies in $\mathcal{N}$, but we can transform a non-doctrinaire belief over behavior strategies into one over mixed strategies and vice versa when $\mathcal{G}$ has perfect recall, by Kuhn's theorem.

Proposition 6. *The patiently stable profiles of* $(\mathcal{G}, \mathcal{P})$ *are the same as the patiently stable profiles of* $\mathcal{N}$ *with the* $\mathcal{P}-$*equivalent terminal node partitions.*

[15] As usual with simultaneous-move games, the order of the players does not matter.

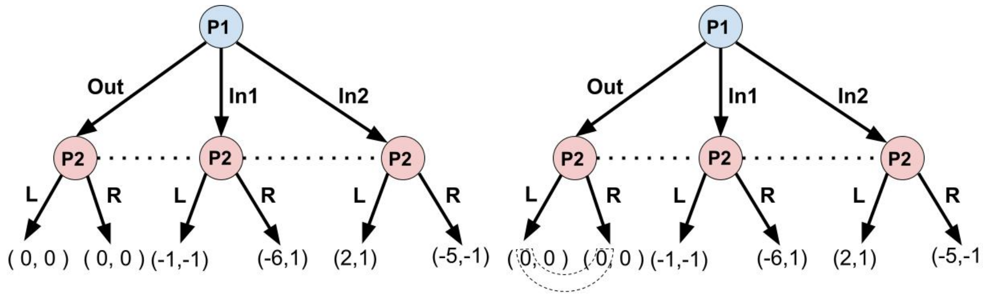


Figure 7: In the game on the left, both players observe the terminal node. In the game on the right, P1 does not observe P2's play if they choose **Out**. The (**Out**, **R**) profile is patiently stable for the game on the left, but not for the game on the right.

Intuitively, the definition of $\hat{\mathcal{P}}_i$ implies agents have the same feedback in the two games, so the problems are information invariant as well as decision invariant. We formally show this in Appendix A.8.[16]

7.4 Coarser Terminal Partitions May Eliminate Patiently Stable Profiles

Sections 7.1 and 7.2 show that coarser observations of opponents' strategies can expand the set of patiently stable profiles. But this is not always true, and coarser terminal node partitions can shrink rather than expand the set of patiently stable profiles in other games.

Consider the two games in Figure 7 that only differ in the terminal node partition of P1. In the game on the left with the finer terminal node partitions, (**Out**, **R**) is patiently stable. It is easy to see that if P1 and P2 both start with a strong prior belief in the (**Out**, **R**) equilibrium and P2 thinks **In1** is more likely than **In2** when they have only seen P1s play **Out**, then it is a steady state under any $0 \leq \delta, \gamma < 1$ for (**Out**, **R**) to be played in every match.

But, (**Out**, **R**) is not patiently stable in the game on the right with the coarser terminal node partitions, as we show in Claim 5.[17] The proof idea, which we rigorously

[16]Note that unlike the "normal form information sets" of Mailath, Samuelson, and Swinkels (1993), the equivalent terminal node partition cannot be derived from the normal form alone.

[17]Technically, Claim 5 imposes additional restrictions on the P2 prior, but the class of priors allowed is broad and includes those with densities that are strictly positive and continuous everywhere as well as Dirichlet distributions.

demonstrate in Appendix A.9, is that patient P1 players will spend many periods experimenting with **In2**, since they cannot learn P2's play by choosing **Out**. This teaches P2s that P1s are much more likely to use **In2** than **In1**, so that they should not play **R**.

Claim 5. *For the game on the right in Figure 7, suppose P2's prior belief satisfies Condition $\mathcal{P}$ from Fudenberg, He, and Imhof (2017). Then, every patiently stable profile satisfies* $\pi(R) = 0$.

Note that **In1** is strictly dominated for P1, and if P2 thinks that P1 never plays **In1**, then **L** is strictly better than **R** for P2 given any conjecture that puts positive probabilities on both **Out** and **In2**. Thus (**Out**, **R**) is ruled out by iterated elimination of weakly dominated strategies, and stable learning outcomes in the example violate this refinement with a finer terminal node partition, but not with a coarser one. Combined with the results of Sections 6.1 and 6.2, this illustrates how the details of the learning environment matter for the effects of various sorts of transformations of the game. In particular, there is no reason to expect predictions about a game's outcome to be invariant to transformations of the game unless the transformations preserve information invariance.

8 Conclusion

The implications of learning depend crucially on the structure of the game and on what agents observe about others' play. When the game and the feedback structure make it profitable for patient players to experiment with dominated strategies (for instance, when agents get no information from choosing a safe action but can use a worse safe action to learn about the consequences of a risky action), patiently stable profiles may violate forward induction or iterated weak dominance. When agents must experiment to learn about off-path play, patiently stable profiles may violate backward induction. But if agents observe opponents' strategies regardless of their own play, patiently stable profiles always satisfy backward induction in simple games. This shows that ruling out some Nash equilibria requires close attention to the details of the game and the learning environment.

As in previous work, we have assumed that agents have non-doctrinaire priors in order to appeal to the Diaconis and Freedman (1990) result on the speed of convergence of Bayesian posteriors to the empirical distribution. Fudenberg, Lanzani, and Strack (2021) extends their convergence result to priors without full support, but if the true state is outside of the support of the priors then agents need not stop experimenting in finite time, as shown by Fudenberg, Romanyuk, and Strack (2017). This raises a suite of new issues, as patient stability might not imply Nash equilibrium.

Also, the assumption that agents have non-doctrinaire priors over aggregate play in the other populations rules out settings where agents place probability 0 on opponent strategies that they believe are strictly dominated. Since much of the refinements literature implicitly assumes all players know the payoff functions of the others, it is natural to wonder if adding some forms of restrictions on the priors would bring the predictions of patient stability closer to those of classic equilibrium refinements. In the case of signaling games with independent priors, Fudenberg and He (2020) shows that the answer is "yes," but the implications of payoff information in general games are unclear. One issue is that, as we have seen, agents may choose to use dominated strategies for their information value, and an agent whose prior gave these strategies probability 0 would be unable to form a Bayesian posterior.[18] Of course, this problem does not arise with myopic players, for they will never pay a current cost to obtain information. But with myopic players there is no reason to expect learning to lead to a Nash equilibrium, let alone a refinement of it.

[18] This problem does not arise in signaling games with independent priors, as there the senders would never experiment with dominated strategies, and receivers never experiment at all.

References

Banks, J. S. and J. Sobel (1987). "Equilibrium selection in signaling games." *Econometrica*, 55, 647–661.

Börgers, T. (1994). "Weak dominance and approximate common knowledge." *Journal of Economic Theory*, 64, 265–276.

Börgers, T. and L. Samuelson (1992). ""Cautious" utility maximization and iterated weak dominance." *International Journal of Game Theory*, 21, 13–25.

Brandenburger, A., A. Friedenberg, and H. J. Keisler (2008). "Admissibility in games." *Econometrica*, 76, 307–352.

Cho, I.-K. and D. M. Kreps (1987). "Signaling games and stable equilibria." *Quarterly Journal of Economics*, 102, 179–221.

Clark, D. and D. Fudenberg (2021). "Justified communication equilibrium." *American Economic Review*, 111, 3004–34.

Dekel, E. and D. Fudenberg (1990). "Rational behavior with payoff uncertainty." *Journal of Economic Theory*, 52, 243–267.

Diaconis, P. and D. Freedman (1990). "On the uniform consistency of Bayes estimates for multinomial probabilities." *Annals of Statistics*, 18, 1317–1327.

Elmes, S. and P. J. Reny (1994). "On the strategic equivalence of extensive form games." *Journal of Economic Theory*, 62, 1–23.

Esponda, I. (2013). "Rationalizable conjectural equilibrium: A framework for robust predictions." *Theoretical Economics*, 8, 467–501.

Fudenberg, D. and D. M. Kreps (1995). "Learning in Extensive-Form Games I. Self-Confirming Equilibria." *Games and Economic Behavior*, 8, 20–55.

Fudenberg, D. and K. He (2018). "Learning and type compatibility in signaling games." *Econometrica*, 86, 1215–1255.

— (2020). "Payoff information and learning in signaling games." *Games and Economic Behavior*, 120, 96–120.

— (2021). "Player-compatible learning and player-compatible equilibrium." *Journal of Economic Theory*, 194, 105238.

Fudenberg, D., K. He, and L. A. Imhof (2017). "Bayesian posteriors for arbitrarily rare events." *Proceedings of the National Academy of Sciences*, 114, 4925–4929.

Fudenberg, D. and Y. Kamada (2015). "Rationalizable partition-confirmed equilibrium." *Theoretical Economics*, 10, 775–806.

— (2018). "Rationalizable partition-confirmed equilibrium with heterogeneous beliefs." *Games and Economic Behavior*, 109, 364–381.

Fudenberg, D., G. Lanzani, and P. Strack (2021). "Pathwise concentration bounds for Bayesian beliefs." *Working Paper*.

Fudenberg, D. and D. K. Levine (1993). "Steady state learning and Nash equilibrium." *Econometrica*, 61, 547–573.

— (2006). "Superstition and rational learning." *American Economic Review*, 96, 630–651.

Fudenberg, D., G. Romanyuk, and P. Strack (2017). "Active learning with a misspecified prior." *Theoretical Economics*, 12, 1155–1189.

Govindan, S. and R. Wilson (2009). "On forward induction." *Econometrica*, 77, 1–28.

Kohlberg, E. and J.-F. Mertens (1986). "On the strategic stability of equilibria." *Econometrica*, 54, 1003–1037.

Mailath, G. J., L. Samuelson, and J. M. Swinkels (1993). "Extensive form reasoning in normal form games." *Econometrica*, 61, 273–302.

Rubinstein, A. and A. Wolinsky (1994). "Rationalizable conjectural equilibrium: between Nash and rationalizability." *Games and Economic Behavior*, 6, 299–311.

Appendix

A Omitted Proofs

A.1 Proof of Claim 4

We prove the claim for the case where $\hat{\mathcal{G}}$ can be obtained by applying COA to $\mathcal{G}$. The argument for the case where it can be obtained by applying INT is similar.

Suppose extensive form $\hat{\mathcal{G}}$ is obtained by coalescing two consecutive information sets $h_i^{'}$ and $h_i^{''}$ of i in $\mathcal{G}$ into one information set $h_i^{\star}$ in $\hat{\mathcal{G}}$ (according to Elmes and Reny (1994)'s "COA" definition of coalescence). Suppose $A_{h_i^{'}} = \{a_1, ..., a_m, a_{\text{pass}}\}$ with the action a_{pass} leading to h_i'', $A_{h_i''} = \{a_{m+1}, ..., a_{m+n}\}$, and $A_{h_i^{\star}} = \{a_1, ..., a_m, a_{m+1}, ..., a_{m+n}\}$. Let $\Delta^{\circ}(X)$ the distributions on X that assign strictly positive probability to each very element in X. We define $\phi : \Delta^{\circ}(A_{h_i^{'}}) \times \Delta^{\circ}(A_{h_i''}) \to \Delta^{\circ}(A_{h_i^{\star}})$, such that $\phi(\alpha_{h_i^{'}}, \alpha_{h_i^{''}})(a_k) = \alpha_{h_i'}(a_k)$ for $1 \leq k \leq m$, while $\phi(\alpha_{h_i^{'}}, \alpha_{h_i^{''}})(a_k) = \alpha_{h_i'}(a_{\text{pass}}) \cdot \alpha_{h_i''}(a_k)$ for $m+1 \leq k \leq m+n$. That is, $\phi(\alpha_{h_i^{'}}, \alpha_{h_i^{''}})$ is a way to choose an element of $A_{h_i^{\star}}$ by using $\alpha_{h_i^{'}}$ and $\alpha_{h_i^{''}}$ sequentially: first draw an element from $A_{h_i^{'}}$ according to $\alpha_{h_i^{'}}$ and then, if the chosen element is a_{pass}, draw an element from $A_{h_i^{''}}$ according to $\alpha_{h_i^{''}}$. The map ϕ is one-to-one, because $\phi(\alpha_{h_i^{'}}, \alpha_{h_i''})$ and $\phi(\beta_{h_i^{'}}, \beta_{h_i''})$ generate different distributions on $\{a_1, ..., a_m)$ if $\alpha_{h_i^{'}} \neq \beta_{h_i^{'}}$, while $\phi(\alpha_{h_i^{'}}, \alpha_{h_i^{''}})$ and $\phi(\alpha_{h_i^{'}}, \beta_{h_i^{''}})$ generate different distributions on $\{a_{m+1}, ..., a_{m+n}\}$ if $\alpha_{h_i^{''}} \neq \beta_{h_i^{''}}$ and $\alpha_{h_i^{'}}$ a positive probability to a_{pass}. Also, ϕ is onto, because for a given $\alpha_{h_i^{\star}} \in \Delta^{\circ}(A_{h_i^{\star}})$, let $\alpha_{h_i'} \in \Delta^{\circ}(A_{h_i^{'}})$ be such that $\alpha_{h_i'}(a_k) = \alpha_{h_i^{\star}}(a_k)$ for $1 \leq k \leq m$, $\alpha_{h_i'}(a_{\text{pass}}) = 1 - \sum_{k=1}^{m} \alpha_{h_i^{\star}}(a_k)$, and $\alpha_{h_i''}(a_k) = \frac{\alpha_{h_i^{\star}}(a_k)}{\sum_{j=m+1}^{m+n} \alpha_{h_i^{\star}}(a_j)}$ for $m+1 \leq k \leq m+n$. It is clear that by construction, $\phi(\alpha_{h_i'}, \alpha_{h_i''}) = \alpha_{h_i^{\star}}$. We have $\phi(\alpha_{h_i^{'}}, \alpha_{h_i^{''}}) = \alpha_{h_i^{\star}}$ if and only if $(\alpha_{h_i^{'}}, \alpha_{h_i^{''}})$ and $\alpha_{h_i^{\star}}$ generate the same choice probabilities over the final actions $\{a_1, ..., a_m, a_{m+1}, ..., a_{m+n}\}$.

For each agent j in game $\mathcal{G}$, let $g_j : \times_{h \in \mathcal{H}_{-j}} \Delta(A_h) \to \mathbb{R}_+$ be j's prior prior density over $-j$'s strategies. Let $\hat{\mathcal{H}}_i$ represent i's information sets in $\hat{\mathcal{G}}$, and continue to use $\mathcal{H}_j$ for j's information sets in $\hat{\mathcal{G}}$ for agents $j \neq i$. Let $\hat{g}_j : \times_{h \in \hat{\mathcal{H}}_{-j}} \Delta(A_h) \to \mathbb{R}_+$ be a density of j's belief about $-j$'s play in $\hat{\mathcal{G}}$ such that (1) if $j \neq i$, then $\hat{g}_j(\alpha_{h_i^\star}, (\alpha_h)_{h \in \hat{\mathcal{H}}_{-j} \setminus \{h_i^\star\}}) = g_j(\phi^{-1}(\alpha_{h_i^\star}), (\alpha_h)_{h \in \mathcal{H}_{-j} \setminus \{h_i', h_i''\}}) / (\phi^{-1}(\alpha_{h_i^\star})(a_{\text{pass}}))^{n-1}$ for all strictly mixed actions $\alpha_{h_i^\star}, (\alpha_h)_{h \in \hat{\mathcal{H}}_{-j} \setminus \{h_i^\star\}}$; (2) for i, $\hat{g}_i = g_i$. That is, $\hat{g}_j$ is over a different domain than g_j since i has one fewer information set in $\hat{\mathcal{G}}$ than $\mathcal{G}$, but we identify each strictly mixed $\alpha_{h_i^\star}$ in the domain of $\hat{g}_j$ with $\phi^{-1}(\alpha_{h_i^\star})$ in the domain of g_j and re-normalize appropriately. Note g_j is strictly positive on the interior and $0 < \phi^{-1}(\alpha_{h_i^\star})(a_{\text{pass}}) < \infty$, so $\hat{g}_j$ is also strictly positive on the interior. This shows the constructed prior $\hat{g}$ is non-doctrinaire.

By the definition of ϕ, each action in $\{a_1, ..., a_m, a_{m+1}, ..., a_{m+n}\}$ has the same likelihood under $\alpha_{h_i^\star}$ and $\phi^{-1}(\alpha_{h_i^\star})$. Also, for every open set $E \subseteq \Delta^\circ(A_{h_i'}) \times \Delta^\circ(A_{h_i''})$, the probability that g_j assigns to E is the same as the probability that $\hat{g}_j$ assigns to $\phi(E) \subseteq \Delta^\circ(A_{h_i^\star})$. Note that for each $\alpha_{h_i'} \in \Delta^\circ(A_{h_i'})$, the projection $E_{\alpha_{h_i'}} := \{\alpha_{h_i''} \in \Delta^\circ(A_{h_i''}) : (\alpha_{h_i'}, \alpha_{h_i''}) \in E\}$ can be viewed as a subset of $\Delta_n^1 := \{x_{m+1}, ..., x_{m+n-1} \geq 0 \text{ s.t. } x_{m+1} + ... + x_{m+n-1} \leq 1\} \subseteq \mathbb{R}^{n-1}$. On the other hand, the image of this projection $\phi(\{\alpha_{h_i'}\} \times E_{\alpha_{h_i'}})$ can be viewed as a subset of $\Delta_n^{\alpha_{h_i'}(a_{\text{pass}})} := \{x_{m+1}, ..., x_{m+n-1} \geq 0 \text{ s.t. } x_{m+1} + ... + x_{m+n-1} \leq \alpha_{h_i'}(a_{\text{pass}})\} \subseteq \mathbb{R}^{n-1}$. Both Δ_n^1 and $\Delta_n^{\alpha_{h_i'}(a_{\text{pass}})}$ are $n-1$ dimensional polytopes, and the latter's volume is $(\alpha_{h_i'}(a_{\text{pass}}))^{n-1}$ that of the former. The normalizing factor $1/(\phi^{-1}(\alpha_{h_i^\star})(a_{\text{pass}}))^{n-1}$ ensures $g_j(E) = \hat{g}_j(\phi(E))$,

Combining the two observations in the previous paragraph, we see that no matter which terminal node is observed, the posterior of g_j will again assign the same probability to E as the posterior of $\hat{g}_j$ assigns to $\phi(E)$. This discussion shows that for any $0 \leq \delta, \gamma < 1$, the set of steady states with $\hat{g}$ in $\hat{\mathcal{G}}$ is the same as the set of steady states with g in $\mathcal{G}$.

Conversely, given a prior density $\hat{g}_j$ for every agent j in the game $\hat{\mathcal{G}}$, we can consider a prior density g_j in $\mathcal{G}$ where $g_j(\alpha_{h_i'}, \alpha_{h_i''}, (\alpha_h)_{h \in \mathcal{H}_{-j} \setminus \{h_i', h_i''\}}) = \hat{g}_j(\phi(\alpha_{h_i'}, \alpha_{h_i''}), (\alpha_h)_{h \in \hat{\mathcal{H}}_{-j} \setminus \{h_i^\star\}}) \cdot (\alpha_{h_i'}(a_{\text{pass}}))^{n-1}$ for $j \neq i$. The same argument as above shows for any $0 \leq \delta, \gamma < 1$, the set of steady states with g in $\mathcal{G}$ is the same as the set of steady states with $\hat{g}$ in $\hat{\mathcal{G}}$.

A.2 Patiently Stable Profiles for Figure 2 Are Nash Equilibria

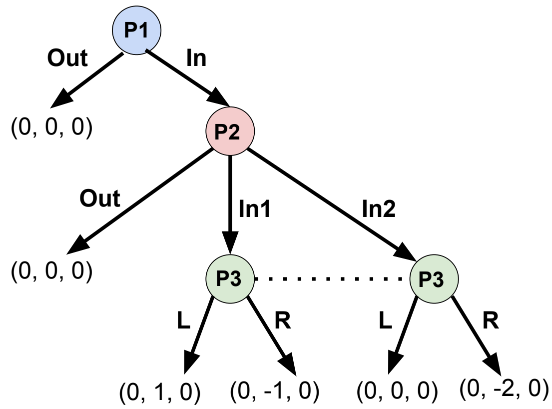


Figure 8: A game with discrete terminal node partitions in which the learning problem of a P2 agent with a given prior is identical to the learning problem of a P2 agent with the same prior in the Figure 2 game.

Proof. Consider the auxiliary game depicted in Figure 8, which modifies payoffs to make P1 and P3 indifferent between all terminal nodes, and ends the game immediately if P2 chooses **Out**. For any $0 \leq \delta, \gamma < 1$, any P1 or P3 policy used in the original steady state is an optimal policy for the corresponding agent in the auxiliary game, so any steady state profile $\pi^* \in \Pi^*(g, \delta, \gamma)$ for the original game is also a steady state profile of the auxiliary game. And P2 faces the same learning problem in the auxiliary game as in the original game, since each strategy profile gives them the same payoff and the same feedback in both games. But we know in every patiently stable profile in the auxiliary game, P2 must not have a profitable deviation, so the same must apply to the patiently stable profiles of the original game. Likewise, we can construct auxiliary games for P1 and P3 to show that they must not have profitable deviations in patiently stable profiles of the original game. ■

A.3 Examples of Games with Terminal Node Partitions That Have Auxiliary Games with Discrete Partitions

The argument in the previous section can be used to show that patient stability selects only Nash equilibria whenever the game and feedback structure are such that, for each player role, there is an auxiliary game with a discrete terminal node partition that leads to the same learning problem for agents in that role. We give some examples of games from the previous literature that meet this condition below.

In Fudenberg and Kamada (2015), Figure 1 Game B and Figure 3 present two games where three players P1, P2, and P3 simultaneously choose actions, P2 and P3 always see the terminal node, and P1 sees the terminal node when they choose **In** but not when they choose **Out**. P1 always gets 0 payoff from choosing **Out**. For P2 and P3, consider the auxiliary game where every player always observes the terminal node. This clearly does not affect P2 and P3's learning problems. For P1, consider an auxiliary game where P1 moves first and the game ends with P1 getting 0 payoff if P1 chooses **Out**. If P1 chooses **In**, then P2 and P3 choose their actions simultaneously as before. All players observe terminal nodes. P1's learning problem in the auxiliary game is the same as in the original game. Thus for Figure 1 Game B and Figure 3 with their original terminal node partitions, patiently stable profiles are Nash equilibria.

Figure 5 of Fudenberg and Kamada (2015) is a three-player game where P1, P2 and P3 simultaneously choose **In** or **Out**. When P1 chooses **In**, they learn P2 and P3's choices, but P1 does not learn how others play if they choose **Out**. P2 always learns how P1 plays, but they only learn how P3 plays if they choose **In** rather than **Out**. Similarly, P3 always learns how P1 plays, but they only learn how P2 plays if they choose **In** rather than **Out**. Players who choose **Out** always get 0. For P1, consider the auxiliary game where they move first and choose **In** or **Out**. If they choose **Out**, the game ends with them getting 0. If they choose **In**, then P2 and P3 simultaneously choose **In** or **Out**. All players observe terminal nodes. This is the same learning problem as in the original game for P1. Next, consider the auxiliary game where P1 and P2 move simultaneously at the start of the game. If P2 chooses **Out**, the game ends with P2 getting 0. If P2 chooses **In**, then P3 chooses between **In** or **Out** without knowing P1's choice. All players observe terminal nodes. This is the same learning problem as in the original game for P2, because the terminal node always reveals P1's play, even when P2 chooses **Out**. But if P2 chooses **Out**, then the terminal node does not show what P3 would have played. Similarly, we can construct an analogous auxiliary game for P3. This shows that for the game in Figure 5, patiently stable profiles are Nash equilibria.

Section 5.1.1 of Fudenberg and He (2021) studies the "restaurant game" where three players P1, P2, and P3 move simultaneously. P1 is a restaurant that chooses between **high** and **low** ingredient qualities, while P2 and P3 are two potential customers who decide whether to go to the restaurant (**In**) or eat at home (**Out**). P1 always sees P2 and P3's choices. P2 sees how others play if they choose **In**, but not if they choose **Out**. Similarly, P3 sees how others play if they choose **In**, but not if they choose **Out**. Choosing **Out** always gives 0 payoff. For P1, the auxiliary game where everyone sees the terminal node does not affect their learning problem. For P2, consider the auxiliary game where they move first, choosing between **In** and **Out**. If they choose **Out**, the game ends with payoff 0 for them. If they choose **In**, then P1 and P3 move simultaneously. All players observe terminal nodes. This auxiliary game presents the same learning problem for P2 as in the original game. Similarly, there is an analogous auxiliary game with discrete terminal node partitions that preserves P3's learning problem, so patiently stable profiles are Nash equilibria in this game.

A.4 Proof of Proposition 2

Proof. Throughout this proof, we think of P1 actions that end the game as leading to singleton P2 information sets where P2 only has one action. Denote the P2 information set reached when P1 plays a_1^* with h_2^*, and use $\mathcal{H}_2^{off} = \mathcal{H}_2 \setminus \{h_2^*\}$ to denote the set of P2 information sets that are off-path under π^*. Let $\widetilde{\Pi}_1 = \{\pi_1 \in \Pi_1 : \forall h_2 \in$

$\mathcal{H}_2^{off},\ \pi_1(a_1) = 0\ \forall a_1 \in \rho(h_2) \setminus \{a_1^*(h_2)\}\}$ be the set of P1 behavior strategies that, for every $h_2 \in \mathcal{H}_2^{off}$, put probability 0 on any action in $\rho(h_2)$ that is not $a_1^*(\alpha_1)$. Further, let $\widetilde{\Pi}_2 = \{\pi_2 \in \Pi_2 : \forall h_2 \in \mathcal{H}_2^{off},\ \pi_2(a_2^*(h_2)|h_2) = 1\}$ be the set of P2 behavior strategies which respond with $a_2^*(h_2)$ at any off-path information set h_2. Throughout the proof, we restrict attention to strategy profiles $\pi \in \widetilde{\Pi}_1 \times \widetilde{\Pi}_2$.

By continuity, there is an $\eta > 0$ such that (1) for any $\pi_1 \in \Pi_1$ satisfying $\pi_1(a_1^*) \geq 1 - \eta$, the unique optimal action for P2 to play at h_2^* is $a_2^*(h_2^*)$, and (2) for any $\pi_2 \in \widetilde{\Pi}_2$ for which $\pi_2(a_2^*(h_2^*)|h_2^*) \geq 1 - \eta$, the unique P1 best response is a_1^*. We focus on steady state profiles in which the aggregate probabilities that P1 plays a_1^* and that P2 plays $a_2^*(h_2^*)$ at h_2^* both exceed $1 - \eta$. We argue that such steady state profiles exist in the limit, and that the corresponding aggregate probabilities that P1 plays a_1^* and P2 plays $a_2^*(h_2)$ in response to any information set h_2 converge to 1.

Let $\xi : \widetilde{\Pi}_1 \to \widetilde{\Pi}_1$ be the continuous mapping given by

$$\xi(\pi_1)(a_1) = \begin{cases} \max\{\pi_1(a_1^*), 1 - \eta\} & \text{if } a_1 = a_1^* \\ \left(1 - \mathbb{1}(\pi_1(a_1^*) < 1 - \eta)\frac{1-\eta-\pi_1(a_1^*)}{1-\pi_1(a_1^*)}\right)\pi_1(a_1) & \text{if } a_1 \neq a_1^* \end{cases}.$$

This function transforms each π_1 into a P1 behavior strategy that puts probability at least $1 - \eta$ on a_1^* and satisfies $\xi(\pi_1) = \pi_1$ whenever $\pi_1(a_1^*) \geq 1 - \eta$. Similarly, let $\phi : \widetilde{\Pi}_2 \to \widetilde{\Pi}_2$ be the continuous mapping such that

$$\phi(\pi_2)(a_2|h_2^*) = \begin{cases} \max\{\pi_2(a_2^*(h_2^*)|h_2^*), 1 - \eta\} & \text{if } a_2 = a_2^*(h_2^*) \\ \left(1 - \mathbb{1}(\pi_2(a_2^*(h_2^*)|h_2^*) < 1 - \eta)\frac{1-\eta-\pi_2(a_2^*(h_2^*)|h_2^*)}{1-\pi_2(a_2^*(h_2^*)|h_2^*)}\right)\pi_2(a_2|h_2^*) & \text{if } a_2 \neq a_2^*(h_2^*) \end{cases}.$$

This takes each π_2 into a P2 behavior strategy that uses $a_2^*(h_2^*)$ at h_2^* with probability at least $1 - \eta$. Note that ϕ coincides with the identity mapping whenever $\pi_2(a_2^*(h_2^*)|h_2^*) \geq 1 - \eta$.

Since g_1 is supportive of π^*, for any $\pi_2 \in \widetilde{\Pi}_2$, $\mathscr{R}_1^{\delta,\gamma}(\pi_2)(a_1) = 0$ for all $a_1 \in \rho(h_2)$ for all $h_2 \in \mathcal{H}_2^{off}$. This means that $\mathscr{R}_1^{\delta,\gamma}(\pi_2) \in \widetilde{\Pi}_1$ for all $\pi_2 \in \widetilde{\Pi}_2$. Likewise, since g_2

is supportive of π^*, for any $\pi_1 \in \widetilde{\Pi}_1$, $\mathscr{R}_2^{\delta,\gamma}(\pi_1)(a_2|h_2) = 0$ for all $a_2 \neq a_2^*(h_2)$ for all $h_2 \in \mathcal{H}_2^{off}$. Thus, $\mathscr{R}_2^{\delta,\gamma}(\pi_1) \in \widetilde{\Pi}_2$ for all $\pi_1 \in \widetilde{\Pi}_1$. Consequently, $\mathscr{R}^{\delta,\gamma}$ maps $\widetilde{\Pi}_1 \times \widetilde{\Pi}_2$ into itself regardless of $\delta, \gamma \in [0,1)$, so the mapping $\widetilde{\mathscr{R}}^{\delta,\gamma} : \widetilde{\Pi}_1 \times \widetilde{\Pi}_2 \to \widetilde{\Pi}_1 \times \widetilde{\Pi}_2$ given by $\widetilde{\mathscr{R}}^{\delta,\gamma}(\pi_1,\pi_2) = (\xi(\mathscr{R}_1^{\delta,\gamma}(\pi_2)), \phi(\mathscr{R}_2^{\delta,\gamma}(\pi_1)))$ is well-defined. Since this mapping is continuous, Brouwer's fixed point theorem guarantees the existence of a fixed point $\pi^{\delta,\gamma} = (\pi_1^{\delta,\gamma}, \pi_2^{\delta,\gamma})$ for any $\delta, \gamma \in [0,1)$.

Consider any parameter sequence $\{\delta_j\}_{j\in\mathbb{N}}$, $\{\gamma_{j,k}\}_{j,k\in\mathbb{N}}$ such that $\lim_{j\to\infty} \delta_j = 1$, $\lim_{k\to\infty} \gamma_{j,k} = 1$ for all $j \in \mathbb{N}$, and $\lim_{j\to\infty} \lim_{k\to\infty} \pi^{\delta_j,\gamma_{j,k}} = \hat{\pi}$ for some $\hat{\pi} \in \widetilde{\Pi}_1 \times \widetilde{\Pi}_2$. Using the fact that the unique P1 best response is a_1^* to any $\pi_2 \in \widetilde{\Pi}_2$,

$$\lim_{j\to\infty} \lim_{k\to\infty} \mathscr{R}_1^{\delta_j,\gamma_{j,k}}(\pi_2^{\delta_j,\gamma_{j,k}})(a_1^*) = 1$$

can be shown using a similar auxiliary game argument to the one given when arguing that patient stability selects Nash equilibria in the Figure 2 game. Thus, $\lim_{j\to\infty} \lim_{k\to\infty} \mathscr{R}_1^{\delta_j,\gamma_{j,k}}(\pi_2^{\delta_j,\gamma_{j,k}}) = \pi_1^*$. As $\xi(\pi_1) = \pi_1$ if $\pi_1(a_1^*) \geq 1-\eta$, it follows that there is some $\overline{j} \in \mathbb{N}$ and $\overline{k} : \mathbb{N} \to \mathbb{N}$ such that $\mathscr{R}_1^{\delta_j,\gamma_{j,k}}(\pi_2^{\delta_j,\gamma_{j,k}}) = \pi_1^{\delta_j,\gamma_{j,k}}$ if $j > \overline{j}$ and $k > \overline{k}(j)$. Similarly, since $a_2^*(h_2^*)$ is the uniquely optimal action at h_2^* given any $\pi_1 \in \widetilde{\Pi}_1$, $\lim_{j\to\infty} \lim_{k\to\infty} \mathscr{R}_2^{\gamma_{j,k}}(\pi_1^{\delta_j,\gamma_{j,k}})(a_2^*(h_2^*)|h_2^*) = 1$ must hold. This means that $\lim_{j\to\infty} \lim_{k\to\infty} \mathscr{R}_2^{\gamma_{j,k}}(\pi_1^{\delta_j,\gamma_{j,k}}) = \pi_2^*$. Moreover, as $\phi(\pi_2) = \pi_2$ if $\pi_2(a_2^*(h_2^*)|h_2^*) \geq 1-\eta$, it follows that $\mathscr{R}_2^{\delta_j,\gamma_{j,k}}(\pi_1^{\delta_j,\gamma_{j,k}}) = \pi_2^{\delta_j,\gamma_{j,k}}$ for $j > \overline{j}$ and $k > \overline{k}(j)$. Collecting these findings reveals that $\pi^{\delta_j,\gamma_{j,k}}$ is a fixed point of the aggregate response mapping $\mathscr{R}^{\delta_j,\gamma_{j,k}}$, and thus a steady state profile by Proposition 1, whenever $j > \overline{j}$ and $k > \overline{k}(j)$. Since $\lim_{j\to\infty} \lim_{k\to\infty} \pi^{\delta_j,\gamma_{j,k}} = \pi^*$, we conclude that π^* is stable. ■

A.5 Proof of Proposition 3

We first establish three lemmas.

Lemma 1. *Fix $\delta \in [0,1)$ and a non-doctrinaire P1 prior g_1. For each $\gamma \in [0,1)$, fix a P1 policy that is optimal given g_1 and never prescribes **In** after it has previously prescribed **Out**. There is some $\kappa \in \mathbb{R}_+$ such that, for arbitrary $\gamma \in [0,1)$, when the* *aggregate P3 strategy puts probability* $\pi_3^{\delta,\gamma}(\boldsymbol{L}) \leq 1/4$ *on* $\boldsymbol{L}$*, the aggregate P1 strategy satisfies* $\pi_1^{\delta,\gamma}(\boldsymbol{In})/(1-\gamma) \leq \kappa$.

Proof. We show that there is some $N \in \mathbb{N}$ such that all P1 agents who have lived at $t \geq N$ periods and have been matched with P3 agents that would play **L** fewer than $t/3$ periods would play **Out**. For each N, consider a P1 agent who has played **In** at least N times and who has observed their P3 opponents play **R** at least 2/3 of the times when **In** is used. By Theorem 4.2 of Diaconis and Freedman (1990), there is an $N \in \mathbb{N}$ so that such a P1 agent will put probability at least $3/(4-\delta)$ on the true probability with which a randomly selected P3 agent plays **R** being weakly more than 2/3. Such an agent thus puts at least probability $3/(4-\delta)$ on aggregate opponent behavior strategy profiles for which the expected payoff from playing **In** is no more than $-1/3$. We claim that this N has the desired properties. To see this, consider a P1 agent who has lived at least N periods and for whom the fraction of time periods where they were matched with a P3 agent that would play **L** that periods is less than 1/3. Then, either that agent has played **Out** in the past, in which case they will again play **Out**, or they have played **In** at least N times. Restricting attention to the latter case, the agent must have a posterior belief that puts probability $p > 3/(4-\delta)$ on aggregate opponent behavior strategy profiles for which the expected payoff from playing **In** is no more than $-1/3$. An upper bound on the agent's expected discounted future lifetime payoff from playing **In** is $(1-\delta)(1-4p/3)+\delta(1-p)$, since the expected current period payoff to playing **In** is weakly less than $p(-1/3)+1-p = 1-4p/3$ and the agent's continuation payoff is bounded above by $\delta(1-p)$, since P1's maximum payoff is 1. As $p > 3/(4-\delta)$, it follows that $(1-\delta)(1-4p/3)+\delta(1-p) < 0$, so the agent must play **Out**.

We now combine this fact with Hoeffding's inequality to derive the desired constraint on the P1 aggregate strategy. By Hoeffding's inequality, there is some $c > 0$ such that, for any aggregate P3 strategy satisfying $\pi_3^{\delta,\gamma}(\mathbf{L}) \leq 1/4$, the share of P1 agents who have lived n periods and for whom the fraction of time periods where they were matched with a P3 agent that play **L** is less than 1/3 is at least $1 - e^{-cn}$. Thus,

$$\begin{aligned}\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma} &\leq \frac{1}{1-\gamma}\left(1-\gamma^N+\sum_{t=N}^{\infty}(1-\gamma)\gamma^t e^{-ct}\right)\\ &= \frac{1-\gamma^N}{1-\gamma}+\frac{\gamma^N e^{-cN}}{1-\gamma e^{-c}}.\end{aligned}$$

Observe that the right-hand side of the inequality converges to $N + 1/(e^{cN} - e^{c(N-1)})$ as $\gamma \to 1$. Since $\pi_1^{\delta,\gamma}(\mathbf{In})/(1-\gamma)$ can never be more than $1/(1-\gamma)$, it follows that $\pi_1^{\delta,\gamma}(\mathbf{In})/(1-\gamma)$ must be uniformly bounded from above by some $\kappa \in \mathbb{R}_+$. ■

Lemma 2. *Fix $\delta \in [0, 1)$ and a non-doctrinaire P2 prior g_2 under which the expected probability of **L** is strictly less than 1/2. Consider a sequence of steady states such that the probability of **In** under the aggregate P1 strategy satisfies $\pi_1^{\delta,\gamma}(\boldsymbol{In})/(1-\gamma) \leq \kappa$ for all $\gamma \in [0, 1)$ for some $\kappa \in \mathbb{R}_+$. Then $\lim_{\gamma\to1} \pi_2^{\delta,\gamma}(\boldsymbol{Out}) = 1$.*

Proof. The idea is to construct a sequence $\{N_j\}_{j\in\mathbb{N}}$ so that, regardless of the values of δ and γ, a P2 agent who observes at least N_j observations of P1 choosing **Out** and no more than j observations of P1 choosing **In** will find it optimal to play **Out**. We then leverage this property to show that $\pi_1^{\delta,\gamma}(\mathbf{In})/(1-\gamma) \leq \kappa$ for all γ implies that almost all P2 agents must play **Out** as $\gamma \to 1$.

Fix an $\epsilon > 0$ such that the expected probability of **L** under g_2 is weakly less than $(1-\epsilon)/2$. We first establish that there is some $N_0 \in \mathbb{N}$ such that, regardless of $\gamma \in [0, 1)$, every P2 agent who has lived at least N_0 periods and has never observed a P1 agent play **In** will play **Out**. Theorem 4.2 of Diaconis and Freedman (1990) implies that there is an $N \in \mathbb{N}$ such that, under the posterior belief of a P2 agent who has at least N observations of P1 agents playing **Out** and no observations of a P1 agent playing **In**, the expected value of the probability with which a randomly selected P1 agent will play **In**, ν, is strictly less than $(1-\delta)\epsilon/\delta$. This N satisfies the desired properties given at the beginning of the paragraph. To see this, consider a P2 agent who has lived at least N periods and has never observed a P1 agent play **In**. An upper bound on the agent's expected discounted future lifetime payoff from playing **In1** is $-(1-\delta)\epsilon + \delta\nu$, which is strictly negative since $\nu < (1-\delta)\epsilon/\delta$, so such an agent must play **Out**.

Inductively applying similar arguments shows there is a sequence of $\{N_j\}_{j\in\mathbb{N}} \subseteq \mathbb{N}$ such that the following holds: For all $\gamma \in [0,1)$ and $j \in \mathbb{N}$, every P2 agent who has lived at least N_j periods, has at most j observations of P1 agents playing **In**, and witnessed no P1 agents playing **In** in their first N_j observations will play **Out**. Observe that, when the probability of a randomly selected P1 agent playing **In** is $\pi_1(\mathbf{In})$, the share of P2 agents who have lived at least $N_j + j$ periods and have exactly j observations of P1 agents playing **In**, all of which came after their first N_j periods, is

$$\begin{aligned}
&\sum_{t=N_j+j}^{\infty}(1-\gamma)\gamma^t \frac{(t-N_j)!}{(t-N_j-j)!j!}\pi_1^{\delta,\gamma}(\mathbf{In})^j\left(1-\pi_1^{\delta,\gamma}(\mathbf{In})\right)^{t-j} \\
=&\gamma^{N_j+j}\left(1-\pi_1^{\delta,\gamma}(\mathbf{In})\right)^{N_j}\frac{\pi_1^{\delta,\gamma}(\mathbf{In})^j}{\left(1-\gamma+\gamma\pi_1^{\delta,\gamma}(\mathbf{In})\right)^{j+1}} \\
=&\gamma^{N_j+j}\left(1-\pi_1^{\delta,\gamma}(\mathbf{In})\right)^{N_j}\left(\frac{1+\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}}{1+\gamma\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}}\right)^{j+1}\frac{\left(\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}\right)^j}{\left(1+\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}\right)^{j+1}}.
\end{aligned}$$

It thus follows that, for a given $\gamma \in [0,1)$ and steady-state strategy profile,

$$\begin{aligned}
\pi_2^{\delta,\gamma}(\mathbf{Out}) \geq &\sum_{j=0}^{\infty}\gamma^{N_j+j}\left(1-\pi_1^{\delta,\gamma}(\mathbf{In})\right)^{N_j}\left(\frac{1+\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}}{1+\gamma\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}}\right)^{j+1}\frac{\left(\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}\right)^j}{\left(1+\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}\right)^{j+1}} \\
=&1-\sum_{j=0}^{\infty}\left(1-\gamma^{N_j+j}\left(1-\pi_1^{\delta,\gamma}(\mathbf{In})\right)^{N_j}\left(\frac{1+\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}}{1+\gamma\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}}\right)^{j+1}\right)\frac{\left(\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}\right)^j}{\left(1+\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}\right)^{j+1}} \\
\geq &1-\sum_{j=K+1}^{\infty}\frac{\left(\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}\right)^j}{\left(1+\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}\right)^{j+1}} \\
&-\sup\left\{1-\gamma^{N_j+j}\left(1-\pi_1^{\delta,\gamma}(\mathbf{In})\right)^{N_j}\left(\frac{1+\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}}{1+\gamma\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}}\right)^{j+1}\right\}_{j\in\{0,1,\ldots,K\}}
\end{aligned}$$

for arbitrary $K \in \mathbb{N}$. Fix an arbitrary $\eta > 0$ and take K to be large enough so that $\sum_{j=K+1}^{\infty} \kappa^j/(1+\kappa)^{j+1} < \eta$. Then the right-hand side of the first line of the final inequality is greater than $1-\eta$ for all $\pi_1^{\delta,\gamma}(\mathbf{In})/(1-\gamma) \leq \kappa$. Observe that the elements of the set over which the supremum is taken in the final line converge to 0 as $\gamma \to 1$ uniformly over $\pi_1^{\delta,\gamma}(\mathbf{In})/(1-\gamma) \leq \kappa$. Thus, $\liminf_{\gamma\to 1} \pi_2^{\delta,\gamma}(\mathbf{Out}) \geq 1-\eta$. Since this holds for all $\eta > 0$, we have $\lim_{\gamma\to 1} \pi_2^{\delta,\gamma}(\mathbf{Out}) = 1$. ■

Lemma 3. *Fix $\delta \in [0,1)$ and a non-doctrinaire P3 prior g_3 that leads a P3 agent to play **L** only when they have previously observed a P2 agent play **In1**. Consider a sequence of steady-states such that the probability of **In** under the aggregate P1 strategy satisfies $\pi_1^{\delta,\gamma}(\boldsymbol{In})/(1-\gamma) \leq \kappa$ for all $\gamma \in [0,1)$ for some $\kappa \in \mathbb{R}_+$ and $\lim_{\gamma\to 1} \pi_2^{\delta,\gamma}(\boldsymbol{In1}) = 0$. Then $\lim_{\gamma\to 1} \pi_3^{\delta,\gamma}(\boldsymbol{L}) = 0$.*

Proof. Observe that

$$
\begin{aligned}
\pi_3^{\delta,\gamma}(\mathbf{L}) &\leq \sum_{t=0}^{\infty}(1-\gamma)\gamma^t(1-(1-\pi_1^{\delta,\gamma}(\mathbf{In})\pi_2^{\delta,\gamma}(\mathbf{In1}))^t) \\
&= 1 - \frac{1-\gamma}{1-\gamma(1-\pi_1^{\delta,\gamma}(\mathbf{In})\pi_2^{\delta,\gamma}(\mathbf{In1}))} \\
&= 1 - \frac{1}{1+\gamma\frac{\pi_1^{\delta,\gamma}(\mathbf{In})}{1-\gamma}\pi_2^{\delta,\gamma}(\mathbf{In1})}
\end{aligned}
$$

since the right-hand side of the inequality is the share of P3 agents who have previously observed a P2 agent play **In1**. By Lemma 1, there is some $\kappa \in \mathbb{R}_+$ such that $\limsup_{\gamma\to 1} \pi_1^{\delta,\gamma}/(1-\gamma) \leq \kappa$, while $\lim_{\gamma\to 1} \pi_2^{\delta,\gamma}(\mathbf{In1}) = 0$ by Lemma 2. It follows that $\lim_{\gamma\to 1} 1+\gamma(\pi_1^{\delta,\gamma}(\mathbf{In}))/(1-\gamma)\pi_2^{\delta,\gamma}(\mathbf{In1}) = 1$, which implies $\lim_{\gamma\to 1} \pi_3^{\delta,\gamma}(\mathbf{L}) = 0$. ■

Proof of Proposition 3. Lemmas 1, 2, and 3 together imply that, for fixed $\delta \in [0,1)$, the aggregate response mapping maps the set of aggregate strategy profiles where $\pi_3^{\delta,\gamma}(L) \leq 1/4$ into itself when γ is close enough to 1. Brouwer's fixed point theorem then guarantees the existence of a steady state profile satisfying this inequality for all sufficiently high γ. Lemmas 1, 2, and 3 further imply that, in the $\gamma \to 1$ limit of such a sequence of steady state profiles, $\lim_{\gamma\to 1} \pi_1^{\delta,\gamma}(\mathbf{Out}) = 1$, $\lim_{\gamma\to 1} \pi_2^{\delta,\gamma}(\mathbf{Out}) = 1$, and

$\lim_{\gamma \to 1} \pi_3^{\delta,\gamma}(\mathbf{R}) = 1$ must be satisfied. Since $\delta \in [0, 1)$ is arbitrary, we conclude that $(\mathbf{Out}, \mathbf{Out}, \mathbf{In})$ is patiently stable. ■

A.6 Proof of Proposition 5

We first state a supporting lemma that shows that, with enough data, an agent's posterior belief about others' play will put high probability on the empirical distribution they have observed.

Lemma 4. *For any fixed non-doctrinaire prior and every $\eta > 0$, there is some M such that, whenever an agent has at least M observations, for each opponent population, the agent's posterior belief puts probability $1 - \eta$ on strategy distributions within η, under the sup norm, of the empirical distribution they have observed.*

This follows from the Fudenberg, Lanzani, and Strack (2021) extension of the pathwise concentration result of Diaconis and Freedman (1990) to priors that do not have full support. The support restriction arise because the agent's prior is concentrated on distributions that can be generated by independent randomizations of their opponents.

Proof of Proposition 5. Fix a discount factor $\delta \in [0, 1)$, and consider a sequence of survival probabilities $\{\gamma_k\}_{k=1}^{\infty}$ with an associated sequence of steady-state strategy profiles $\{\pi_k\}_{k=1}^{\infty}$ such that $\pi_k \to \pi^*$. We show that $\pi_i^*(s_i) = 0$ for each i and each $s_i \notin S_i^*$.

First, it is clear that $\pi_i^*(s_i) = 0$ for each i and each $s_i \notin S_i^{(0)}$. This is because agents have full-support posterior beliefs after every history and their observations do not depend on their play, so they never use weakly dominated strategies.

By Lemma 4, for a given $\eta > 0$, there is some M such that, whenever a player has at least M observations, their posterior belief over the prevailing strategy distribution in each of their opponent populations puts probability $1 - \eta$ on strategy distributions within η of the empirical distribution they have observed. By the law of large numbers, we can choose this M to be such that the posterior beliefs of an agent in an arbitrary player role i who has lived at least M periods will be accurate with high probability in the following sense. With probability $1-2\eta$, at the end of the period, following any possible observation in the period itself, the agent's posterior belief puts probability at least $1-2\eta$ on strategy distributions within 2η of the true prevailing distribution for each opponent role $j \neq i$.

Now we show by induction that π_i^* puts probability 1 on the decreasing subsets $S_i^{(0)}, S_i^{(1)}, S_i^{(2)}, \ldots$. We have shown that in the base case $\pi_i^*(s_i) = 0$ for each i and each $s_i \notin S_i^{(0)}$. Suppose inductively that $\pi_i^*(s_i) = 0$ for each i and $s_i \notin S_i^{(m)}$ for some m. Fix arbitrary $\epsilon, \nu > 0$ and restrict attention to k large enough so that $|\pi_{i,k} - \pi_i^*| < \epsilon/2$ for all player roles i. By the preceding argument, we know that for all sufficiently large k, there is more than $1-\nu$ share of player i agents whose posterior beliefs at the end of the period for each opponent role j put probability at least $1-\epsilon$ on strategy distributions within $1-\epsilon$ of $\Delta(S_j^{(m)})$. Let $\sigma \in \Delta(S_{-i})$ be the expectation held by such an agent about the play of their opponents in the current period. The properties of the agent's beliefs imply that σ is full support and that $\sigma(S_j^{(m)}|s_{-ij}) \geq (1-\epsilon)^2$ for all $s_{-ij} \in S_{-ij}$. For all sufficiently small ϵ, every $s_i \notin S_i^{(m)}$ must be suboptimal for such an agent, so $\pi_i^*(s_i) \leq \nu$ must hold for each i and $s_i \notin S_i^{(m+1)}$. Since this is true for all $\nu > 0$, we conclude that $\pi_i^*(s_i) = 0$ for each i and $s_i \notin S_i^{(m+1)}$.

■

A.7 Proof of Proposition 4

Proof. Let $D_i^{(m)}$ be the extensive-form strategies of i that choose an action inconsistent with backward induction at a decision node that is $m+1$ steps away from the terminal nodes, but do not do so at any decision nodes closer to the terminal nodes. We show this is a valid elimination sequence as defined by Definition 4, so this result follows from Proposition 5. To begin, note that the elements of $D_i^{(0)}$ are weakly dominated for i: For any $s_i \in D_i^{(0)}$, consider a different strategy s_i' that changes one of the non-backward-induction actions at one of i's decision nodes h_i one step away from terminal nodes to a backward-induction action. Then $u_i(s_i', s_{-i}) \geq u_i(s_i, s_{-i})$ for all s_{-i}. Moreover, there exists at least one s^*_{-i} such that h_i is reached (since the game is simple), so $u_i(s'_i, s^*_{-i}) > u_i(s_i, s^*_{-i})$.

By definition, $S_i^{(m)}$ are the strategies where i uses the backward-induction action at all decision nodes $m+1$ steps or fewer away from terminal nodes. To see that each $s_i \in D_i^{(m+1)}$ fails to be a best response for i to full-support conjectures of their opponents' play that put high conditional probabilities on $S_j^{(m)}$ for each $j \neq i$, let h_i be a decision node $m+2$ steps away from terminal nodes where s_i does not choose the backward-induction action, and let s'_i be the strategy that only differs from s_i in that s'_i by selecting a backward-induction action at h_i. Let J be the set of players who have decision nodes in the subgame starting at h_i, not counting h_i itself. Because the game is simple, $i \notin J$. For the same reason, whether play reaches h_i does not depend on the strategy of i or the strategies of the players in J. Let $-iJ$ denote all players $k \notin i \cup J$, and let $S^{reach}_{-iJ} \subseteq S_{-iJ}$ be the set of strategies of $-iJ$ that reach h_i.

Because i's payoff in the subgame starting at h_i only depends on i's action at h_i and on the strategies of players in J, for any $s_{-i} \in S_{-i}^{(m)}$, i's payoff $u_i(s'_i, s_{-i} \mid h_i)$ from playing s'_iin the subgame starting at h_i strictly higher than their payoff $u_i(s_i, s_{-i} \mid h_i$ from s_i. Consider any strictly mixed profile σ_{-i} where $\sigma_{-i}(S_J^{(m)} | s_{-iJ}) \geq 1 - \epsilon$ for each $s_{-iJ} \in S_{-iJ}$. When $-iJ$ play a strategy profile in S^{reach}_{-iJ}, i's payoff is equal to i's payoff in the subgame starting at h_i. When $-iJ$ choose a strategy outside of S^{reach}_{-iJ}, i is indifferent between s_i and s'_i. Therefore, $u_i(s_i, \sigma_{-i}) - u_i(s'_i, \sigma_{-i}) = \sigma_{-i}(S^{reach}_{-iJ}) \cdot [\mathbb{E}[u_i(s_i, s_{-i} \mid h_i) | S^{reach}_{-iJ}] - u_i(s'_i, s_{-i} \mid h_i) | S^{reach}_{-iJ}]]$. We have $\sigma_{-iJ}(S^{reach}_{-iJ}) > 0$ since each opponent's strategy is strictly mixed, and so $u_i(s_i, (\sigma_j)_{j \in J} \mid h_i) - u_i(s'_i, (\sigma_j)_{j \in J} \mid h_i) < 0$ for all sufficiently small $\epsilon > 0$. ■

A.8 Proof of Proposition 6

Proof. First note that for every non-doctrinaire prior g over behavior strategies in $\mathcal{G}$, there is a non-doctrinaire prior $\hat{g}$ over mixed strategies in $\mathcal{G}$ that generates the same set of steady states for every $0 \leq \delta, \gamma < 1$. This is because each $-i$ behavior strategy

$(\alpha_{h_{-i}})_{h_{-i}\in\mathcal{H}_i}$ is associated with an equivalent mixed strategy $\sigma_{-i} \in \Delta(\mathbb{S}_{-i})$, defined by $\sigma_{-i}(s_{-i}) = \times_{h_{-i}\in\mathcal{H}_{-i}}\alpha_{h_{-i}}(s_{-i}(h_{-i}))$ for each $s_{-i} \in \mathbb{S}_{-i}$. This association maps the interior of the set of behavior strategies onto the interior of the set of mixed strategies, so the non-doctrinaire g_i generates a non-doctrinaire $\hat{g}_i$ over $-i$'s mixed strategies. Conversely, if we start with a non-doctrinaire prior $\hat{g}$ over mixed strategies in $\mathcal{G}$, then by applying Kuhn's theorem in a game with perfect recall, we can identify a non-empty set of equivalent behavior strategies for every mixed strategy. Consider the prior g_i over $-i$ behavior strategies where i believes $-i$ first draw a mixed strategy σ_{-i} according to $\hat{g}_i$, and then randomize uniformly over all behavior strategies equivalent to it. Then g_i is strictly positive on the interior because $\hat{g}_i$ enjoys the same property.

Learning with a non-doctrinaire prior $\hat{g}$ over mixed strategies in $\mathcal{G}$ with terminal node partitions $\mathcal{P}$ and learning with the same $\hat{g}$ in $\mathcal{N}$ with the $\mathcal{P}$-equivalent partitions generate the same set of steady states for every $0 \le \delta, \gamma < 1$. This is because both environments generate the same dynamic optimization problem for each agent: in both environments, they start with the same prior beliefs, receive the same payoffs for each strategy profile (s_i, s_{-i}) played, and observe the same information (up to identifying elements of the $\mathcal{P}_i$ partition with those in the equivalent $\hat{\mathcal{P}}_i$ partition.) ■

A.9 Proof of Claim 5

Proof. Suppose there is a prior g satisfying the hypotheses of the claim, parameters $\{\delta_j\}_{j\in\mathbb{N}}$, $\{\gamma_{j,k}\}_{j,k\in\mathbb{N}}$, and associated steady-state profiles $\{\pi_{j,k} \in \Pi^*(g, \delta_j, \gamma_{j,k})\}$ such that $\lim_{j\to\infty}\delta_j = 1$, $\lim_{k\to\infty}\gamma_{j,k} = 1$ for each j, and $\lim_{j\to\infty}\lim_{k\to\infty}\pi_{j,k} = \pi$.

For arbitrary $\epsilon > 0$, we will show $\pi(\mathbf{R}) < 2\epsilon$. The idea is to show that all P2 agents except the very young and those with unusual samples will have seen enough instances of P1 choosing **In2** and no instance of P1 choosing **Out** as to play **L** at their information set.

By Proposition 1 from Fudenberg, He, and Imhof (2017), there exists some $x \ge 1$ so that if a P2 agent has n observations of P1's play and in each of their observations

P1 never chose **In1,** then their mean posterior probability of P1 choosing **In1** is lower than $\frac{2x}{n+x}$. By Theorem 1 from Fudenberg, He, and Imhof (2017), there exists $N \geq 1$ such that in any steady state $\hat{\pi}$ where $\hat{\pi}(\mathbf{In2}) \geq q$, with probability at least $1-\epsilon$ a P2 agent with age at least N/q will have a mean posterior belief of P1 playing **In2** that is higher than $(1-\epsilon)q$.

Define the constant $K = \frac{16Nx}{\epsilon}$ and find some $\beta < 1$ so that, whenever $\delta \geq \beta$ and $\gamma \geq \beta$, a P1 agent will always choose **In2** in the first K periods of life. Consider any j large enough so that $\delta_j \geq \beta$. For large k, using the fact that P1 agents experiment with **In2** for at least K periods, $\pi_{j,k}(\mathbf{In2}) \geq (1+\gamma_{j,k}+...+\gamma_{j,k}^{K-1}) \cdot (1-\gamma_{j,k}) \geq \frac{1}{2}(1-\gamma_{j,k}) \cdot K$. So, a P2 agent aged at least $\frac{N}{\frac{1}{2}(1-\gamma_{j,k}) \cdot K} = (\frac{\epsilon}{2}\frac{1}{1-\gamma_{j,k}}) \cdot \frac{1}{4 \cdot x}$ has at least $1-\epsilon$ chance of believing that P1 plays **In2** with probability at least $\frac{1}{4}(1-\gamma_{j,k}) \cdot K$. This age is no larger than $\frac{\epsilon}{2}$ times the expected P2 lifespan, which contains at least $1-\epsilon$ fraction of the P2 population. Also, a P2 agent with age at least $\frac{\epsilon}{2}\frac{1}{1-\gamma_{j,k}}$ has a mean posterior belief about P1 playing **In1** that is always smaller than $\frac{2x}{\frac{\epsilon}{2}\frac{1}{1-\gamma_{j,k}}+x} = \frac{2x(1-\gamma_{j,k})}{\frac{1}{2}\epsilon+(1-\gamma_{j,k})x} \leq \frac{4x}{\epsilon}(1-\gamma_{j,k})$. Taking the ratio of the mean posterior beliefs assigned to P1 playing **In2** and **In1**, we get $\frac{\frac{1}{4}(1-\gamma_{j,k}) \cdot K}{\frac{4x}{\epsilon}(1-\gamma_{j,k})} = \frac{1}{16}K \cdot \frac{\epsilon}{x} = N \geq 1$.

Therefore, except for a mass of smaller than ϵ of P2s younger than $\frac{\epsilon}{2} \cdot \frac{1}{1-\gamma_{j,k}}$ and another mass ϵ of P2s with unusual samples, P2s respond to **In1** and **In2** with **L**. This shows in the steady state with δ_j and k large enough, $\pi_{j,k}(\mathbf{R}) < 2\epsilon$. It implies also that $\lim_{k\to\infty} \pi_{j,k}(\mathbf{R}) \leq 2\epsilon$ for all large enough j, therefore $\pi(\mathbf{R}) \leq 2\epsilon$. ■