Player-Compatible Learning and Player-Compatible Equilibrium*

Drew Fudenberg† Kevin He‡

First version: September 23, 2017
This version: May 27, 2020

Abstract

Player-Compatible Equilibrium (PCE) imposes cross-player restrictions on the magnitudes of the players' "trembles" onto different strategies. These restrictions capture the idea that trembles correspond to deliberate experiments by agents who are unsure of the prevailing distribution of play. PCE selects intuitive equilibria in a number of examples where trembling-hand perfect equilibrium (Selten, 1975) and proper equilibrium (Myerson, 1978) have no bite. We show that rational learning and weighted fictitious play imply our compatibility restrictions in a steady-state setting.

Keywords: non-equilibrium learning, equilibrium refinements, trembling-hand perfect equilibrium, weighted fictitious play.

*We thank Alessandro Bonatti, Dan Clark, Glenn Ellison, Ben Golub, Shengwu Li, Dave Rand, Alex Wolitzky, and Muhamet Yildiz for valuable conversations and comments. We thank National Science Foundation grant SES 1643517 for financial support, and Giacomo Lanzani for research assistance.

†Department of Economics, MIT. Email: drew.fudenberg@gmail.com

‡California Institute of Technology and University of Pennsylvania. Email: hesichao@gmail.com

1 Introduction

Starting with Selten (1975), a number of papers have used the device of vanishingly small "trembles" to refine the set of Nash equilibria. This paper introduces *player-compatible equilibrium* (PCE), which extends the tremble-based approach. PCE requires that the trembles used to support an equilibrium satisfy *player compatibility*, which imposes restrictions on how one player's trembles compare to those of another. We say player i is more player-compatible with strategy s_i^* than player j is with strategy s_j^* if whenever s_j^* is optimal for j against some correlated profile σ, s_i^* is optimal for i against any profile $\hat{\sigma}$ matching σ in terms of the strategies of players other than i and j. PCE is invariant to the utility representations of players' preferences over game outcomes, and provides a link between tremble-based refinements and learning-in-games. As we will explain, PCE interprets "trembles" not as errors, but as players' deliberate experiments to learn how others play. Its cross-player tremble restrictions derive from an analysis of the relative frequencies of experiments that different players choose to undertake over time under a number of commonly used learning policies.

Section 2 defines player compatibility and PCE, studies their basic properties, and proves that PCE exist in all finite games. The compatibility relation is easiest to satisfy when i and j are "non-interacting," meaning that their payoffs do not depend on each other's play. But PCE can have bite even when all players interact with each other, provided that the interactions are not too strong. Moreover, as shown by the examples in Section 3, PCE can rule out seemingly implausible equilibria that other tremble-based refinements such as trembling-hand perfect equilibrium (Selten, 1975) and proper equilibrium (Myerson, 1978) cannot eliminate.

One of these examples is a "link-formation game," where players are split into two sides, and each player decides whether or not to pay a cost to be **Active** and form links with all of the active players on the other side. Players with lower costs are more compatible with **Active** and so experiment with it more. In the "anti-monotonic" version of the game, players who incur a higher private cost of link formation give lower benefits to their linked partners; in the "co-monotonic" version, higher cost players give others higher benefits. In the anti-monotonic version the only PCE outcome is for all players to choose **Active**, because the experimentation of the low-cost players induces all players on the other side to be **Active** as well. On the other hand, both "all **Active**" and "all **Inactive"** are PCE outcomes in the co-monotonic case. In contrast to PCE making different predictions in the two versions of the game, other equilibrium refinements make the same predictions whether payoffs are anti-monotonic or co-monotonic.

To provide motivation for player-compatible trembles, we study a learning framework where agents are born into different player roles and repeatedly play the stage game. They face some time-invariant distribution of opponents' play, as they would in a steady state of a model where a continuum of anonymous agents are randomly matched each period. We compare the experimentation behavior of agents in different player roles who follow "index learning policies." These policies assign a numerical index to each strategy that only depends on data from periods when that strategy was used, and play the strategy with the highest index. We formulate an *index compatibility* condition for index policies, and use a coupling argument to show that any index policies for i and j satisfying this index-compatibility condition for strategies s_i^* and s_j^* will lead to i experimenting relatively more with s_i^* than j with s_j^* over their lifetimes against any distribution of opponents' play.

Index compatibility provides a general condition for i to choose s_i^* more often than j chooses s_j^*. This condition applies across a range of observation structures and (not necessarily optimal) learning policies. We link player compatibility with index compatibility for two canonical learning policies in a class of "factorable games." In these games, playing a strategy s_i reveals how opponents played at all the information sets that are relevant for $i's$ payoff when they play s_i, but gives no information about the payoffs of $i's$ other strategies. We show that player compatibility implies index compatibility for the rational learning policy given by the Gittins index, and for the weighted fictitious play heuristic (Cheung and Friedman, 1997). Interpreting trembles as play frequencies during a learning process, our analysis provides a learning foundation for PCE's cross-player tremble restrictions — in the link-formation game, for example, it justifies the idea that low-cost agents "tremble onto" **Active** more frequently than high-cost ones do.

1.1 Related Work

1.1.1 Tremble-Based Refinements

Tremble-based solution concepts date back to Selten (1975), who thanks Harsanyi for suggesting them. These solution concepts consider totally mixed strategy profiles where players do not play an exact best reply to the strategies of others, but may assign positive probability to some or all strategies that are not best replies. Different solution concepts in this class consider different kinds of "trembles," but they all make predictions based on the limits of these non-equilibrium strategy profiles as the probability of trembling tends to zero. Since we compare PCE to these refinements below, we summarize them here for the reader's convenience.

An *ϵ-perfect equilibrium* is a totally mixed strategy profile where every non-best reply has weight less than ϵ. A limit of ϵ_t-perfect equilibria where $\epsilon_t \to 0$ is called a *trembling-hand* *perfect equilibrium.* An *ϵ-proper equilibrium* is a totally mixed strategy profile σ where for every player i and strategies s_i and s_i' if $u_i(s_i, \sigma_{-i}) < u_i(s_i', \sigma_{-i})$ then $\sigma_i(s_i) < \epsilon \cdot \sigma_i(s_i')$. A limit of ϵ_t-proper equilibria where $\epsilon_t \to 0$ is called a *proper equilibrium;* in this limit a more costly tremble is infinitely less likely than a less costly one, regardless of the cost difference. *Approachable equilibrium* (Van Damme, 1987) is also based on the idea that strategies with worse payoffs are played less often. It too is the limit of ϵ_t-perfect equilibria, but where the players pay control costs to reduce their tremble probabilities. When these costs are "regular," all of the trembles are of the same order. Because PCE does not require that the less likely trembles are infinitely less likely than more likely ones, it is closer to approachable equilibrium than to proper equilibrium. The *strategic stability* concept of Kohlberg and Mertens (1986) is also defined using trembles, but applies to components of Nash equilibria as opposed to single strategy profiles.

Unlike PCE, proper equilibrium and approachable equilibrium do not impose cross-player restrictions on the relative probabilities of various trembles. For this reason, these equilibrium concepts reduce to perfect Bayesian equilibrium in signaling games with two possible signals, such as the beer-quiche game of Cho and Kreps (1987), when each type of the sender is viewed as a different player. They do impose restrictions when applied to the ex-ante form of the game, i.e., at the stage before the sender has learned their type. However, as Cho and Kreps (1987) point out, evaluating the cost of mistakes at the ex-ante stage of a signaling game means that the interim losses are weighted by the prior distribution over sender types, so that less likely types are more likely to tremble. In addition, applying a different positive linear rescaling to each type's utility function preserves every type's preference over lotteries on outcomes, but changes the sets of proper and approachable equilibria, while such utility rescalings have no effect on the set of PCE. In light of these issues, we always apply tremble-based refinements at the interim stage in Bayesian games.

Like PCE, *extended proper equilibrium* (Milgrom and Mollner, 2019) places restrictions on the relative probabilities of tremble by different players, but it does so in a different way: An extended proper equilibrium is the limit of $(\boldsymbol{\beta},\epsilon_t)-$proper equilibria, where $\boldsymbol{\beta} = (\beta_1, ...\beta_I)$ is a strictly positive vector of utility re-scaling, and $\sigma_i(s_i) < \epsilon_t \cdot \sigma_j(s_j)$ if player i's rescaled loss from s_i (compared to the best response) is less than j's loss from s_j. In a signaling game with only two possible signals, every Nash equilibrium where each sender type strictly prefers not to deviate from their equilibrium signal is an extended proper equilibrium at the interim stage, because suitable utility rescalings for the types can lead to any ranking of their utility costs of deviating to the off-path signal. By contrast, Proposition 4 shows every PCE must satisfy the compatibility criterion of Fudenberg and He (2018), which has bite even in binary signaling games such as the beer-quiche example of Cho and Kreps (1987). So an extended proper equilibrium need not be a PCE, a fact that Examples 1 and 2 further demonstrate. Conversely, because extended proper equilibrium makes some trembles infinitely less likely than others, it can eliminate some PCE.[1]

1.1.2 The Learning Foundations of Equilibrium

This paper builds on the work of Fudenberg and Levine (1993) and Fudenberg and Kreps (1995, 1994) on learning foundations for self-confirming and Nash equilibrium. It is also related to recent work that that provides explicit learning foundations for various equilibrium concepts that reflect ambiguity aversion, misspecified priors, or model uncertainty, such as Battigalli, Cerreia-Vioglio, Maccheroni, and Marinacci (2016), Battigalli, Francetich, Lanzani, and Marinacci (2019), Esponda and Pouzo (2016), Fudenberg, Lanzani, and Strack (2020)and Lehrer (2012). Unlike those papers, we focus on characterizing the relative rates with which different players experiment with strategies that are not myopically optimal. For this reason our analysis of learning is closer to Fudenberg and Levine (2006) and Fudenberg and He (2018).

Our investigation of learning dynamics significantly expands on that of Fudenberg and He (2018), which focused on a particular learning policy (rational Bayesians) in a restricted set of games (signaling games). In contrast, our analysis applies more broadly to any index policies that satisfy an *index compatibility* condition. We show that two strategies of i and j ranked by player compatibility lead to index-compatible learning policies in the class of "factorable games" defined in Section 5, under both rational learning and weighted fictitious play. We develop new tools to deal with new issues that arise in these more general games. For instance, Fudenberg and He (2018) compare the Gittins indices of different sender types in signaling games using the fact that any stopping time (for the auxiliary optimal-stopping problem defining the index) of the less-compatible type is also feasible for the more-compatible type. But our general setting allows player roles to interact, so it is not always valid to exchange the stopping times of two different roles. A feasible stopping time for i in the auxiliary problem only conditions on past observations of $-i$'s play, but the optimal stopping time for $j \neq i$ may condition on past observations of i's play in environments where i and j interact.

We deal with this problem by showing how i can nevertheless construct a feasible stopping time that *mimics* an infeasible one of j. Moreover, when a player faces more than one opponent, their optimal experimentation policy may lead them to observe a correlated distribution of opponents' play, even though the opponents do no actually play correlated

[1] Example available on request.

strategies. We discuss this issue of *endogenous correlation* in Section 9.2; it is the reason we define PCE in terms of correlated play.

In methodology the paper is related to other work on active learning and experimentation. In single-agent settings, these include Doval (2018), Francetich and Kreps (2020a,b), and Fryer and Harms (2017). In multi-agent settings additional issues arise such as free-riding and encouraging others to learn, see e.g., Bolton and Harris (1999), Keller et al. (2005), Klein and Rady (2011), Heidhues, Rady, and Strack (2015), Frick and Ishii (2015), Halac, Kartik, and Liu (2016), Strulovici (2010), and the survey by Hörner and Skrzypacz (2016). Unlike most models of multi-agent bandit problems, our agents only learn from personal histories, not from the actions or histories of others. Our focus is the comparison of experimentation policies under different payoff parameters, which is central to PCE's cross-player tremble restrictions.

2 Player-Compatible Equilibrium

In this section, we develop a concept of the relative "compatibility" between two player-strategy pairs and discuss its properties. We then introduce PCE, which builds cross-player tremble restrictions based on this compatibility relation into an equilibrium concept.

Like proper equilibrium, PCE is defined on the strategic form of a game. Of course many extensive forms can have the same strategic form, and the learning motivation for PCE and player-compatible trembles does depend on the underlying extensive form and the feedback structure, but we postpone these issues until Section 4.

2.1 Player Compatibility

Consider a game in its strategic form with a finite set of players $i \in \mathbb{I}$, finite strategy sets $|\mathbb{S}_i| \geq 2$,[2] and utility functions $u_i : \mathbb{S} \to \mathbb{R}$, where $\mathbb{S} := \times_i \mathbb{S}_i$. For each i, let $\Delta(\mathbb{S}_i)$ denote the set of mixed strategies for i. Let $\Delta^\circ(\mathbb{S})$ represent the interior of $\Delta(\mathbb{S})$, the set of full-support correlated strategy profiles. We now define an incomplete or partial order on strategy-player pairs.

Definition 1. For player $i \neq j$ and strategies $s_i^* \in \mathbb{S}_i$, $s_j^* \in \mathbb{S}_j$, *i is more player-compatible with s_i^* than j is with s_j^**, abbreviated as $s_i^* \succsim s_j^*$,[3] if for every totally mixed correlated

[2] If $\mathbb{S}_i = \{s_i^*\}$ is a singleton, we would have $s_i^* \succsim s_j$ and $s_j \succsim s_i^*$ for any strategy s_j of any player j if we follow the convention that the maximum over an empty set is $-\infty$.

[3] This notation is unambiguous provided i and j have disjoint strategy sets. In the event that i and j share some strategies, we will clarify this notation by attaching player subscripts.

strategy profiles $\sigma \in \Delta^\circ(\mathbb{S})$ with

$$\sum_{s \in \mathbb{S}} u_j(s_j^*, s_{-j}) \cdot \sigma(s) = \max_{s_j' \in \mathbb{S}_j} \sum_{s \in \mathbb{S}} u_j(s_j', s_{-j}) \cdot \sigma(s),$$

we get

$$\sum_{s \in \mathbb{S}} u_i(s_i^*, s_{-i}) \cdot \tilde{\sigma}(s) > \max_{s_i'' \in \mathbb{S}_i \setminus \{s_i^*\}} \sum_{s \in \mathbb{S}} u_i(s_i'', s_{-i}) \cdot \tilde{\sigma}(s)$$

for every totally mixed correlated strategy profile $\tilde{\sigma} \in \Delta^\circ(\mathbb{S})$ satisfying $\mathrm{marg}_{-ij}(\sigma) = \mathrm{marg}_{-ij}(\tilde{\sigma})$.

In words, if s_j^* is weakly optimal for the less-compatible j against σ, then s_i^* is strictly optimal for the more-compatible i against any $\tilde{\sigma}$ whose marginal on $-ij$'s play agrees with the marginal of σ. The compatibility condition does not depend on the particular expected utility functions used to represent the players' preferences over probability distributions on $\mathbb{S}$.

We argue below in Theorem 2 that play in the learning model is constrained by the compatibility relation, and that result is stronger when the compatibility relation is more complete. Since $\Delta^\circ(\mathbb{S}) \subseteq \Delta(\mathbb{S})$, the compatibility relation is more complete than an alternative definition that replaces totally mixed strategy profiles with any correlated strategy profile, and Theorem 2 would continue to hold with this alternative definition. We restrict to totally mixed strategies here to get a sharper conclusion. The restriction fits with our assumptions in the learning model that all agents have full-support prior beliefs about opponents' strategies (for rational Bayesians) or strictly positive initial counts (for weighted fictitious play). Conversely, since $\times_i \Delta^\circ(\mathbb{S}_i) \subseteq \Delta^\circ(\mathbb{S})$, our definition of compatibility ranks fewer strategy-player pairs than an alternative definition that only considers mixed strategy profiles with independent mixing between different opponents.[4] We need to use the more stringent definition to match the microfoundations of our compatibility-based cross-player restrictions: the definition that only considers independent mixing imposes restrictions that the learning model does not imply.[5]

The compatibility relation is transitive, as the next proposition shows.

Proposition 1. *Suppose $s_i^* \succsim s_j^* \succsim s_k^*$ where s_i^*, s_j^*, s_k^* are strategies of i, j, k. Then $s_i^* \succsim s_k^*$.*

The compatibility relation is also asymmetric, except in some "corner cases." Say that a

[4] Formally, this alternative definition would replace "totally mixed correlated strategy profiles" with "independently and totally mixed strategy profiles" in the definition of $s_i^* \succsim s_j^*$.

[5] One form of our microfoundation for player-compatible trembles considers rational learners who choose strategies based on their Gittins index. Even for learners who hold independent beliefs about opponents' play at different information sets, a strategy's Gittins index need not be its expected payoff against independent randomizations by the opponents, but we show that the index is always the expected payoff against some correlated strategy profile.

strategy is *strictly interior dominant* if it is strictly better than any other strategy versus any totally mixed strategy profile of the opponents, and similarly say that it is *strictly interior dominated*[6] if it is strictly dominated versus totally mixed opponent strategy profiles.

Proposition 2. *If $s_i^* \succsim s_j^*$, then at least one of the following is true: (i) $s_j^* \not\succsim s_i^*$; (ii) s_i^* is strictly interior dominated for i and s_j^* is strictly interior dominated for j; (iii) s_i^* is strictly interior dominant for i and s_j^* is strictly interior dominant for j.*[7]

The proofs of Propositions 1 and 2 are straightforward; they can be found in the Online Appendix.

The definition of player compatibility simplifies in the following special case. A game has a *multipartite structure* if the set of players $\mathbb{I}$ can be divided into C mutually exclusive classes, $\mathbb{I} = \mathbb{I}_1 \cup ... \cup \mathbb{I}_C$, in such a way that whenever i and j belong to the same class $i, j \in \mathbb{I}_c$, (1) they are *non-interacting*, meaning neither player's payoff depends on the other's strategy; and (2) they have the same strategy set, $\mathbb{S}_i = \mathbb{S}_j$. Every Bayesian game has a multipartite structure when each type is viewed as a different player. As another example, we will later use a complete-information game with a multipartite structure, the link-formation game (Example 2), to illustrate both PCE and the learning motivation for player-compatible trembles.

In a game with multipartite structure with $i, j \in \mathbb{I}_c$, we can write $u_i(s_c, s_{-ij})$ without ambiguity for $s_c \in \mathbb{S}_i$, since all augmentations of the strategy profile s_{-ij} with a strategy by player j lead to the same payoff for i. For $s_c^* \in \mathbb{S}_i = \mathbb{S}_j$, the definition of $s_{ic}^* \succsim s_{jc}^*$ reduces[8] to: For every totally mixed correlated σ with $\sigma_{-ij} \in \Delta^\circ(\mathbb{S}_{-ij})$,

$$\sum_{s \in \mathbb{S}} u_j(s_{jc}^*, s_{-ij}) \cdot \sigma(s) = \max_{s_j' \in \mathbb{S}_j} \sum_{s \in \mathbb{S}} u_j(s_j', s_{-ij}) \cdot \sigma(s)$$

implies

$$\sum_{s \in \mathbb{S}} u_i(s_{ic}^*, s_{-ij}) \cdot \sigma(s) > \max_{s_i'' \in \mathbb{S}_i \setminus \{s_{ic}^*\}} \sum_{s \in \mathbb{S}} u_i(s_i'', s_{-ij}) \cdot \sigma(s).$$

2.2 Player-Compatible Trembles and PCE

PCE is a tremble-based solution concept. It builds on and modifies Selten (1975)'s definition of trembling-hand perfect equilibrium (in the strategic form) as the limit of equilibria of

[6]Recall that a strategy can be strictly dominated even though it is not strictly dominated by any pure strategy.

[7]The converse of this statement is not true since the relation $\succsim$ is not in general complete: we could have neither $s_i^* \succsim s_j^*$ nor $s_j^* \succsim s_i^*$.

[8]We use s_{ic}^* to refer to i's copy of s_c^* and s_{jc}^* to refer to j's copy.

perturbed games in which agents are constrained to tremble, so we begin by defining our notation for the trembles and the associated constrained equilibria.

Definition 2. A *tremble profile* $\boldsymbol{\epsilon}$ assigns a positive number $\boldsymbol{\epsilon}(s_i) > 0$ to every player i and every pure strategy $s_i \in \mathbb{S}_i$. Given a tremble profile $\boldsymbol{\epsilon}$, write $\Pi_i^{\boldsymbol{\epsilon}}$ for the set of $\boldsymbol{\epsilon}$-*strategies* of player i, namely:

$$\Pi_i^{\boldsymbol{\epsilon}} := \{\sigma_i \in \Delta(\mathbb{S}_i) \text{ s.t. } \sigma_i(s_i) \geq \boldsymbol{\epsilon}(s_i) \ \forall s_i \in \mathbb{S}_i\}.$$

We call σ° an $\boldsymbol{\epsilon}$-*equilibrium* if for each i,

$$\sigma_i^\circ \in \underset{\sigma_i \in \Pi_i^{\boldsymbol{\epsilon}}}{\arg\max}\ u_i(\sigma_i, \sigma_{-i}^\circ).$$

Note that $\Pi_i^{\boldsymbol{\epsilon}}$ is compact and convex. It is also non-empty when $\boldsymbol{\epsilon}$ is close enough to $\mathbf{0}$. By standard results, whenever $\boldsymbol{\epsilon}$ is small enough so that $\Pi_i^{\boldsymbol{\epsilon}}$ is non-empty for each i, an $\boldsymbol{\epsilon}$-equilibrium exists.

The key building block for PCE is $\boldsymbol{\epsilon}$-PCE, which is an $\boldsymbol{\epsilon}$-equilibrium where the tremble profile is "co-monotonic" with $\succsim$ in the following sense:

Definition 3. Tremble profile $\boldsymbol{\epsilon}$ is *player-compatible* if $\boldsymbol{\epsilon}(s_i^*) \geq \boldsymbol{\epsilon}(s_j^*)$ for all i, j, s_i^*, s_j^* such that $s_i^* \succsim s_j^*$. An $\boldsymbol{\epsilon}$-equilibrium where $\boldsymbol{\epsilon}$ is player-compatible is called a *player-compatible* $\boldsymbol{\epsilon}$-*equilibrium* (or $\boldsymbol{\epsilon}$-*PCE*).

The condition on $\boldsymbol{\epsilon}$ says the minimum weight i could assign to s_i^* is no smaller than the minimum weight j could assign to s_j^* in the constrained game,

$$\min_{\sigma_i \in \Pi_i^{\boldsymbol{\epsilon}}} \sigma_i(s_i^*) \geq \min_{\sigma_j \in \Pi_j^{\boldsymbol{\epsilon}}} \sigma_j(s_j^*).$$

This is a "cross-player tremble restriction," that is, a restriction on the relative probabilities of trembles by different players. Note that this restriction, like the player compatibility relation, depends on the players' preferences over distributions on $\mathbb{S}$ but not on the particular utility representation. This invariance property distinguishes player-compatible trembles from other models of stochastic behavior such as the stochastic terms in logit best responses. Our learning foundation will interpret these trembles not as mistakes, but as deliberate experiments by agents trying to learn how others play.

As is usual for tremble-based equilibrium refinements, we now define PCE as the limit of a sequence of $\boldsymbol{\epsilon}$-PCE where $\boldsymbol{\epsilon} \to \mathbf{0}$.

Definition 4. A strategy profile σ^* is a *player-compatible equilibrium (PCE)* if there exists a sequence of player-compatible tremble profiles $\boldsymbol{\epsilon}^{(t)} \to \mathbf{0}$ and an associated sequence of strategy profiles $\sigma^{(t)}$, where each $\sigma^{(t)}$ is an $\boldsymbol{\epsilon}^{(t)}$-PCE, such that $\sigma^{(t)} \to \sigma^*$.

We think of PCE as primarily a solution concept for games with three or more players, where the relative tremble probabilities of $i \neq j$ affect some third party's best response. In two-player games, $s_i^* \succsim s_j^*$ only when s_j^* is strictly interior dominated or s_i^* is strictly interior dominant.

The cross-player restrictions embodied in player-compatible trembles translate into analogous restrictions on PCE, as shown in the next result.

Proposition 3. *For any PCE σ^*, player k, and strategy $\bar{s}_k$ such that $\sigma_k^*(\bar{s}_k) > 0$, there exists a sequence of totally mixed strategy profiles $\sigma_{-k}^{(t)} \to \sigma_{-k}^*$ such that*

(i) for every pair $i, j \neq k$ with $s_i^* \succsim s_j^*$,

$$\liminf_{t\to\infty} \frac{\sigma_i^{(t)}(s_i^*)}{\sigma_j^{(t)}(s_j^*)} \geq 1;$$

and (ii) $\bar{s}_k$ is a best response for k against every $\sigma_{-k}^{(t)}$.

The proof is in the Appendix, as are the proofs of subsequent results except where otherwise stated.

Treating each $\sigma_{-k}^{(t)}$ as a totally mixed approximation to σ_{-k}^*, in a PCE each player k essentially best responds to totally mixed opponent play that respects player compatibility.

It is easy to show that every $\boldsymbol{\epsilon}$-PCE respects player compatibility up to the "adding up constraint" that probabilities on different strategies must sum up to 1 and i must place probability no smaller than $\boldsymbol{\epsilon}(s_i^{'})$ on strategies $s_i^{'} \neq s_i^*$. The "up to" qualification disappears in the $\boldsymbol{\epsilon}^{(t)} \to \mathbf{0}$ limit because the required probabilities on $s_i^{'} \neq s_i^*$ tend to 0.

Since PCE is defined as the limit of $\boldsymbol{\epsilon}$-equilibria for a restricted class of trembles, the set of PCE form a subset of trembling-hand perfect equilibria; the next result shows this subset is not empty. It uses the fact that the tremble profiles with the same lower bound on the probability of each action satisfy the compatibility condition in any game.

Theorem 1. *A PCE exists in every finite strategic-form game.*

2.3 Learning and Player-Compatible Trembles

Sections 4 and 5 provide a microfoundation for player-compatible trembles, which form the core innovation of PCE. To preview the results, Sections 4 presents a general sufficient condition for i to experiment more with s_i^* than j does with s_j^* over their lifetimes that is applicable across a range of learning environments and learning policies. Sections 5 completes the story by showing that in a class of games that includes our Section 3 examples, the player-compatibility condition $s_i^* \succsim s_j^*$ implies Sections 4's sufficient condition for the rational learning policy and for weighted fictitious play. For analyzing rational behavior, we consider agents who start with the same priors over the play of their opponents. We believe we could extend this conclusion to agents with slightly different priors using a stronger notion of player compatibility, but we do not pursue this result here.[9]

PCE incorporates cross-player tremble restrictions derived from learning into an equilibrium framework. It also imposes some extra restrictions that we do not microfound, so the map from learning models to PCE is inexact, but PCE does capture some novel implications of learning. Like any game-theoretic equilibrium concept, it provides a reduced form that allows analysts to study comparative statics in various applications without needing to solve the dynamic learning problem anew in each of them.

In Section 6, we expand the game to include duplicate copies of some of the original strategies, where two strategies are duplicates if they provide exactly the same payoff and exactly the same information.[10] If $s_i^* \succsim s_j^*$ in the original game, then in the expanded game we impose the cross-player tremble restriction that the probability of i trembling onto the set of copies of s_i^* is larger than the probability of j trembling onto the set of copies of s_j^*. The way we update our PCE definition in the presence of duplicates fits our interpretation of trembles as experimentation frequencies: As we show, the sum of i's lifetime experimentation frequencies with all duplicates of s_i^* is independent of the number of duplicates under both rational behavior and weighted fictitious play. We show that the set of PCE in the expanded game with these new tremble restrictions is the same as the set of PCE in the original game.

3 Examples of PCE

In this section, we study examples of games where PCE rules out unintuitive Nash equilibria. We will also use these examples to distinguish PCE from existing refinements.

3.1 The Restaurant Game

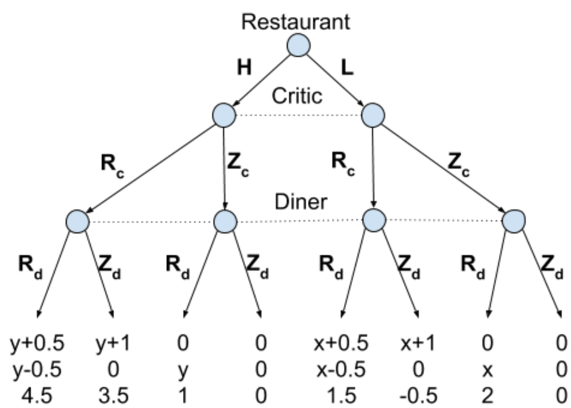
We start with a complete-information game where PCE differs from other solution concepts.

Example 1. There are three players in the game: a food critic, a regular diner, and a restaurant. Simultaneously, the restaurant decides between ordering high-quality (**H**) or

[9]To do this, we would measure the "strength" of the compatibility ranking by saying that i *is* λ *more player-compatible with* s_i^* *than* j *is with* s_j^* if the inequality in the definition $s_i^* \succsim s_j^*$ holds for all $\tilde{\sigma} \in \Delta^\circ(\mathbb{S})$ satisfying $||\text{marg}_{-ij}(\sigma) - \text{marg}_{-ij}(\tilde{\sigma})|| \leq \lambda$. We believe that our learning foundation would extend to cases where the agents' priors are sufficiently close compared to λ.

[10]Two strategies with the same payoffs that give different information about opponents' play are not equivalent in our learning model.

low-quality (**L**) ingredients, while the critic and the diner decide whether to go eat at the restaurant (**R**) or order pizza (**Z**) and eat at home. The utility from **Z** is normalized to 0. If both customers choose **Z**, the restaurant also gets 0 payoff. Otherwise, the restaurant's payoff depends on the ingredient quality and clientele. Choosing **L** yields a profit of +2 per customer while choosing **H** yields a profit of +1 per customer. In addition, if the food critic is present she will write a review based on ingredient quality, which affects the restaurant's payoff by ± 2.5. Each customer gets a payoff of $x < -1$ from consuming food made with low-quality ingredients and a payoff of $y > 0.5$ from consuming food made with high-quality ingredients, while the critic gets an additional +1 payoff from going to the restaurant and writing a review (regardless of food quality). Customers each incur a 0.5 congestion cost if they both go to the restaurant. We depict this situation in the game tree below, with c and d subscripts denoting strategies of the critic and the diner.



The strategies of the two customers affect each other's payoffs, so the critic and the diner are not non-interacting players. In particular, they cannot be mapped into two types of the same agent in a Bayesian game.

The strategy profile ($\boldsymbol{Z_c}$, $\boldsymbol{Z_d}$, **L**) is a proper equilibrium, sustained by the restaurant's belief that when at least one customer plays **R**, it is far more likely that the diner deviated to patronizing the restaurant than the critic, even though the critic has a greater incentive to go to the restaurant since she gets paid for writing reviews. It is also an extended proper equilibrium.[11]

We claim that $\boldsymbol{R_c} \succsim \boldsymbol{R_d}$. Note that for any profile σ of totally mixed, correlated play that makes the diner indifferent between $\boldsymbol{Z_d}$ and $\boldsymbol{R_d}$, we must have $u_1(\boldsymbol{R_c}, \tilde{\sigma}_{-c}) \geq 0.5$ for any profile $\tilde{\sigma}$ that agrees with σ in terms of the restaurant's play. The critic's utility from $\boldsymbol{R_c}$ is minimized when the diner chooses $\boldsymbol{R_d}$ with probability 1, but even then the critic gets

[11]($\boldsymbol{Z_c}$, $\boldsymbol{Z_d}$, **L**) is an extended proper equilibrium, because scaling the critic's payoff by a large positive constant makes it more costly for the critic to deviate to **R1** than for the diner to deviate to **R2**.

0.5 higher utility from going to a crowded restaurant than the diner gets from going to an empty restaurant, holding fixed food quality at the restaurant. This shows $\boldsymbol{R_c} \succsim \boldsymbol{R_d}$.

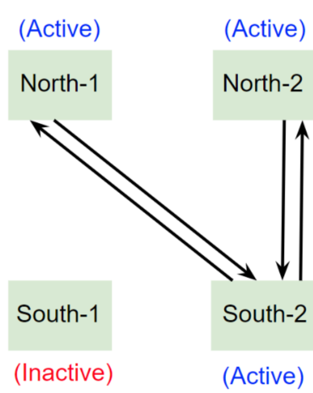
Whenever $\sigma_c^{(t)}(\boldsymbol{R_c})/\sigma_d^{(t)}(\boldsymbol{R_d}) > \frac{1}{4}$, the restaurant strictly prefers **H** over **L**. Thus by Proposition 3, there is no PCE where the restaurant plays **L** with positive probability. ♦

3.2 The Link-Formation Game

In the next example, PCE makes different predictions in two versions of a game with different payoff parameters, while all other solution concepts we know of make the same predictions in both versions.

Example 2. There are 4 players in the game, split into two sides: North and South. The players are named North-1, North-2, South-1, and South-2, abbreviated as N1, N2, S1, and S2.

These players engage in a strategic link-formation game. Each player simultaneously takes an action: either **Inactive** or **Active**. An **Inactive** player forms no links. An **Active** player forms a link with every **Active** player on the opposite side. (Two players on the same side cannot form links.) For example, suppose N1 plays **Active**, N2 plays **Active**, S1 plays **Inactive**, and S2 plays **Active**. Then N1 creates a link to S2, N2 creates a link to S2, S1 creates no links, and S2 creates links to both N1 and N2.



Each player i is characterized by two parameters: cost (c_i) and quality (q_i). Cost refers to the private cost that a player pays for each link they create. Quality refers to the benefit that a player provides to others when they link to her. A player who forms no links gets a payoff of 0. In the above example, the payoff to North-1 is $q_{\mathrm{S2}} - c_{\mathrm{N1}}$ and the payoff to South-2 is $(q_{\mathrm{N1}} - c_{\mathrm{S2}}) + (q_{\mathrm{N2}} - c_{\mathrm{S2}})$.

We consider two specifications of the payoff functions. In the *anti-monotonic* version on the left, players with a higher cost have a lower quality. In the *co-monotonic* version on the right, players with a higher cost have a higher quality. There are two pure-strategy Nash outcomes for each version: all links form or no links form. "All links form" is the unique PCE outcome in the anti-monotonic case, while both "all links" and "no links" are PCE outcomes under co-monotonicity.

Anti-Monotonic		
Player	**Cost**	**Quality**
North-1	14	30
North-2	19	10
South-1	14	30
South-2	19	10

Co-Monotonic		
Player	**Cost**	**Quality**
North-1	14	10
North-2	19	30
South-1	14	10
South-2	19	30

PCE makes different predictions in these two versions of the game because the compatibility structure with respect to own quality is reversed. In both versions, $\text{Active}_{N1} \succsim \text{Active}_{N2}$, but N1 has high quality in the anti-monotonic version, and low quality in the co-monotonic version. Thus, in the anti-monotonic version but not in the co-monotonic version, player-compatible trembles lead to the high-quality counterparty choosing **Active** at least as often as the low-quality counterparty, which means **Active** has a positive expected payoff even when one's own cost is high.

In contrast, the set of equilibria that satisfy extended proper equilibrium, proper equilibrium, trembling-hand perfect equilibrium, p-dominance, Pareto efficiency, and strategic stability do not depend on whether payoffs are anti-monotonic or co-monotonic, as shown in the Online Appendix. ♦

3.3 Signaling Games

Recall that a signaling game is a two-player Bayesian game, where P1 is a sender who knows their own type θ, and P2 only knows that P1's type is drawn according to the distribution $\lambda \in \Delta(\Theta)$ on a finite type space Θ. After learning their type, the sender sends a signal $s \in S$ to the receiver. Then, the receiver responds with an action $a \in A$. Utilities depend on the sender's type θ, the signal s, and the action a.

Fudenberg and He (2018)'s compatibility criterion is defined only for signaling games. It does not use limits of games with trembles, but instead restricts the beliefs that the receiver can have about the sender's type. That sort of restriction does not seem easy to generalize beyond games with observed actions, while using trembles allows us to define PCE for general strategic-form games. As we will see, the more general PCE definition implies the compatibility criterion in signaling games.

With each sender type viewed as a different player, this game has $|\Theta| + 1$ players, $\mathbb{I} = \Theta \cup \{2\}$, where the strategy set of each sender type θ is $\mathbb{S}_\theta = S$ while the strategy set of the receiver is $\mathbb{S}_2 = A^S$, the set of signal-contingent plans. So a mixed strategy of θ is a possibly mixed signal choice $\sigma_1(\cdot \mid \theta) \in \Delta(S)$, while a mixed strategy $\sigma_2 \in \Delta(A^S)$ of the receiver is a mixed plan about how to respond to each signal.

Fudenberg and He (2018) define *type compatibility* for signaling games. A signal s^* is more type-compatible with θ than with $\theta^{'}$ if for every behavioral strategy σ_2,

$$u_1(s^*, \sigma_2; \theta^{'}) \geq \max_{s^{'} \neq s^*} u_1(s^{'}, \sigma_2; \theta^{'})$$

implies

$$u_1(s^*, \sigma_2; \theta) > \max_{s^{'} \neq s^*} u_1(s^{'}, \sigma_2; \theta).$$

They also define the *compatibility criterion*, which imposes restrictions on off-path beliefs in signaling games. Consider a Nash equilibrium σ_1^*, σ_2^*. For any signal s^* and receiver action a with $\sigma_2^*(a \mid s^*) > 0$, the compatibility criterion requires that a best responds to some belief $p \in \Delta(\Theta)$ about the sender's type such that, whenever s^* is more type-compatible with θ than with $\theta^{'}$ and s^* is not equilibrium dominated[12] for θ, p satisfies $\frac{p(\theta^{'})}{p(\theta)} \leq \frac{\lambda(\theta^{'})}{\lambda(\theta)}$.

Since every totally mixed strategy of the receiver is payoff-equivalent to a behavioral strategy, it is easy to see that type compatibility implies $s_\theta^* \succsim s_{\theta'}^*$.[13] The next result shows that when specialized to signaling games, all PCE pass the compatibility criterion.

Proposition 4. *In a signaling game, every PCE σ^* is a Nash equilibrium satisfying the compatibility criterion of Fudenberg and He (2018).*

This proposition in particular implies that in the beer-quiche game of Cho and Kreps (1987), the quiche-pooling equilibrium is not a PCE, as it does not satisfy the compatibility criterion.

4 Index Learning Policies and Index Compatibility

This section characterizes a general class of "index learning policies" that lead i to experiment more with s_i^* than j does with s_j^*. The next section shows that optimal learning behavior and weighted fictitious play belong to this class in "factorable" games, when $s_i^* \succsim s_j^*$. Together,

[12] Signal s^* is not equilibrium dominated for θ if $\max_{a \in A} u_1(s^*, a; \theta) > u_1(s_1, \sigma_2^*; \theta)$ for every s_1 with $\sigma_1^*(s_1 \mid \theta) > 0$.

[13] The converse does not hold. We defined type compatibility to require testing against all receiver strategies and not just the totally mixed ones, so it is possible that $s_\theta^* \succsim s_{\theta'}^*$ but s^* is not more type-compatible with θ than with $\theta^{'}$, so type-compatibility is harder to satisfy than player compatibility. We now realize that we could have restricted type compatibility to only consider totally mixed strategies, and all of the results of Fudenberg and He (2018) would still hold.

these sections link the player-compatibility relation with agents' learning behavior under various learning policies, providing a learning foundation for the tremble restrictions central to PCE.

The learning problem the players face depends on what they observe about the play of others, which in turn depends on the extensive form of the game, denoted by Γ. This game has a set of players $i \in \mathbb{I}$ and also a player 0 that we will use to model Nature's moves. The collection of information sets of player $i \in \mathbb{I}$ is written as $\mathcal{H}_i$. At each $h \in \mathcal{H}_i$, player i chooses an action a_h from the finite set of possible actions A_h. An extensive-form pure strategy of i specifies an action at each information set $h \in \mathcal{H}_i$. We denote by $\mathbb{S}_i$ the set of all such strategies. Let Z be the set of terminal vertices of Γ. Also, let $Z(s)$ denote the terminal vertex reached under the pure strategy profile $s \in \times_i \mathbb{S}_i$.

We consider an agent born into player role i who maintains this role throughout their life. They have has a geometrically distributed lifetime with $0 \leq \gamma < 1$ probability of survival between periods. All agents share the same survival chance. Each period, the agent plays the stage game Γ, choosing a strategy $s_i \in \mathbb{S}_i$. Then, with probability γ, they continue into the next period and play the stage game again. With complementary probability they exit the system.

Each player is equipped with a finite set of *observations* $\mathbb{O}_i$ and a *feedback function* $\mathfrak{o}_i : Z \to \mathbb{O}_i$ that maps the terminal node reached to an observation. We assume each player has perfect recall and remembers their chosen strategy.

Definition 5. The set of all finite *histories* of all lengths for i is $Y_i := \cup_{t \geq 0} (\mathbb{S}_i \times \mathbb{O}_i)^t$. For a history $y_i \in Y_i$ and $s_i \in \mathbb{S}_i$, the *subhistory* y_{i,s_i} is the (possibly empty) subsequence of y_i where the agent played s_i.

To compare players i and j's relative experimentation probabilities, we need the agents in these two player roles to face "equivalent" learning environments. The next definition formalizes what this means; it is a joint restriction on the game tree Γ and the feedback structures $(\mathbb{O}_i, \mathfrak{o}_i)$, $(\mathbb{O}_j, \mathfrak{o}_j)$ of the two player roles.

Definition 6. Agents i and j face *isomorphic learning problems* if:

- There exists an isomorphism $\varphi : \mathbb{S}_i \to \mathbb{S}_j$
- For each $s_i \in \mathbb{S}_i$, the union $(\{s_i\} \times \mathbb{O}_i) \cup (\{s_j\} \times \mathbb{O}_j)$ of possible observations of i after s_i and possible observations of j after s_j can be partitioned into a set of equivalence classes $\mathbb{E}_{s_i,s_j}$, where elements $E \in \mathbb{E}_{s_i,s_j}$ are disjoint subsets.
- For each pure strategy profile $\tilde{s}$ and $s_i \in \mathbb{S}_i$ with $s_j = \varphi(s_i)$, $(s_i, \mathfrak{o}_i(Z(s_i, \tilde{s}_{-i})))$ and $(s_j, \mathfrak{o}_j(Z(s_j, \tilde{s}_{-j})))$ belong to the same equivalence class in $\mathbb{E}_{s_i,s_j}$.

Whenever we say i and j face isomorphic learning problems, it is always with respect to some isomorphism φ and some collection of equivalence classes $\mathbb{E} = (\mathbb{E}_{s_i,\varphi(s_i)})_{s_i \in S_i}$. To make this explicit, sometimes we will say i and j face $(\varphi, \mathbb{E})$-*isomorphic learning problems*.

If i and j face isomorphic learning problems, there is a bijection φ between the strategies of the two agents and a notion of equivalence between the one-period histories they might observe after choosing s_i and s_j, with $s_j = \varphi(s_i)$. This equivalence puts i's observation after $(s_i, \tilde{s}_{-i})$ in the same equivalence class as j's observation after $(s_j, \tilde{s}_{-j})$, for any profile $\tilde{s}$.

Consider Example 1 when the Critic and the Diner observe all other players' actions if they choose **R**, but observe nothing if they choose **Z**. That is,

$$\mathbb{O}_C = \mathbb{O}_D = \{(L,R),(L,Z),(H,R),(H,Z),\varnothing\}.$$

Consider the natural isomorphism $\varphi(R_c) = R_d$ and $\varphi(Z_c) = Z_d$, and partition histories after R_c and R_d as

$$\begin{aligned}\mathbb{E}_{R_c,R_d} = &\{\{(R_c,(L,R)),(R_d,(L,R)),(R_c,(L,Z)),(R_d,(L,Z))\}, \\ &\{(R_c,(H,R)),(R_d,(H,R)),(R_c,(H,Z)),(R_d,(H,Z))\}\}.\end{aligned}$$

The two equivalence classes in $\mathbb{E}_{R_c,R_d}$ represent whether the restaurant is observed to play **L** or **H**. Also let $\mathbb{E}_{Z_c,Z_d} = \{\{(Z_c,\varnothing),(Z_d,\varnothing)\}\}$ contain just one equivalence class, which corresponds to no observations. It is clear that given any pure strategy profile s, (R_c, s_{-c}) and (R_d, s_{-d}) lead to the same histories, up to equivalence classes (i.e., the same observation of the restaurant's play.)

We extend the notion of equivalence to histories with more than one period in the natural way.

Definition 7. Suppose i and j face $(\varphi, \mathbb{E})$-isomorphic learning problems with $\varphi(s_i) = s_j$. Say i's subhistory y_{i,s_i} is *equivalent* to j's subhistory y_{j,s_j}, written as $y_{i,s_i} \sim y_{j,s_j}$, if they are equivalent period by period according to the equivalence classes $\mathbb{E}_{s_i,\varphi(s_i)}$.

Equivalence of y_{i,s_i} and y_{j,s_j} says i has played s_i as many times as j has played s_j, and that the sequence of observations that i encountered from experimenting with s_i are the "same" as those that j encountered from experimenting with s_j.

The following histories for the critic and the diner of the restaurant game are equivalent for the strategy R (under the $\varphi, \mathbb{E}$ previously given). This equivalence arises because the subhistories $y_{c,R}$ and $y_{d,R}$ contain the same sequences of the restaurant's play (even though the two agents have different observations in terms of how often the other patron goes to the restaurant).

	period	1	2	3	4	5
y_c:	own strategy	R	Z	Z	Z	R
	observation	(L, Z)	$\varnothing$	$\varnothing$	$\varnothing$	(H, Z)
y_d:	own strategy	Z	R	Z	R	
	observation	$\varnothing$	(L, R)	$\varnothing$	(H, Z)	

Table 1: The two histories y_c (for the critic, with length 5) and y_d (for the diner, with length 4) have equivalent subhistories for R.

The agent decides which strategy to use in each period based on their history so far. We assume that this *learning policy* is a deterministic map (which is w.l.o.g. for expected-utility maximizers), and denote it $r_i : Y_i \to \mathbb{S}_i$.

Definition 8. A learning policy is an *index policy* if there are *index functions* $(\iota_{s_i})_{s_i \in \mathbb{S}_i}$ with each ι_{s_i} mapping subhistories of s_i to real numbers, such that $r_i(y_i) \in \arg\max_{s_i \in \mathbb{S}_i} \{\iota_{s_i}(y_{i,s_i})\}$.

If an agent uses an index policy, we can think of their behavior in the following way. At each history, they compute an index for each strategy $s_i \in \mathbb{S}_i$ based on the subhistory of those periods where they chose s_i, and play a strategy with the highest index.[14] The best-known example of an index policy is the Gittins index (Gittins, 1979). Some heuristics for learning problems, such as weighted fictitious play (Cheung and Friedman, 1997), are also index policies. The key restriction in an index policy is that each strategy's index depends only on the observations when that strategy was played. Note that index policies are deterministic, unlike some heuristics such as Thompson sampling (Thompson, 1933).

Finally, we define a notion of the relative compatibility of index policies r_i and r_j with various strategies.

Definition 9. Suppose i and j face $(\varphi, \mathbb{E})$-isomorphic learning problems with $\varphi(s_i^*) = s_j^*$. For two index policies r_i and r_j, r_i is *more index-compatible* with s_i^* than r_j is with s_j^* if for any histories y_i, y_j and any strategy $s_i' \in \mathbb{S}_i$, $s_i' \neq s_i^*$ satisfying

- $y_{i,s_i^*} \sim y_{j,s_j^*}$ and $y_{i,s_i'} \sim y_{j,\varphi(s_i')}$
- s_j^* has weakly the highest index for j,

we have that s_i' does not have the weakly highest index for i.

The key result of this section, Proposition 5, shows that index compatibility is a sufficient condition for average agents in the role of player i to play s_i^* more frequently than those

[14] To handle possible ties, we can introduce a strict order over each agent's strategy set, and specify that if two strategies have the same index the agent plays the one that is higher ranked.

in the j-role play s_j^*. This result is not obvious because the index-compatibility relation only applies when two agents have equivalent histories, which typically does not hold during the dynamic process of experimentation. To formalize the result, suppose that each agent is randomly matched with opponents in other player roles, and that the play of these opponents corresponds to a fixed mixed strategy profile σ, as it would in a steady state. We call σ the *social distribution*. It, together with the agent's learning policy, generates a stochastic process X_i^t describing i's strategy in period t; denote its distribution by $\mathbb{P}_{r_i,\sigma}$.

Definition 10. Let X_i^t be the $\mathbb{S}_i$-valued random variable representing i's play in period t. Player i's *discounted lifetime play under* the social distribution σ and learning policy r_i is $\phi_i(\cdot; r_i, \sigma) : \mathbb{S}_i \to [0, 1]$, where for each $s_i \in \mathbb{S}_i$

$$\phi_i(s_i; r_i, \sigma) := (1 - \gamma) \sum_{t=1}^{\infty} \gamma^{t-1} \cdot \mathbb{P}_{r_i,\sigma}\{X_i^t = s_i\}.$$

Each newborn player i expects to play each of s_i a share $\phi_i(s_i; r_i, \sigma)$ of their lifetime.

Proposition 5. *Suppose i and j face $(\varphi, \mathbb{E})$-isomorphic learning problems with $\varphi(s_i^*) = s_j^*$, and r_i, r_j are index policies, with r_i more index-compatible with s_i^* than r_j is with s_j^*. Then $\phi_i(s_i^*; r_i, \sigma_{-i}) \geq \phi_j(s_j^*; r_j, \sigma_{-j})$ for any $0 \leq \gamma < 1$ and $\sigma \in \times_k \Delta(\mathbb{S}_k)$.*

Note that Proposition 5 holds whenever there exists any $(\varphi, \mathbb{E})$ such that i, j face $(\varphi, \mathbb{E})$-isomorphic learning problems and r_i, r_j satisfy index-compatibility with respect to $(\varphi, \mathbb{E})$.

The proof extends the coupling argument in the proof of Fudenberg and He (2018)'s Lemma 2, which only applies to the Gittins index in signaling games, and also fills in a missing step (Lemma 5) that the earlier proof implicitly assumed. To deal with the issue that i and j learn from endogenous data that diverge as they undertake different experiments, we couple the learning problems of i and j using what we call *response paths* $\mathfrak{S} \in ((\mathbb{S})^N)^\infty$ where $N = \max_i |\mathbb{S}_i|$. We can think of $\mathfrak{S}$ as a two-dimensional array of strategy profiles, $\mathfrak{S} = ((\mathfrak{S}_{1,1}, \mathfrak{S}_{1,2}, ..., \mathfrak{S}_{1,N}), (\mathfrak{S}_{2,1}, \mathfrak{S}_{2,2}, ..., \mathfrak{S}_{2,N}), ...)$, where $\mathfrak{S}_{t,n_i} \in \mathbb{S}$ for every $t \geq 1$, $1 \leq n_i \leq N$. We may enumerate each player's strategy set $\mathbb{S}_i$ and interchangeably refer to each strategy $s_i \in \mathbb{S}_i$ with its assigned number $n_{s_i} \in \{1, ..., N\}$. For a given path and learning policy r_i for player i, imagine running the policy against the data-generating process where the t-th time i plays the n_i-th strategy in $\mathbb{S}_i$, i is matched up with opponents who play the strategies $\mathfrak{S}_{t,n_i}$. Given a learning policy r_i, each $\mathfrak{S}$ induces a deterministic infinite history of i's strategies $y_i(\mathfrak{S}, r_i) \in (\mathbb{S}_i)^\infty$. (For $n_i > |\mathbb{S}_i|$, the values of $(\mathfrak{S}_{t,n_i})_{t \geq 1}$ do not matter for the induced history.) We show that under the hypothesis that r_i is more index-compatible with s_i^* than r_j is with s_j^*, the weighted lifetime frequency of s_i^* in $y_i(\mathfrak{S}, r_i)$ is larger than the frequency of s_j^* in $y_j(\mathfrak{S}, r_j)$ for every $\mathfrak{S}$, where play in different periods of the infinite histories $y_i(\mathfrak{S}, r_i), y_j(\mathfrak{S}, r_j)$ are weighted by the probabilities of surviving into these periods, just as in the definition of discounted lifetime play.

Lemma 5 in the Appendix shows that when i and j face i.i.d. draws of opponents' play from a fixed learning environment σ, the discounted lifetime play are the same as if they each faced a random response path $\mathfrak{S}$ drawn at birth according to the (infinite) product measure over $((\mathbb{S})^N)^\infty$ whose marginals correspond to $\sigma^{|\mathbb{S}_i|}$.

5 Index Compatibility and Player Compatibility in Factorable Games

Section 4 proves that whenever index-strategy pairs (r_i, s_i^*) and (r_j, s_j^*) satisfy index compatibility, and i and j face isomorphic learning problems, index policy r_i uses s_i^* more often than r_j uses s_j^* against any social distribution σ. Index compatibility is a joint restriction on the agents' learning policy and the game's feedback structure $(\mathbb{O}, \mathfrak{o})$, which gives the domain that the learning policies are defined on. This section shows that player compatibility implies index compatibility for rational behavior and weighted fictitious play in a class of *factorable* games. Factorability applies to the examples discussed in Section 3 for players ranked by compatibility.

5.1 Factorability and Isomorphic Factoring

In factorable games, agent i's observation is just their utility: $\mathfrak{o}_i(s_i, s_{-i}) = u_i(s_i, s_{-i})$. In general, i's payoff $u_i(s_i, s_{-i})$ needs not reveal the actions that others' strategies s_{-i} pick at all $-i$ information sets in the game tree. The definition of factorability puts restrictions on the extensive-form game tree Γ to discipline what i can learn from own payoffs.

Factorability says that playing strategy $s_i \in \mathbb{S}_i$ against any strategy profile of $-i$ identifies all of the opponents' actions that can be payoff-relevant for s_i, but does not reveal any information about the payoff consequences of choosing any other strategy $s_i' \neq s_i$ against the social distribution. From i's perspective, it is as if the game tree can be "factored" into disjoint parts based on the elements of $\mathbb{S}_i$, and playing each $s_i \in \mathbb{S}_i$ lets i learn how s_{-i} play at all payoff-relevant $-i$ information sets in the s_i-part of the game tree, but provides no information about s_{-i} in any other part of the tree. We now make this idea formal.

For an information set h of j with $j \neq i$, write P_h for the partition on $\mathbb{S}_{-i}$ where two strategy profiles s_{-i}, s'_{-i} are in the same element of the partition if they prescribe the same play on h. That is, the partition elements in P_h are $\{s_{-i} \in \mathbb{S}_{-i} : s_{-i}(h) = a_h\}$ for $a_h \in A_h$. Thus partition P_h contains perfect information about play on h, but no other information.

Definition 11. For each player i and strategy $s_i \in \mathbb{S}_i$, let $\Pi_i[s_i]$ be the coarsest partition of $\mathbb{S}_{-i}$ that makes $s_{-i} \mapsto u_i(s_i, s_{-i})$ measurable. The game Γ is *factorable for i* if:

1. For each $s_i \in \mathbb{S}_i$ there exists a (possibly empty) collection of $-i$'s information sets $F_i[s_i] \subseteq \mathcal{H}_{-i}$ so that $\Pi_i[s_i] = \bigvee_{h \in F_i[s_i]} P_h$. (The meet across an empty collection is the coarsest possible partition on $\mathbb{S}_{-i}$, i.e. no information).

2. For two strategies $s_i \neq s_i'$, $F_i[s_i] \cap F_i[s_i'] = \emptyset$.

When Γ is factorable for i, we refer to $F_i[s_i]$ as the *s_i-relevant information sets*, a terminology we now justify. In general, i's payoff from playing s_i can depend on the profile of $-i$'s actions at all opponent information sets. Condition (1) implies that only opponents' actions on $F_i[s_i]$ matter for i's payoff after choosing s_i, and furthermore this dependence is one-to-one. That is,

$$u_i\left(s_i, s_{-i}\right) = u_i\left(s_i, s_{-i}'\right) \Leftrightarrow \left(\forall h \in F_i\left[s_i\right] \quad s_{-i}\left(h\right) = s_{-i}'\left(h\right)\right).$$

The one-to-one mapping from s_{-i} to i's payoff implies that i's learning cannot be blocked by another player: By choosing s_i, i can always use their own payoffs to identify actions on $F_i[s_i]$ regardless of what happens elsewhere in the game tree.[15] It also shows that if Γ is factorable for i, then $F_i[s_i]$ are uniquely defined for all s_i. Suppose there were two collections $(F_i[s_i])_{s_i \in \mathbb{S}_i}$ and $(\tilde{F}_i[s_i])_{s_i \in \mathbb{S}_i}$ with $F_i[s_i] \backslash \tilde{F}_i[s_i] \neq \varnothing$ for some $s_i \in \mathbb{S}_i$ that both satisfy Condition (1) of Definition 11. Then there are two $-i$ profiles s_{-i}, s_{-i}' that match on $\tilde{F}_i[s_i]$ but not on $F_i[s_i]$. But then we get both $u_i(s_i, s_{-i}) = u_i(s_i, s_{-i}')$ and $u_i(s_i, s_{-i}) \neq u_i(s_i, s_{-i}')$, a contradiction.

Condition (2) implies that i cannot extrapolate the payoff consequence of a different strategy $s_i' \neq s_i$ through playing s_i (provided i's prior is independent about opponents' play on different information sets). This is because there is no intersection between the s_i-relevant information sets and the s_i'-relevant ones. In particular this means that player i cannot "free ride" on others' experiments and learn about the consequences of various risky strategies while playing a safe one that is myopically optimal.

In short, Condition (1) ensures i gets information about play in the same part of the game tree every time she plays s_i (instead of learning about play in two different parts of the tree depending on someone else's strategy), while Condition (2) guarantees that there is no interaction between learning about different strategies.

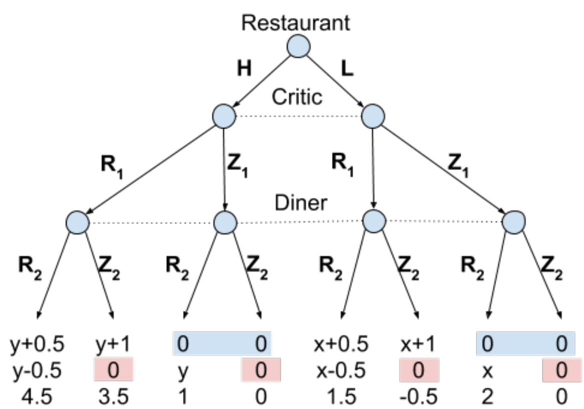
If $F_i[s_i]$ is empty, then s_i is a kind of "opt out" action for i. After choosing s_i, i receives the same utility from every reachable terminal node and gets no information about the payoff consequences of any of their other strategies.

[15] It is easy but expositionally costly to extend this to the case where several actions on A_h lead to the same payoff for i.

5.1.1 Examples of Factorable Games

We now illustrate factorability using the examples from Section 3 and some other general classes of games.

The Restaurant Game Consider the restaurant game from Example 1. Since $x < -1$ and $y > 0.5$, we have $x \neq y$ and $x \neq y + 0.5$. By choosing **R**, the customer's payoff perfectly reveals others' play. By choosing **Z**, the customer always gets 0 payoff (these nodes are colored in the diagram below) and so cannot infer anyone else's play.



The restaurant game is factorable for the Critic and the Diner. Let $F_i[\mathbf{R}_i]$ consist of the two information sets of $-i$ and let $F_i[\mathbf{Z}_i]$ be the empty set for each $i \in \{1, 2\}$. It is easy to verify that the two conditions of factorability are satisfied.

It is important for factorability that a customer who takes the "outside option" of ordering pizza gets the same payoff regardless of the restaurant's play, and does not observe the restaurant's quality choice even if the other customer patronizes the restaurant. Factorability rules out this sort of "free information," so that when we analyze the non-equilibrium learning problem we know that each agent can only learn a strategy's payoff consequences by playing it themselves.

The Link-Formation Game Consider the link-formation game from Example 2. The payoff for a player choosing **Inactive** is always 0, whereas the payoff for a player choosing **Active** exactly identifies the play of the two players on the opposite side. We can let $F_i[\mathbf{Active}_i]$ consist of the information sets of the other two agents on the other side of i and let $F_i[\mathbf{Inactive}_i]$ be empty. This specification of the s_i-relevant information sets shows the stage game is factorable for every player.

Binary Participation Games More generally, Γ is factorable for i whenever it is a *binary participation game* for i.

Definition 12. Γ is a *binary participation game* for i if the following are satisfied.

1. i has a unique information set with two actions, without loss labeled **In** and **Out.**

2. All paths of play in Γ pass through i's information set.

3. All paths of play where i plays **In** pass through the same information sets.

4. Terminal vertices associated with i playing **Out** all have the same payoff for i.

5. Terminal vertices associated with i playing **In** all have different payoffs for i.

Action **Out** is an outside option for i that leads to a constant payoff regardless of others' play. We are implicitly assuming in part (5) of the definition that the game has generic payoffs for i after choosing **In**, in the sense that changing the action at any one information set on the path of play will change i's payoff.

If Γ is a binary participation game for i, let $F_i[\mathbf{In}]$ be the collection of $-i$ information sets encountered in paths of play where i chooses **In**. Let $F_i[\mathbf{Out}]$ be the empty set. We see that Γ is factorable for i. Clearly $F_i[\mathbf{In}] \cap F_i[\mathbf{Out}] = \emptyset$, so Condition (2) of factorability is satisfied. When i chooses the strategy **In**, the tree structure of Γ implies different profiles of play on $F_i[\mathbf{In}]$ must lead to different terminal nodes, and the generic payoff condition means Condition (1) of factorability is satisfied for strategy **In**. When i plays **Out**, i gets the same payoff regardless of the others' play, so Condition (1) of factorability is satisfied for strategy **Out**.

The restaurant game is a binary participation game for the critic and the diner, where ordering pizza is the outside option. The link-formation game is a binary participation game for every player, where **Inactive** is the outside option.

Signaling to Multiple Audiences To give a different class of examples of factorable games, consider a game of signaling to one or more audiences. To be precise, Nature moves first and chooses a type for the sender, drawn according to some known distribution over a finite set of types, Θ. The sender then chooses a signal $s \in S$, observed by all receivers $r_1, ..., r_{n_r}$. Each receiver then simultaneously chooses an action. The profile of receiver actions, together with the sender's type and signal, determine payoffs for all players. Viewing different types of senders as different players, this game is factorable for all sender types, provided payoffs are generic. This factorability arises because for each type i, $F_i[s]$ is the set of n_r information sets for the receivers after seeing signal s.

5.1.2 Examples of Non-Factorable Games

The next result gives a necessary condition for factorability, which we then use to provide examples of non-factorable games. Suppose H is an information set of player $j \neq i$. Player i's payoff is *independent* of h if $u_i(a_h, a_{-h}) = u_i(a'_h, a_{-h})$ for all a_h, a'_h, a_{-h}, where a_h, a'_h are actions on information set h, and a_{-h} is a profile of actions on all other information sets in the game tree. In games where each player moves at most once along any path of play, if i's payoff is not independent of the action taken at some information set h, then i can always put h onto the path of play via a unilateral deviation at one of their information sets.

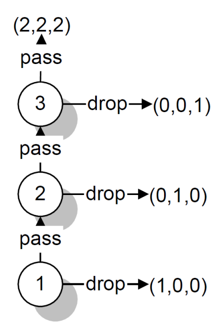
Proposition 6. *Suppose each player moves at most once along any path of play in Γ, the game is factorable for i, and i's payoff is not independent of h^*. For any strategy profile, either h^* is on the path of play, or i has a deviation at one of their information sets that puts h^* onto the path of play.*

This result follows from two lemmas.

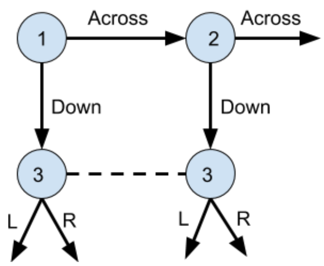
Lemma 1. *For any game that is factorable for i and any information set h^* for player $j \neq i$ where j has at least two different actions, if $h^* \in F_i[s_i]$ for some extensive-form strategy $s_i \in \mathbb{S}_i$, then h^* is always on the path of play when i chooses s_i.*

Lemma 2. *For any game that is factorable for i and any information set h^* of player $j \neq i$, suppose i's payoff is not independent of h^*. Then 1) j has at least two different actions on h^*; and (2) there exists some extensive-form strategy $s_i \in \mathbb{S}_i$ so that $h^* \in F_i[s_i]$.*

We can combine these two lemmas to prove the proposition. Consider the centipede game for three players below.



Each player moves at most once on each path, and 1 and 2's payoffs are not independent of the (unique) information set of player 3. But, if both 1 and 2 choose "drop", then no one step deviation by either 1 or 2 can put the information set of 3 onto the path of play. Proposition 6 thus implies the centipede game is not factorable for either 1 or 2. Moreover, Fudenberg and Levine (2006) showed that in this game even very patient player 2's may not learn to play a best response to player 3, so that the strategy profile (drop, drop, pass) can persist even though it is not trembling-hand perfect. Intuitively, if the player 1's only play pass as experiments, then when the fraction of new players is very small, the player 2's may not get to play often enough to make experimentation with pass worthwhile.



As another example, the Selten's horse game displayed above is not factorable for 1 or 2 if the payoffs are generic, even though the conclusion of Proposition 6 is satisfied. The information set of 3 must belong to both F_1[Down] and F_1[Across] because 3's play can affect 1's payoff even if 1 chooses Across, since 2 could choose Down. This violates the factorability requirement that $F_1[\text{Down}] \cap F_1[\text{Across}] = \varnothing$. The same argument shows the information set of 3 must belong to both F_2[Down] and F_2[Across], since when 1 chooses Down the play of 3 affects 2's payoff regardless of 2's play. So, again, $F_2[\text{Down}] \cap F_2[\text{Across}] = \varnothing$ is violated.

Condition (2) of factorability also rules out games where i has two strategies that give the same information, but one strategy always has a worse payoff under all profiles of opponents' play. In this case, we can think of the worse strategy as an informationally equivalent but more costly experiment than the better strategy. Reasonable learning policies (including rational learning) will not use such strategies, but we do not capture this feature in the general definition of PCE because our setup there only considers abstract strategy spaces $\mathbb{S}_i$ and not an extensive-form game tree.[16]

5.1.3 Isomorphic Factoring

In order to compare the learning behavior of agents i and j, it is not enough that the game is factorable for each of them. We must also ensure that they face isomorphic learning problems in order to apply Proposition 5. To do this we define the notion of *isomorphic factoring.*

Definition 13. When Γ is factorable for both i and j, the factoring is *isomorphic* for i and j if there exists a bijection $\varphi : \mathbb{S}_i \to \mathbb{S}_j$ such that $F_i[s_i] \cap \mathcal{H}_{-ij} = F_j[\varphi(s_i)] \cap \mathcal{H}_{-ij}$ for every $s_i \in \mathbb{S}_i$.

[16]It would be interesting to try to refine the definition of PCE to capture this, perhaps using the "signal function" approach of Battigalli and Guaitoli (1997) and Rubinstein and Wolinsky (1994).

This says the s_i-relevant information sets (for i) are the same as the $\varphi(s_i)$-relevant information sets (for j), insofar as the actions of $-ij$ are concerned. For example, the restaurant game is isomorphically factorable for the critic and the diner (under the isomorphism $\varphi(\mathbf{R1})=\mathbf{R2}$, $\varphi(\mathbf{Z1})=\mathbf{Z2}$) because $F_1[\mathbf{R1}] \cap \mathcal{H}_3 = F_2[\mathbf{R2}] \cap \mathcal{H}_3 =$ the singleton set containing the unique information set of the restaurant. As another example, all signaling games (with possibly many receivers as in Section 5.1.1) are isomorphically factorable for the different types of the sender. Similarly, the link-formation game is isomorphically factorable for pairs (N1, N2), and (S1, S2), but note that it is not isomorphically factorable for (N1, S1).

Factorability and isomorphic factoring let us construct $(\varphi, \mathbb{E})$ so that i and j face $(\varphi, \mathbb{E})$-isomorphic learning problems. The equivalence classes $\mathbb{E}_{s_i,\varphi(s_i)}$ for $s_i \in \mathbb{S}_i$ are such that $(s_i, u_i(s_i, \tilde{s}_{-i})) \sim (\varphi(s_i), u_j(\varphi(s_i), \hat{s}_{-j})$ if and only if $\tilde{s}_{-i}|_{F_i[s_i]\cap\mathcal{H}_{-ij}} = \hat{s}_{-j}|_{F_j[\varphi(s_i)]\cap\mathcal{H}_{-ij}}$.

5.2 Rational Learning in Factorable Games

We first consider rational agents who maximize expected discounted payoffs. This learning rule requires two additional elements: a Bayesian prior belief over others' play and a discount factor. We assume that each agent i starts with a *regular independent prior:*

Definition 14. Agent i has a *regular independent prior* if their beliefs g_i on $\times_{h\in\mathcal{H}_{-i}}\Delta(A_h)$ can be written as the product of full-support marginal densities on $\Delta(A_h)$ across different $h \in \mathcal{H}_{-i}$, so that $g_i((\alpha_h)_{h\in\mathcal{H}_{-i}}) = \prod_{h\in\mathcal{H}_{-i}} g_i^h(\alpha_h)$ with $g_i^h(\alpha_h) > 0$ for all $\alpha_h \in \Delta^\circ(A_h)$.

Thus, the agent thinks the social distribution assigns an unknown mixed action at each $-i$ information set, and thinks actions at different $-i$ information sets are generated independently from these underlying mixed actions, whether the information sets belong to the same player or to different players. Furthermore, the agent holds independent beliefs about the mixed actions at different information sets.[17] The agent updates g_i by applying Bayes rule to their history y_i. If the stage game is a signaling game, for example, this independence assumption means that the senders only update their beliefs about the receiver's response to a given signal based on the responses received to that signal, and that the senders' beliefs about this response do not depend on the responses they have observed to other signals.

[17] We assume that agents do not know Nature's mixed actions, which must be learned just as the play of other players. If agents know Nature's move, then a regular independent prior would be a density g_i on $\times_{h\in\mathcal{H}_{\mathbb{I}\setminus\{i\}}}\Delta(A_h)$, so that $g_i((\alpha_h)_{\mathcal{H}_{\mathbb{I}\setminus\{i\}}}) = \prod_{h\in\mathcal{H}_{\mathbb{I}\setminus\{i\}}} g_i^h(\alpha_h)$ with $g_i^h(\alpha_h) > 0$ for all $\alpha_h \in \Delta^\circ(A_h)$. As Fudenberg and Kreps (1993) point out, an agent who believes two opponents are randomizing independently may nevertheless have subjective correlation in their uncertainty about the randomizing probabilities of these opponents. Here we study the natural special case where the agents' prior beliefs about the opponents are independent, i.e., a product measure.

In addition to the survival chance $0 \leq \gamma < 1$ between periods, the agent further discounts future payoffs according to their patience $0 \leq \delta < 1$, so their overall effective discount factor is $0 \leq \delta\gamma < 1$. All agents have the same patience level.

Given a belief about the distribution of play at each opponent information set, we can calculate the Gittins index of each strategy $s_i \in \mathbb{S}_i$. Let $\nu_{s_i} \in \times_{h \in F_i[s_i]} \Delta(\Delta(A_h))$ be a belief over opponents' mixed actions at the information sets in $F_i[s_i]$. The Gittins index of s_i under belief ν_{s_i} is given by the maximum value of the following auxiliary optimization problem:

$$\sup_{\tau \geq 1} \frac{\mathbb{E}_{\nu_{s_i}} \left\{ \sum_{t=1}^{\tau} (\delta\gamma)^{t-1} \cdot u_i(s_i, (a_h(t))_{h \in F_i[s_i]}) \right\}}{\mathbb{E}_{\nu_{s_i}} \left\{ \sum_{t=1}^{\tau} (\delta\gamma)^{t-1} \right\}},$$

where the supremum is taken over all positive-valued stopping times $\tau \geq 1$. Here $(a_h(t))_{h \in F_i[s_i]}$ means the profile of actions that $-i$ plays on $F_i[s_i]$ the t-th time that i uses s_i — by assumption about factorable games, only these actions and not actions elsewhere in the game tree determine i's payoff from playing s_i, and i can always infer these actions from their own payoffs. The distribution over the infinite sequence of profiles $(a_h(t))_{t=1}^{\infty}$ is given by i's belief ν_{s_i}, that is, there is some fixed mixed action in $\times_{h \in F_i[s_i]} \Delta(A_h)$ that generates profiles $(a_h(t))$ i.i.d. across periods t. The event $\{\tau = T\}$ for $T \geq 1$ corresponds to using s_i for T periods, observing the first T elements $(a_h(t))_{t=1}^{T}$, then stopping.

A learning policy that chooses a strategy s_i with the highest Gittins index after each history y_i solves the rational agent's dynamic optimization problem. We denote any such policy as OPT_i, suppressing its dependence on δ and g_i.

5.3 Weighted Fictitious Play in Factorable Games

Next we consider the weighted fictitious play heuristic, a generalization of Brown (1951)'s fictitious play.[18] Agent i keeps track of *counts* for actions at the opponent information sets in the game tree,

$$\{N_h^{a_h} \in \mathbb{R}_{++} : h \in \mathcal{H}_{-i}, a_h \in A_h\}.$$

The $N_h^{a_h}$ values of a newcomer agent start at some *initial counts*, $N_h^{a_h}(0) > 0$, and the counts update as i learns.

After history y_i of i where s_i has been used $T \geq 0$ times, i's subhistory for s_i can be viewed as $y_{i,s_i} = (s_i, s_{-i}^{(t)}(h)_{h \in F_{-i}[s_i]})_{t=1}^{T}$ where $s_{-i}^{(t)}(h)_{h \in F_{-i}[s_i]}$ is the observed $-i$'s play on $F_i[s_i]$ the t-th time that s_i was used. (This is because there is a one-to-one relationship between s_{-i}'s play on $F_i[s_i]$ and $u_i(s_i, s_{-i})$.) The updated count on (h, a_h) for $h \in F_i[s_i]$ and $a_h \in A_h$

[18] This heuristic was first estimated on lab data by Cheung and Friedman (1997). It was generalized by Camerer and Ho (1999) and later analyzed by Benaïm, Hofbauer, and Hopkins (2009).

is

$$N_h^{a_h}(y_i) = \sum_{t=1}^{T} \mathbf{1}(s_{-i}^{(t)}(h) = a_h) \cdot \rho^{(T-t)} + \rho^T N_h^{a_h}(0)$$

for some $\rho \in (0, 1]$. That is, i calculates a weighted sum for the total number of times that $-i$ have played a_h in the history y_i, where past observations on $F_i[s_i]$ are discounted at a rate ρ between successive uses of the strategy s_i. All agents share the same weight factor ρ.

Following history y_i, i assigns an index to s_i equal to its expected payoff when opponents play the mixed action $\alpha_h(a_h; y_i) = \frac{N_h^{a_h}(y_i)}{\sum_{a'_h \in A_h} N_h^{a'_h}(y_i)}$ on information sets $h \in F_i[s_i]$. Write WFP_i for a learning policy that chooses a strategy with the highest weighted fictitious play index after every history (suppressing its dependence on ρ and the initial counts $\{N_h^{a_h}(0)\}$).

When $\rho = 1$, the counts are updated according to the unweighted fictitious play, and the limit of $\rho \to 0$ corresponds to myopically best replying to the observed play when each strategy was most recently used. The special case of the Gittins index where the prior g_i marginalized to each $\Delta(A_h)$ is a Dirichlet distribution and $\delta = 0$ is equivalent to the special case of unweighted fictitious play (i.e., $\rho = 1$) with some initial counts that depend on the Dirichlet priors' parameters. In general OPT_i differs from WFP_i outside of these corner cases.

5.4 Player-Compatibility Implies Index-Compatibility of OPT and WFP under Isomorphic Factoring

The main result of this paper, Theorem 2, shows that if $s_i^* \succsim s_j^*$ in a game isomorphically factorable for i and j with $\varphi(s_i^*) = s_j^*$, then i uses s_i^* more frequently than j uses s_i^* both under rational experimentation and under weighted fictitious play. This comparison holds under the hypothesis that i and j start their learning processes with the same "initial conditions." For OPT, this means i's prior g_i marginalized to the s_i-relevant $-ij$ information sets equals to j's prior g_j marginalized to the $\varphi(s_i)$-relevant $-ij$ information sets for every $s_i \in \mathbb{S}_i$. For WFP, this means i and j start with the same initial counts about $-ij$'s actions.

Theorem 2. *Suppose $s_i^* \succsim s_j^*$ and the game is isomorphically factorable for i and j with $\varphi(s_i^*) = s_j^*$. For any common survival chance $0 \le \gamma < 1$ and any social distribution σ, we have $\phi_i(s_i^*; r_i, \sigma_{-i}) \ge \phi_j(s_j^*; r_j, \sigma_{-j})$ under either of the following conditions:*

- $r_i = \mathrm{OPT}_i$, $r_j = \mathrm{OPT}_j$, and priors g_i, g_j are regular and equivalent[19]: that is, they satisfy $g_i|_{\Delta(A_h):h \in F_i[s_i] \cap \mathcal{H}_{-ij}} = g_j|_{\Delta(A_h):h \in F_j[\varphi(s_i)] \cap \mathcal{H}_{-ij}}$ for every $s_i \in \mathbb{S}_i$.

[19]The theorem easily generalizes to the case where i starts with one of $L \ge 2$ possible priors $g_i^{(1)}, ..., g_i^{(L)}$

- $r_i = \text{WFP}_i$, $r_j = \text{WFP}_j$, and i and j have the same initial counts $N_h^{a_h}(0)$ for every $s_i \in \mathbb{S}_i$, $h \in F_i[s_i] \cap \mathcal{H}_{-ij}$, and $a_h \in A_h$.

The proof works by showing that if $s_i^* \succsim s_j^*$, then both rational learning and weighted fictitious play are $(\varphi, \mathbb{E})$-index compatible. This then lets us apply Proposition 5's general conclusion about index compatible learning policies.

6 Replication Invariance of PCE

Fix a *base game* where each i has a finite strategy set $\mathbb{S}_i$ and utility function $u_i : \mathbb{S} \to \mathbb{R}$.

Definition 15. An *extended game with duplicates* is a game with the same players as the base game, where $\bar{\mathbb{S}}_i$ the set of strategies of i is a finite subset of $\mathbb{S}_i \times \mathbb{N}$ such that for all $(s_i, n_i) \in \bar{\mathbb{S}}_i$ and $(s_j, n_j)_{j \neq i} \in \bar{\mathbb{S}}_{-i}$, the payoff in the new game is $\bar{u}_i((s_i, n_i), (s_j, n_j)_{j\neq i}) = u_i(s_i, s_j)$. For every $s_i \in \mathbb{S}_i$, there exists some $n_i \in \mathbb{N}$ so that $(s_i, n_i) \in \bar{\mathbb{S}}_i$.

The interpretation is that i has some copies of every strategy they had in the base game, and could have different numbers of copies of different strategies, where duplicate copies of the same strategy have the same payoff consequences. For the learning foundation we will now assume that the extended game is factorable, which implies that each duplicate provides the same information. As an example, suppose that in the Restaurant Game the critic can arrive at the restaurant by taking the red bus or the blue bus, and the color of the bus is not observed by other players, does not change anyone's payoffs, and does not change what the diner observes. We can then replace R_c with two actions R_c^{red}, R_c^{blue} at every node in the diner's information set, letting R_c^{red} and R_c^{blue} both have the same payoff consequences as R_c in the original game. This modified game is an extended game with duplicates for the original game.

Subsection 6.1 defines player-compatible trembles and PCE in extended games with duplicates. Using the compatibility relation $\succsim$ from the base game, a tremble profile in the extended game with duplicates is player compatible if the *sum* of tremble probabilities assigned to all copies of s_i^* exceeds the sum assigned to all copies of s_j^*, whenever $s_i^* \succsim s_j^*$. PCE is then defined using this restriction on trembles. We show that the set of PCE in the base game coincides with the set of PCE in the extended game with duplicates.

This definition of player-compatible trembles in extended games with duplicates fits with our interpretation of trembles as experimentation frequencies and an analysis of how learning

with probabilities $p_1, ..., p_L$ and j starts with priors $g_j^{(1)}, ..., g_j^{(L)}$ with the same probabilities, and each $g_i^{(l)}, g_j^{(l)}$ is a pair of equivalent regular priors for $1 \leq l \leq L$.

dynamics in the extended game compare with those in the base game. The idea is that if all copies of a strategy s_i give i the same information about others' play, then i should be exactly indifferent between all such copies after all histories in the learning process. Holding fixed initial beliefs and the social distribution, i's weighted lifetime average play of s_i in the base game should then equal the sum of their weighted lifetime average plays of all copies of s_i in the extended game with duplicates. Thus, any comparisons that hold between the "tremble" probabilities of i onto s_i^* and j onto s_j^* in the base game must also hold between the "tremble" probabilities of i onto the copies of s_i^* and j onto the copies of s_j^* in the extended game. We formalize this intuition in binary participation games in Subsection 6.2 for rational learning and weighted fictitious play.

6.1 PCE in Extended Games with Duplicates

A *tremble profile of the extended game* $\bar{\boldsymbol{\epsilon}}$ assigns a positive number $\bar{\boldsymbol{\epsilon}}(s_i, n_i) > 0$ to every player i and every pure strategy $(s_i, n_i) \in \bar{\mathbb{S}}_i$. We define $\bar{\boldsymbol{\epsilon}}$-*strategies* of i and $\bar{\boldsymbol{\epsilon}}$-*equilibrium* of the extended game in the usual way, relative to the strategy sets $\bar{\mathbb{S}}_i$.

Definition 16. Tremble profile $\bar{\boldsymbol{\epsilon}}$ is *player-compatible in the extended game* if $\sum_{n_i} \bar{\boldsymbol{\epsilon}}(s_i^*, n_i) \geq \sum_{n_j} \bar{\boldsymbol{\epsilon}}(s_j^*, n_j)$ for all i, j, s_i^*, s_j^* such that $s_i^* \succsim s_j^*$. An $\bar{\boldsymbol{\epsilon}}$-equilibrium where $\bar{\boldsymbol{\epsilon}}$ is player-compatible is called a *player-compatible* $\bar{\boldsymbol{\epsilon}}$-*equilibrium* (or $\bar{\boldsymbol{\epsilon}}$-*PCE*).

We now relate $\bar{\boldsymbol{\epsilon}}$-equilibria in the extended game to $\boldsymbol{\epsilon}$-equilibria in the base game. Recall the following constrained optimality condition that applies to both the extended game and the base game:

Fact 1. *A feasible mixed strategy of i is* ***not*** *a constrained best response to a $-i$ profile if and only if it assigns more than the required weight to a non-optimal response.*

We associate with a strategy profile $\bar{\sigma} \in \times_{i\in\mathbb{I}}\Delta(\bar{\mathbb{S}}_i)$ in the extended game a *consolidated strategy profile* $\mathscr{C}(\bar{\sigma}) \in \times_{i\in\mathbb{I}}\Delta(\mathbb{S}_i)$ in the base game, given by adding up the probabilities assigned to all copies of each base-game strategy. More precisely, $\mathscr{C}(\bar{\sigma})_i(s_i) := \sum_{n_i} \bar{\sigma}_i(s_i, n_i)$. Similarly, $\mathscr{C}(\bar{\boldsymbol{\epsilon}})$ is the consolidated tremble profile, given by $\mathscr{C}(\bar{\boldsymbol{\epsilon}})(s_i) := \sum_{n_i} \bar{\boldsymbol{\epsilon}}(s_i, n_i)$.

Conversely, given a strategy profile $\sigma \in \times_{i\in\mathbb{I}}\Delta(\mathbb{S}_i)$ in the base game, the extended strategy profile $\mathscr{E}(\sigma) \in \times_{i\in\mathbb{I}}\Delta(\bar{\mathbb{S}}_i)$ is defined by $\mathscr{E}(\sigma)_i(s_i, n_i) := \sigma_i(s_i)/N(s_i)$ for each $i, (s_i, n_i) \in \bar{\mathbb{S}}_i$, where $N(s_i)$ is the number of copies of s_i that $\bar{\mathbb{S}}_i$ contains. Similarly, $\mathscr{E}(\boldsymbol{\epsilon})$ is the extended tremble profile, given by $\mathscr{E}(\boldsymbol{\epsilon})(s_i, n_i) := \boldsymbol{\epsilon}(s_i)/N(s_i)$.

Lemma 3. *If $\bar{\sigma}$ is an $\bar{\boldsymbol{\epsilon}}$-equilibrium in the extended game, then $\mathscr{C}(\bar{\sigma})$ is an $\mathscr{C}(\bar{\boldsymbol{\epsilon}})$-equilibrium in the base game. If σ is an $\boldsymbol{\epsilon}$-equilibrium in the base game, then $\mathscr{E}(\sigma)$ is an $\mathscr{E}(\boldsymbol{\epsilon})$-equilibrium in the extended game.*

The proof of results in this section can be found in the Online Appendix.

PCE is defined as usual in the extended game.

Definition 17. A strategy profile $\bar{\sigma}^*$ is a *player-compatible equilibrium (PCE) in the extended game* if there exists a sequence of player-compatible tremble profiles $\bar{\boldsymbol{\epsilon}}^{(t)} \to \mathbf{0}$ and an associated sequence of strategy profiles $\bar{\sigma}^{(t)}$, where each $\bar{\sigma}^{(t)}$ is an $\bar{\boldsymbol{\epsilon}}^{(t)}$-PCE, such that $\bar{\sigma}^{(t)} \to \bar{\sigma}^*$.

These PCE correspond exactly to PCE of the base game.

Proposition 7. *If $\bar{\sigma}^*$ is a PCE in the extended game, then $\mathscr{C}(\bar{\sigma}^*)$ is a PCE in the base game. If σ^* is a PCE in the base game, then $\mathscr{E}(\sigma^*)$ is a PCE in the extended game.*

In fact, starting from a PCE σ^* of the base game, we can construct more PCE of the extended game than $\mathscr{E}(\sigma^*)$ by shifting around the probabilities assigned to different copies of the same base-game strategy, but all these profiles essentially correspond to the same outcome.

6.2 Learning and Trembles in Binary Participation Games with Duplicates

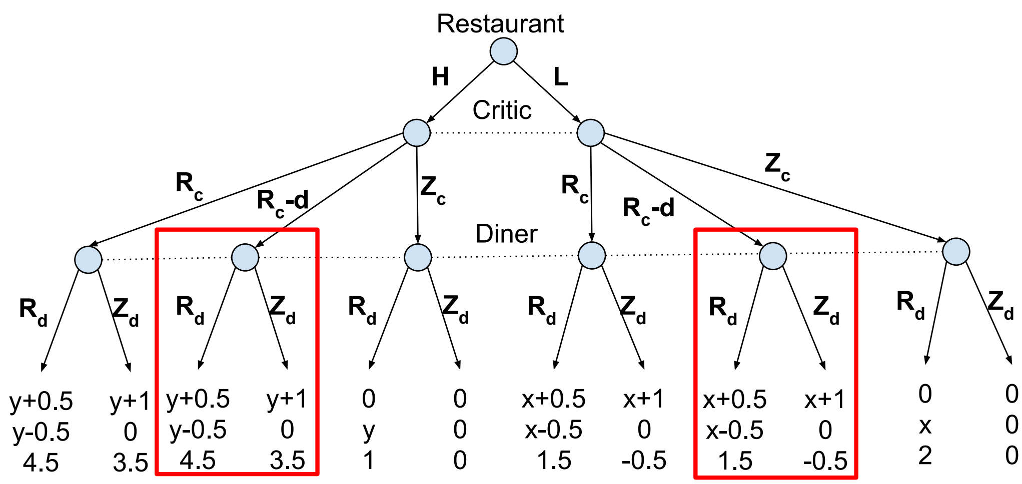
We give the simplest illustration of how learning dynamics in extended games with duplicates relate to those in the base game, using binary participation games. These results can also be developed for other factorable games, but at the cost of more complicated notation.

Consider a binary participation game for i (Definition 12) as the base game and create an extended game with duplicates by adding an extra copy of the **In** strategy for i to the game tree, called **In-d**. We show that when r_i is an optimal learning policy for i or the weighted fictitious play heuristic, the discounted lifetime play $\phi_i(\mathbf{In}; r_i, \sigma_{-i})$ for the base game is equal to the sum $\phi_i(\mathbf{In}; r_i, \sigma_{-i}) + \phi_i(\mathbf{In\text{-}d}; r_i, \sigma_{-i})$ in the new game, for the same social distribution σ.

We modify the original game tree Γ and information sets $\mathcal{H}$ to arrive at a new game tree $\bar{\Gamma}$ with information sets $\bar{\mathcal{H}}$. The basic idea is that **In-d** gives the same payoffs and information to i, and $-i$ cannot tell which one i chose.

By the definition of a binary participation game for i, let h_i be i's unique information set in $\mathcal{H}$. Enumerate the vertices in h_i as $h_i = \{v_1, ..., v_n\}$. Playing **In** at vertex v_k in the original tree leads to some subtree $\Gamma^{(k)} \subseteq \Gamma$. Start with $\bar{\Gamma} = \Gamma$ and add a new move, **In-d**, to every $v_k \in h_i$. Append a new subtree $\hat{\Gamma}^{(k)}$ to $\bar{\Gamma}$ for every $v_k \in h_i$, such that $\hat{\Gamma}^{(k)}$ is a copy of $\Gamma^{(k)}$ (including payoffs at terminal vertices) and playing **In-d** at v_k leads to $\hat{\Gamma}^{(k)}$. Now we give a procedure to construct the information sets $\bar{\mathcal{H}}$ to capture the idea that **In** and **In-d** are indistinguishable to others. Start with $\bar{\mathcal{H}} = \mathcal{H}$ and let $V^{(k)}$ be the set of vertices in $\Gamma^{(k)}$. For every $1 \le k \le n$ and $v \in V^{(k)}$, find the information set $h \in \bar{\mathcal{H}}$ with $v \in h$, then put $h := h \cup \{\tilde{v}\}$, where $\tilde{v}$ is the copy of v in $\hat{\Gamma}^{(k)}$. That is, each vertex reachable after i chooses **In-d** is indistinguishable to others from its "twin" reachable when i chooses **In**.

As discussed before, the Restaurant Game is a binary participation game for the critic and the diner, with going to the restaurant as **In** and ordering pizza as **Out**. We illustrate adding a duplicate copy of R_c for the critic to the game, labeled $R_c - d$. The critic's unique information set contains two vertices, and the new game tree adds two new subtrees to the original game, highlighted in red.



The set of histories in the learning framework for i with the extended stage game is $\tilde{Y}_i = \cup_{t \ge 0}(\{\mathbf{In}, \mathbf{In\text{-}d}, \mathbf{Out}\} \times \mathbb{R})^t$. We now define a notion of equivalence between a *stochastic* learning policy in the extended game $\tilde{r}_i : \tilde{Y}_i \to \Delta(\{\mathbf{In}, \mathbf{In\text{-}d}, \mathbf{Out}\})$ and a (deterministic) learning policy in the original game, $r_i : Y_i \to \{\mathbf{In}, \mathbf{Out}\}$. Basically, $\tilde{r}_i$ behaves just like r_i except it can randomize between **In** and **In-d**.

Definition 18. Let $\zeta : \tilde{Y}_i \to Y_i$ be such that for $\tilde{y}_i \in \tilde{Y}_i$, $\zeta(\tilde{y}_i) \in Y_i$ replaces every instance of instance of **In-d** with **In**. Learning policies $\tilde{r}_i : \tilde{Y}_i \to \Delta(\{\mathbf{In}, \mathbf{In\text{-}d}, \mathbf{Out}\})$ and $r_i : Y_i \to \{\mathbf{In}, \mathbf{Out}\}$ are *equivalent up to duplicates* if for every $\tilde{y}_i \in \tilde{Y}_i$, if $r_i(\zeta(\tilde{y}_i)) = \mathbf{Out}$, then also $\tilde{r}_i(\tilde{y}_i)(\mathbf{Out}) = 1$. If $r_i(\zeta(\tilde{y}_i)) = \mathbf{In}$, then $\tilde{r}_i(\tilde{y}_i)(\mathbf{In}) + \tilde{r}_i(\tilde{y}_i)(\mathbf{In\text{-}d}) = 1$.

The main result of this section shows that rational learning and weighted fictitious play lead to learning policies that are equivalent up to duplicates in the base game and the extended game. Furthermore, any pair of such equivalent policies in the two settings lead to the same lifetime discounted frequencies of playing **In** for the original game as playing **In** and **In-d** for the extended game against the same social distributions of $-i$.

Technically, extensive-form strategies in $(\Gamma, \mathcal{H})$ and $(\bar{\Gamma}, \bar{\mathcal{H}})$ are defined over two different domains. To make sense of i facing the "same" social distribution of $-i$'s play in the two settings, let $\psi : \bar{\mathcal{H}} \to \mathcal{H}$ be the natural isomorphism between the two collections of information sets. Each information set $\tilde{h}$ in the modified game is either equal to an information set $h \in \mathcal{H}$, or it is an old information set with some extra vertices added, that is there is some (unique) h with $\tilde{h} \supsetneq h$. Let $\psi(\tilde{h}) := h$. Two strategy profiles $\sigma, \tilde{\sigma}$ for $(\Gamma, \mathcal{H})$ and $(\bar{\Gamma}, \bar{\mathcal{H}})$ are $-i$ *equivalent* if $\tilde{\sigma}(\tilde{h}) = \sigma(\psi(\tilde{h}))$ for all $\tilde{h} \in \tilde{\mathcal{H}}_{-i}$.

Proposition 8. *Suppose stochastic learning policy $\tilde{r}_i$ in the extended game is equivalent up to duplicates with the learning policy r_i in the base game.*

- *For a fixed patience parameter $0 \leq \delta < 1$ and regular prior g_i over others' play,*[20] *r_i is OPT_i if and only if $\tilde{r}_i$ is an optimal learning policy with the extended game.*
- *For a fixed decay parameter $0 \leq \rho < 1$ and initial counts $N_h^{a_h}(0)$, r_i is WFP_i if and only if after every $\tilde{y}_i \in \tilde{Y}_i$, $\tilde{r}_i(\tilde{y}_i)$ is supported on strategies that maximize payoffs under the weighted fictitious play conjecture of $-i$'s play.*
- *For $-i$ equivalent social distributions $\sigma, \tilde{\sigma}$ for the base game and extended games, $\phi_i(\boldsymbol{In}; r_i, \sigma_{-i}) = \phi_i(\boldsymbol{In}; \tilde{r}_i, \tilde{\sigma}_{-i}) + \phi_i(\boldsymbol{In\text{-}d}; \tilde{r}_i, \tilde{\sigma}_{-i})$.*

7 Concluding Discussion

PCE makes two key contributions. First, it generates new and sensible restrictions on equilibrium play by imposing cross-player restrictions on the relative probabilities that different players assign to certain strategies — namely, those strategy pairs s_i, s_j ranked by the player-compatibility relation $s_i \succsim s_j$. As we have shown through examples, these cross-player restrictions distinguish PCE from other refinement concepts and allows us to make comparative statics predictions in some games where other equilibrium refinements do not.

Second, PCE shows how restricted "trembles" can capture some of the implications of non-equilibrium learning. PCE's cross-player restrictions arise endogenously for a general class of index learning policies, which includes both the standard model of Bayesian agents maximizing their expected discounted lifetime utility, and computationally tractable heuristics like weighted fictitious play. We conjecture that the result that i is more likely to experiment with s_i than j is with s_j when $s_i \succsim s_j$ applies in other natural models of learning

[20]The prior is over $\times_{h \in \mathcal{H}_{-i}} \Delta(A_h)$ in the original stage game and over $\times_{\tilde{h} \in \tilde{\mathcal{H}}_{-i}} \Delta(A_{\tilde{h}})$ in the extended game, but we identify $\Delta(A_{\tilde{h}})$ with $\Delta(A_{\psi(\tilde{h})})$ for each $\tilde{h} \in \tilde{H}_{-i}$. The same identification applies for the initial counts in the original and extended games.

or dynamic adjustment, such as those considered by Francetich and Kreps (2020a,b), and that it may be possible to provide foundations for PCE in other and perhaps larger classes of games.

The strength of the PCE refinement depends on the completeness of the compatibility order $\succsim$, since $\boldsymbol{\epsilon}$-PCE imposes restrictions on i and j's play only when the relation $s_i \succsim s_j$ holds. Our player compatibility definition supposes that player i thinks all mixed strategies of other players are possible, as it considers the set of all totally mixed correlated strategies $\sigma_{-i} \in \Delta^\circ(\mathbb{S}_{-i})$. If the players have some prior knowledge about their opponents' utility functions, player i might deduce *a priori* that the other players will only play strategies in some subset of $\Delta^\circ(\mathbb{S}_{-i})$. As we show in Fudenberg and He (2020), in signaling games imposing this kind of prior knowledge leads to a more complete version of the compatibility order. It may similarly lead to a more refined version of PCE.

PCE is defined for general strategic forms. We have only provided learning foundations for player-compatible trembles in factorable games. Moreover, even in factorable games, PCE imposes some extra restrictions that we do not microfound, but we view this as a first step in connecting together tremble-based refinement concepts with learning-in-games. As we have shown through the link-formation game and other examples, PCE is a convenient reduced form that generates novel comparative statics predictions in various applications without needing the analyst to solve the dynamic learning problem anew in each of them.

References

Battigalli, P., S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci (2016): "Analysis of information feedback and selfconfirming equilibrium," *Journal of Mathematical Economics*, 66, 40–51.

Battigalli, P., A. Francetich, G. Lanzani, and M. Marinacci (2019): "Learning and self-confirming long-run biases," *Journal of Economic Theory*, 183, 740–785.

Battigalli, P. and D. Guaitoli (1997): "Conjectural equilibria and rationalizability in a game with incomplete information," in *Decisions, Games and Markets*, Springer, 97–124.

Benaïm, M., J. Hofbauer, and E. Hopkins (2009): "Learning in games with unstable equilibria," *Journal of Economic Theory*, 144, 1694–1709.

Bolton, P. and C. Harris (1999): "Strategic experimentation," *Econometrica*, 67, 349–374.

Brown, G. W. (1951): "Iterative solution of games by fictitious play," *Activity Analysis of Production and Allocation*, 13, 374–376.

Camerer, C. and T.-H. Ho (1999): "Experience-weighted Attraction Learning in Normal Form Games," *Econometrica*, 67, 827–874.

Cheung, Y.-W. and D. Friedman (1997): "Individual learning in normal form games: Some laboratory results," *Games and Economic Behavior*, 19, 46–76.

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

Doval, L. (2018): "Whether or not to open Pandora's box," *Journal of Economic Theory*, 175, 127–158.

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

Francetich, A. and D. Kreps (2020a): "Choosing a good toolkit, I: Prior-free heuristics," *Journal of Economic Dynamics and Control*, 111, 103813.

——— (2020b): "Choosing a good toolkit, II: Bayes-rule based heuristics," *Journal of Economic Dynamics and Control*, 111, 103814.

Frick, M. and Y. Ishii (2015): "Innovation adoption by forward-looking social learners," *Working Paper.*

Fryer, R. and P. Harms (2017): "Two-armed restless bandits with imperfect information: Stochastic control and indexability," *Mathematics of Operations Research*, 43, 399–427.

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.

Fudenberg, D. and D. M. Kreps (1993): "Learning Mixed Equilibria," *Games and Economic Behavior*, 5, 320–367.

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

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

Fudenberg, D., G. Lanzani, and P. Strack (2020): "Limits Points of Endogenous Misspecified Learning," *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.

Gittins, J. C. (1979): "Bandit Processes and Dynamic Allocation Indices," *Journal of the Royal Statistical Society. Series B (Methodological)*, 148–177.

Halac, M., N. Kartik, and Q. Liu (2016): "Optimal contracts for experimentation," *Review of Economic Studies*, 83, 1040–1091.

Heidhues, P., S. Rady, and P. Strack (2015): "Strategic experimentation with private payoffs," *Journal of Economic Theory*, 159, 531–551.

Hörner, J. and A. Skrzypacz (2016): "Learning, experimentation and information design," in *Advances in Economics and Econometrics: Eleventh World Congress*, ed. by B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge University Press, chap. 2, 63–97.

Jackson, M. O. and A. Wolinsky (1996): "A strategic model of social and economic networks," *Journal of Economic Theory*, 71, 44–74.

Keller, G., S. Rady, and M. Cripps (2005): "Strategic experimentation with exponential bandits," *Econometrica*, 73, 39–68.

Klein, N. and S. Rady (2011): "Negatively correlated bandits," *Review of Economic Studies*, 78, 693–732.

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

Lehrer, E. (2012): "Partially specified probabilities: decisions and games," *American Economic Journal: Microeconomics*, 4, 70–100.

Milgrom, P. and J. Mollner (2019): "Extended Proper Equilibrium," *Working Paper*.

Monderer, D. and L. S. Shapley (1996): "Potential games," *Games and Economic Behavior*, 14, 124–143.

Myerson, R. B. (1978): "Refinements of the Nash equilibrium concept," *International Journal of Game Theory*, 7, 73–80.

Pearce, D. G. (1984): "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, 52, 1029–1050.

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

Selten, R. (1975): "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory*, 4, 25–55.

Strulovici, B. (2010): "Learning while voting: Determinants of collective experimentation," *Econometrica*, 78, 933–971.

Thompson, W. R. (1933): "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, 25, 285–294.

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

Appendix

8 Proofs of Results Stated in the Main Text

8.1 Proof of Proposition 3

We first state an auxiliary lemma.

Lemma 4. If σ° is an $\boldsymbol{\epsilon}$-PCE and $s_i^* \succsim s_j^*$, then

$$\sigma_i^\circ(s_i^*) \geq \min\left[\sigma_j^\circ(s_j^*), 1 - \sum_{s_i^{'} \neq s_i^*} \boldsymbol{\epsilon}(s_i^{'})\right].$$

Proof. Suppose $\boldsymbol{\epsilon}$ is player-compatible and let $\boldsymbol{\epsilon}$-equilibrium σ° be given. For $s_i^* \succsim s_j^*$, suppose $\sigma_j^\circ(s_j^*) = \boldsymbol{\epsilon}(s_j^*)$. Then $\sigma_i^\circ(s_i^*) \geq \boldsymbol{\epsilon}(s_i^*) \geq \boldsymbol{\epsilon}(s_j^*) = \sigma_j^\circ(s_j^*)$, where the second inequality comes from $\boldsymbol{\epsilon}$ being player-compatible. On the other hand, suppose $\sigma_j^\circ(s_j^*) > \boldsymbol{\epsilon}(s_j^*)$. Since σ° is an $\boldsymbol{\epsilon}$-equilibrium, the fact that j puts more than the minimum required weight on s_j^* implies s_j^* is at least a weak best response for j against σ°, with σ° totally mixed due to the trembles.The definition of $s_i^* \succsim s_j^*$ then implies that s_i^* must be a strict best response for i against σ° as well. In the $\boldsymbol{\epsilon}$-equilibrium, i must assign as much weight to s_i^* as possible, so that $\sigma_i^\circ(s_i^*) = 1 - \sum_{s_i^{'} \neq s_i^*} \boldsymbol{\epsilon}(s_i^{'})$. Combining these two cases establishes the desired result. □

Proposition 3: *For any PCE σ^*, player k, and strategy $\bar{s}_k$ such that $\sigma_k^*(\bar{s}_k) > 0$, there exists a sequence of totally mixed strategy profiles $\sigma_{-k}^{(t)} \to \sigma_{-k}^*$ such that*

(i) for every pair $i, j \neq k$ with $s_i^* \succsim s_j^*$,

$$\liminf_{t\to\infty} \frac{\sigma_i^{(t)}(s_i^*)}{\sigma_j^{(t)}(s_j^*)} \geq 1;$$

and (ii) $\bar{s}_k$ is a best response for k against every $\sigma_{-k}^{(t)}$.

Proof. By Lemma 4, for every $\boldsymbol{\epsilon}^{(t)}$-PCE we get

$$\begin{aligned}
\frac{\sigma_i^{(t)}(s_i^*)}{\sigma_j^{(t)}(s_j^*)} &\geq \min\left[\frac{\sigma_j^{(t)}(s_j^*)}{\sigma_j^{(t)}(s_j^*)}, \frac{1 - \sum_{s_i^{'} \neq s_i^*} \boldsymbol{\epsilon}^{(t)}(s_i^{'}|i)}{\sigma_j^{(t)}(s_j^*)}\right] \\
&= \min\left[1, \frac{1 - \sum_{s_i^{'} \neq s_i^*} \boldsymbol{\epsilon}^{(t)}(s_i^{'}|i)}{\sigma_j^{(t)}(s_j^*)}\right] \geq 1 - \sum_{s_i^{'} \neq s_i^*} \boldsymbol{\epsilon}^{(t)}(s_i^{'}|i).
\end{aligned}$$

This says

$$\inf_{t\geq T} \frac{\sigma_i^{(t)}(s_i^*)}{\sigma_j^{(t)}(s_j^*)} \geq 1 - \sup_{t\geq T} \sum_{s_i^{'} \neq s_i^*} \boldsymbol{\epsilon}^{(t)}(s_i^{'}|i).$$

For any sequence of trembles such that $\boldsymbol{\epsilon}^{(t)} \to \mathbf{0}$, $\lim_{T\to\infty} \sup_{t\geq T} \sum_{s_i' \neq s_i^*} \boldsymbol{\epsilon}^{(t)}(s_i'|i) = 0$, so

$$\liminf_{t\to\infty} \frac{\sigma_i^{(t)}(s_i^*)}{\sigma_j^{(t)}(s_j^*)} = \lim_{T\to\infty} \left\{ \inf_{t\geq T} \frac{\sigma_i^{(t)}(s_i^*)}{\sigma_j^{(t)}(s_j^*)} \right\} \geq 1.$$

This shows that if we fix a PCE σ^* and consider a sequence of player-compatible trembles $\boldsymbol{\epsilon}^{(t)}$ and $\boldsymbol{\epsilon}^{(t)}$-PCE $\sigma^{(t)} \to \sigma^*$, then each $\sigma_{-k}^{(t)}$ satisfies $\liminf_{t\to\infty} \sigma_i^{(t)}(s_i^*)/\sigma_j^{(t)}(s_j^*) \geq 1$ whenever $i, j \neq k$ and $s_i^* \succsim s_j^*$. Furthermore, from $\sigma_k^*(\bar{s}_k) > 0$ and $\sigma_k^{(t)} \to \sigma_k^*$, we know there is some $T_1 \in \mathbb{N}$ so that $\sigma_k^{(t)}(\bar{s}_k) > \sigma_k^*(\bar{s}_k)/2$ for all $t \geq T_1$. We may also find $T_2 \in \mathbb{N}$ so that $\boldsymbol{\epsilon}^{(t)}(\bar{s}_k|k) < \sigma_k^*(\bar{s}_k)/2$ for all $t \geq T_2$, since $\boldsymbol{\epsilon}^{(t)} \to \mathbf{0}$. So when $t \geq \max(T_1, T_2)$, $\sigma_k^{(t)}$ places strictly more than the required weight on $\bar{s}_k$, so $\bar{s}_k$ is at least a weak best response for k against $\sigma_{-k}^{(t)}$. Now the subsequence of opponent play $(\sigma_{-k}^{(t)})_{t\geq\max(T_1,T_2)}$ satisfies the requirement of this proposition. □

8.2 Proof of Theorem 1

Theorem 1: *PCE exists in every finite strategic-form game.*

Proof. Consider a sequence of tremble profiles with the same lower bound on the probability of each strategy, that is $\boldsymbol{\epsilon}^{(t)}(s_i|i) = \epsilon^{(t)}$ for all i and s_i, and with $\epsilon^{(t)}$ decreasing monotonically to 0 in t. Each of these tremble profiles is player-compatible (regardless of the compatibility structure $\succsim$) and there is some finite T large enough that $t \geq T$ implies an $\boldsymbol{\epsilon}^{(t)}$-equilibrium exists, and some subsequence of these $\boldsymbol{\epsilon}^{(t)}$-equilibria converges since the space of strategy profiles is compact. By definition these $\boldsymbol{\epsilon}^{(t)}$-equilibria are also $\boldsymbol{\epsilon}^{(t)}$-PCE, which establishes existence of PCE. □

8.3 Proof of Proposition 4

Proposition 4: *In a signaling game, every PCE σ^* is a Nash equilibrium satisfying the compatibility criterion of Fudenberg and He (2018).*

Proof. Since every PCE is a trembling-hand perfect equilibrium and since this latter solution concept refines Nash, σ^* is a Nash equilibrium. To show that it satisfies the compatibility criterion, we need to show that σ_2^* assigns probability 0 to plans in A^S that, for some $s \in S$, do not best respond to an "admissible" belief at signal s under profile σ^*. For any plan assigned positive probability under σ_2^*, by Proposition 3 we may find a sequence of totally mixed signal profiles $\sigma_1^{(t)}$ of the sender, so that whenever $s_\theta \succsim s_{\theta'}$ we have $\liminf_{t\to\infty} \sigma_1^{(t)}(s \mid \theta)/\sigma_1^{(t)}(s \mid \theta') \geq 1$. Write $q^{(t)}(\cdot \mid s)$ as the Bayesian posterior belief about the sender's type after signal

s under $\sigma_1^{(t)}$, which is well defined because each $\sigma_1^{(t)}$ is totally mixed. Whenever $s_\theta \succsim s_{\theta'}$, this sequence of posterior beliefs satisfies $\liminf_{t\to\infty} q^{(t)}(\theta \mid s)/q^{(t)}(\theta' \mid s) \geq \lambda(\theta)/\lambda(\theta')$, so if the receiver's plan best responds to every element in the sequence, it also best responds to an accumulation point $(q^\infty(\cdot \mid s))_{s\in S}$ with $q^\infty(\theta \mid s)/q^\infty(\theta' \mid s) \geq \lambda(\theta)/\lambda(\theta')$ whenever $s_\theta \succsim s_{\theta'}$. Since the player compatibility definition used in this paper is slightly easier to satisfy than the type compatibility definition that the set $P(s', \sigma^*)$ is based on, the plan best responds to $P(s', \sigma^*)$ after every signal s'. □

8.4 Proof of Proposition 5

Let $N = \max_i |\mathbb{S}_i|$. We first show that i's discounted lifetime play is the same whether i plays against strategy profiles drawn i.i.d. in different periods from the social distribution σ_{-i}, or against a response path drawn from a certain distribution η at the start of i's life. The next lemma constructs this η from σ, which is the same for all agents, and does not depend on their (possibly stochastic) learning policies.

Lemma 5. *In a factorable game, for each $\sigma \in \times_k \Delta(\mathbb{S}_k)$, there is a distribution η over response paths, so that for any player i, any possibly random policy $r_i : Y_i \to \Delta(\mathbb{S}_i)$, and any strategy $s_i \in \mathbb{S}_i$, we have*

$$\phi_i(s_i; r_i, \sigma) = (1-\gamma)\mathbb{E}_{\mathfrak{S}\sim\eta}\left[\sum_{t=1}^{\infty} \gamma^{t-1} \cdot (y_i^t(\mathfrak{S}, r_i) = s_i)\right],$$

where $y_i^t(\mathfrak{S}, r_i)$ refers to the t-th period history in $y_i(\mathfrak{S}, r_i)$.

Proof. In fact, we will prove a stronger statement: we will show there is such a distribution that induces the same distribution over period-t histories for every i, every learning policy r_i, and every t.

Think of each response path $\mathfrak{S}$ as a two-dimensional array, $\mathfrak{S} = (\mathfrak{S}_{t,n})_{t\in\mathbb{N}, 1\leq n\leq N}$. For non-negative integers $(m_n)_{n=1}^N$, each finite two-dimensional array of strategy profiles $((s_{t,n})_{t=1}^{m_n})_{n=1}^N$ with each $s_{t,n} \in \mathbb{S}$ defines a "cylinder set" of response paths with the form:

$$\{\mathfrak{S} : \mathfrak{S}_{t,n} = s_{t,n} \text{ for each } 1 \leq n \leq N, 1 \leq t \leq m_n\}.$$

That is, the cylinder set consists of those response paths whose first m_n elements for the n-th strategy match a given sequence of strategy profiles, $(s_{t,n})_{t=1}^{m_n}$. (If $m_n = 0$, then there is no restriction on $\mathfrak{S}_{t,n}$ for any t.) We specify the distribution η by specifying the probability it assigns to these cylinder sets:

$$\eta\left\{((s_{t,n})_{t=1}^{m_n})_{n=1}^{N}\right\} = \prod_{n=1}^{N}\prod_{t=1}^{m_n}\sigma(s_{t,n}),$$

where we have abused notation to write $((s_{t,n})_{t=1}^{m_n})_{n=1}^{N}$ for the cylinder set satisfying this profile of sequences, and we have used the convention that the empty product is defined to be 1.

We establish the claim by induction on t for period-t histories. For $t \geq 0$, let $Y_i[t] \subseteq Y_i$ be the set of possible period-t histories of i, that is $Y_i[t] := (\mathbb{S}_i \times \mathbb{O}_i)^t$. In the base case of $t = 1$, we show playing against a response path drawn according to η and playing against a pure strategy[21] drawn from $\sigma_{-i} \in \times_{k\neq i}\Delta(\mathbb{S}_k)$ generate the same period-1 history. Fixing a learning policy $r_i : Y_i \to \mathbb{S}_i$ of i, the probability of i having the period-1 history $(s_i^{(1)}, o^{(1)}) \in Y_i[1]$ in the random-matching model is $\mathbf{1}(r_i(\emptyset) = s_i^{(1)})\cdot\sigma(s : \mathfrak{o}_i(Z(s_i^{(1)}, s_{-i})) = o^{(1)})$. That is, i's policy must play $s_i^{(1)}$ in the first period of i's life. Then, i must encounter such a pure strategy that generates the required observation $o^{(1)}$, and this has probability $\sigma(s : \mathfrak{o}_i(Z(s_i^{(1)}, s_{-i})) = o^{(1)})$. The probability of this happening against a response path drawn from η is

$$\begin{aligned}
&\mathbf{1}(r_i(\emptyset) = s_i^{(1)}) \cdot \eta(\mathfrak{S}{:}\mathfrak{o}_i(Z(s_i^{(1)}, s_{1,s_i^{(1)},-i})) = o^{(1)}) \\
=&\mathbf{1}(r_i(\emptyset) = s_i^{(1)}) \cdot \sigma(s : \mathfrak{o}_i(Z(s_i^{(1)}, s_{-i})) = o^{(1)}),
\end{aligned}$$

where the second line comes from the probability η assigns to cylinder sets.

We now proceed with the inductive step. By induction, suppose random matching and the η-distributed response path induce the same distribution over the set of period-T histories, $Y_i[T]$, where $T \geq 1$. Write this common distribution as $\phi_{i,T}^{RM} = \phi_{i,T}^{\eta} = \phi_{i,T} \in \Delta(Y_i[T])$. We prove that they also generate the same distribution over length $T+1$ histories.

Suppose random matching generates distribution $\phi_{i,T+1}^{RM} \in \Delta(Y_i[T+1])$ and the η-distributed response path generates distribution $\phi_{i,T+1}^{\eta} \in \Delta(Y_i[T+1])$. Each length $T+1$ history $y_i[T+1] \in Y_i[T+1]$ may be written as $(y_i[T], (s_i^{(T+1)}, o^{(T+1)}))$, where $y_i[T]$ is a length-T history and $(s_i^{(T+1)}, o^{(T+1)})$ is a one-period history corresponding to what happens in period $T+1$. Therefore, we may write for each $y_i[T+1]$,

$$\phi_{i,T+1}^{RM}(y_i[T+1]) = \phi_{i,T}^{RM}(y_i[T]) \cdot \phi_{i,T+1|T}^{RM}((s_i^{(T+1)}, o^{(T+1)})|y_i[T]),$$

[21] In the random matching model agents are facing a randomly drawn pure strategy profile each period (and not a fixed behavior strategy): they are matched with random opponents, who each play a pure strategy in the game as a function of their personal history. From Kuhn's theorem, this is equivalent to facing a fixed profile of behavior strategies.

and

$$\phi^{\eta}_{i,T+1}(y_i[T+1]) = \phi^{\eta}_{i,T}(y_i[T]) \cdot \phi^{\eta}_{i,T+1|T}(((s_i^{(T+1)}, o^{(T+1)})|y_i[T]),$$

where $\phi^{RM}_{i,T+1|T}$ and $\phi^{\eta}_{i,T+1|T}$ are the conditional probabilities of the form "having history $(s_i^{(T+1)}, o^{(T+1)})$ in period $T+1$, conditional on having history $y_i[T] \in Y_i[T]$ in the first T periods." If such conditional probabilities are always the same for the random-matching model and the η-distributed response path model, then from the hypothesis $\phi^{RM}_{i,T} = \phi^{\eta}_{i,T}$, we can conclude $\phi^{RM}_{i,T+1} = \phi^{\eta}_{i,T+1}$.

By argument exactly analogous to the base case, we have for the random-matching model

$$\phi^{RM}_{i,T+1|T}((s_i^{(T+1)}, o^{(T+1)})|y_i[T]) = \mathbf{1}(r_i(y_i(T)) = s_i^{(T+1)}) \cdot \sigma(s : \mathfrak{o}_i(Z(s_i^{(T+1)}, s_{-i})) = o^{(T+1)}),$$

since the matching is independent across periods. In the η-distributed response path model, since a single response path is drawn once and fixed, one must compute the conditional probability that the drawn $\mathfrak{S}$ is such that the observation $o^{(T+1)}$will be seen in period $T+1$, given the history $y_i[T]$ (which is informative about which response path i is facing).

For each $1 \le n \le N$, let the non-negative integer m_n represent the number of times i has used the n-th strategy in $\mathbb{S}_i$ in the history $y_i[T]$. Let $(o_{t,n})_{1\le t \le m_n}$ represent the sequence of observations seen after using the n-th strategy, in chronological order. Consider the following finite union of cylinder sets, $(s_{t,n} : \mathfrak{o}_i(Z(n, s_{t,n,-i})) = o_{t,n})_{1\le t\le m_n, 1\le n\le N}$. This is the set of response sequences consistent with the observations so far.

If $\mathfrak{S}$ is to produce the observation $o^{(T+1)}$ from i's next play of $s_i^{(T+1)}$, then $\mathfrak{S}$ must belong to a more restrictive cylinder set that satisfies the additional restriction $(s_{m_{s_i^{(T+1)}}+1, s_i^{(T+1)}} : \mathfrak{o}_i(Z(s_i^{(T+1)}, s_{-i})) = o_{m_{s_i^{(T+1)}}+1, s_i^{(T+1)}})$. The conditional probability of $\mathfrak{S}$ belonging to this more restrictive cylinder set, given that it falls in $(s_{t,n} : \mathfrak{o}_i(Z(n, s_{t,n,-i})) = o_{t,n})_{1\le t\le m_n, 1\le n\le N}$, is given by the ratio of η-probabilities of these unions of cylinder sets, which from the product structure of η on cylinder sets, must be $\sigma(s : \mathfrak{o}_i(Z(s_i^{(T+1)}, s_{-i})) = o^{(T+1)})$. □

Thus, to prove that $\phi_i(s_i^*; r_i, \sigma_{-i}) \ge \phi_j(s_j^*; r_j, \sigma_{-j})$, it suffices to show that for every $\mathfrak{S}$, the period where s_i^* is played for the k-th time in induced history $y_i(\mathfrak{S}, r_i)$ happens earlier than the period where s_j^* is played for the k-th time in history $y_j(\mathfrak{S}, r_j)$.

Now we turn to the proof of Proposition 5.

Proof. Let $0 \le \gamma < 1$ and the social distribution σ be fixed. Enumerate the strategy sets of i and j so that s_i and $\varphi(s_i)$ are assigned the same number for every $s_i \in \mathbb{S}_i$. Consider the product distribution η on the space of response paths, $((\mathbb{S})^N)^\infty$, as in the proof of Lemma 5.

By Lemma 5, denote the period where s_i^* appears in $y_i(\mathfrak{S}, r_i)$ for the k-th time as $T_i^{(k)}$, the period where s_j^* appears in $y_j(\mathfrak{S}, r_j)$ for the k-th time as $T_j^{(k)}$. The quantities $T_i^{(k)}, T_j^{(k)}$ are defined to be ∞ if the corresponding strategies do not appear at least k times in the infinite histories. Write $\#(s_i^{'}; k) \in \mathbb{N} \cup \{\infty\}$ be the number of times $s_i^{'} \in \mathbb{S}_i$ is played in the history $y_i(\mathfrak{S}, r_i)$ before $T_i^{(k)}$. Similarly, $\#(s_j^{'}; k) \in \mathbb{N} \cup \{\infty\}$ denotes the number of times $s_j^{'} \in \mathbb{S}_j$ is played in the history $y_j(\mathfrak{S}, r_j)$ before $T_j^{(k)}$. Since φ establishes a bijection between $\mathbb{S}_i$ and $\mathbb{S}_j$, it suffices to show that for every $k = 1, 2, 3, ...$ either $T_j^{(k)} = \infty$ or for all $s_i^{'} \neq s_i^*$, $\#(s_i^{'}; k) \leq \#(s_j^{'}; k)$ where $s_j^{'} = \varphi(s_i^{'})$.

We show this by induction on k. First we establish the base case of $k = 1$.

Suppose $T_j^{(1)} \neq \infty$, and, by way of contradiction, suppose there is some $s_i^{'} \neq s_i^*$ such that $\#(s_i^{'}; 1) > \#(\varphi(s_i^{'}); 1)$. Find the subhistory y_i of $y_i(\mathfrak{S}, r_i)$. that leads to $s_i^{'}$ being played for the $(\#(\varphi(s_i^{'}); 1) + 1)$-th time, and find the subhistory y_j of $y_j(\mathfrak{S}, r_j)$ that leads to j playing s_j^* for the first time (y_j is well-defined because $T_j^{(1)} \neq \infty$). Note that $y_{i,s_i^*} \sim y_{j,s_j^*}$ vacuously, since i has never played s_i^* in y_i and j has never played s_j^* in y_j.

Also, $y_{i,s_i^{'}} \sim y_{j,s_j^{'}}$. To see this, note that i has played $s_i^{'}$ for $\#(\varphi(s_i^{'}); 1)$ times and j has played $s_j^{'}$ for the same number of times. The definition of response paths implies they faced the same sequence of opponent strategy profiles, and the definition of isomorphic learning problems implies they have gotten equivalent observations in all these periods.

Since $r_j(y_j) = s_j^*$ and r_j is an index policy, s_j^* must have weakly the highest index at y_j. Since r_i is more compatible with s_i^* than r_j is with s_j^*, $s_i^{'}$ must not have the weakly highest index at y_i. And yet $r_i(y_i) = s_i'$ contradiction.

Now suppose this statement holds for all $k \leq K$ for some $K \geq 1$. We show it also holds for $k = K+1$. If $T_j^{(K+1)} = \infty$ or $T_j^{(K)} = \infty$, we are done. Otherwise, by way of contradiction, suppose there is some $s_i^{'} \neq s_i^*$ so that $\#(s_i'; K+1) > \#(\varphi(s_i^{'}); K+1)$. Find the subhistory y_i of $y_i(\mathfrak{S}, r_i)$. that leads to $s_i^{'}$ being played for the $(\#(\varphi(s_i^{'}); K+1) + 1)$-th time. Since $T_j^{(K)} \neq \infty$, from the inductive hypothesis $T_i^{(K)} \neq \infty$ and $\#(s_i'; K) \leq \#(\varphi(s_i^{'}); K)$. That is, i must have played $s_i^{'}$ no more than $\#(\varphi(s_i^{'}); K)$ times before playing s_i^* for the K-th time. Since $\#(\varphi(s_i^{'}); K+1) + 1 > \#(\varphi(s_i^{'}); K)$, the subhistory y_i must extend beyond period $T_i^{(K)}$, so it contains K instances of i playing s_i^*.

Next, find the subhistory y_j of $y_j(\mathfrak{S}, r_j)$ that leads to j playing s_j^* for the $(K+1)$-th time. (This is well-defined because $T_j^{(K+1)} \neq \infty$.) Note that $y_{i,s_i^*} \sim y_{j,s_j^*}$, since i and j have played s_i^*, s_j^* for K times each, and they were facing the same response paths. Also, $y_{i,s_i^{'}} \sim y_{j,s_j^{'}}$ since i has played $s_i^{'}$ for $\#(\varphi(s_i^{'}); K+1)$ times and j has played $s_j^{'}$ for the same number of times. Since $r_j(y_j) = s_j^*$ and r_j is an index policy, s_j^* must have weakly the highest index at y_j. Since r_i is more compatible with s_i^* than r_j is with s_j^*, $s_i^{'}$ must not have the weakly highest index at y_i. And yet $r_i(y_i) = s_i'$ contradiction. $\square$

8.5 Proof of Lemma 1

Proof. By way of contradiction, suppose there is some profile of moves by $-i$, $(a_h)_{h\in\mathcal{H}_{-i}}$, so that h^* is off the path of play in $(s_i,(a_h)_{h\in\mathcal{H}_{-i}}) = (s_i, a_{h^*}, (a_h)_{h\in\mathcal{H}_{-i}\backslash h^*})$. Find a different action of j on h^*, $a^{'}_{h^*} \neq a_{h^*}$. Since h^* is off the path of play, both $(s_i, a_{h^*}, (a_h)_{h\in\mathcal{H}_{-i}\backslash h^*})$ and $(s_i, a^{'}_{h^*}, (a_h)_{h\in\mathcal{H}_{-i}\backslash h^*})$ lead to the same payoff for i. But by Condition (1) in the definition of factorability and the fact that $h^* \in F_i[s_i]$, we will have found two $-i$ action profiles $s_{-i}, s^{'}_{-i}$ in two different blocks of $\Pi_i[s_i]$ with $u_i(s_i, s_{-i}) = u_i(s_i, s^{'}_{-i})$. This contradicts $\Pi_i[s_i]$ being the coarsest partition of $\mathbb{S}_{-i}$ that makes $u_i(s_i,\cdot)$ measurable. □

8.6 Proof of Lemma 2

Proof. Since i's payoff is not independent of h^*, there exist actions $a_{h^*} \neq a^{'}_{h^*}$ on h^* and a profile a_{-h^*} of actions elsewhere in the game tree, so that $u_i(a_{h^*}, a_{-h^*}) \neq u_i(a^{'}_{h^*}, a_{-h^*})$. Consider the strategy s_i for i that matches a_{-h^*} in terms of play on i's information sets, so we may equivalently write

$$u_i(s_i, a_{h^*}, (a_h)_{h\in\mathcal{H}_{-i}\backslash h^*}) \neq u_i(s_i, a^{'}_{h^*}, (a_h)_{h\in\mathcal{H}_{-i}\backslash h^*}),$$

where $(a_h)_{h\in\mathcal{H}_{-i}\backslash h^*}$ are the components of a_{-h^*} corresponding to information sets of $-i$. If $h^* \notin F_i[s_i]$, then by Condition (1) of factorability, $(a_{h^*}, (a_h)_{h\in\mathcal{H}_{-i}\backslash h^*})$ and $(a^{'}_{h^*}, (a_h)_{h\in\mathcal{H}_{-i}\backslash h^*})$ belong to the same block in $\Pi_i[s_i]$. Yet, they give different payoffs to i, which contradicts that i's payoff after s_i must be measurable with respect to $\Pi_i[s_i]$. □

8.7 Proof of Proposition 6

Proof. Combining Lemmas 1 and 2 implies there is an extensive-form strategy $s_i \in \mathbb{S}_i$ such that h^* is on the path of play whenever i chooses s_i. Consider some strategy profile $(s_i^\circ, s_{-i}^\circ)$ where h^* is off the path. Then i can unilaterally deviate to s_i, and h^* is on the path of (s_i, s_{-i}°). Furthermore, i's play differs on the new path relative to the old path on exactly one information set, since i plays at most once on any path. So instead of deviating to s_i, i can deviate to s_i' that matches s_i in terms of this information set where i's play is modified, but otherwise is the same as s_i°. So h^* is also on the path of play for $(s_i' s_{-i}^\circ)$, where $s^{'}_i$ differs from s_i° only on one information set. □

9 Index Compatibility of OPT and WFP when $s_i^* \succsim s_j^*$

In this section, we show that OPT and WFP are index compatible under the conditions of Theorem 2. This conclusion, when combined with Proposition 5, implies Theorem 2.

Let the equivalence classes $\mathbb{E}$ be such that $(s_i, u_i(s_i, \tilde{s}_{-i})) \sim (\varphi(s_i), u_j(\varphi(s_i), \hat{s}_{-j})$ if and only if $\tilde{s}_{-i}|_{F_i[s_i] \cap \mathcal{H}_{-ij}} = \hat{s}_{-j}|_{F_j[\varphi(s_i)] \cap \mathcal{H}_{-ij}}$. Conditions on factorability and isomorphic factoring ensure that i, j face $(\varphi, \mathbb{E})$-isomorphic learning problems. Indeed, if i and j faced the same pure profile $\tilde{s}$, then $\tilde{s}_{-i}|_{F_i[s_i] \cap \mathcal{H}_{-ij}} = \tilde{s}_{-j}|_{F_j[\varphi(s_i)] \cap \mathcal{H}_{-ij}}$ since $F_i[s_i] \cap \mathcal{H}_{-ij} = F_j[\varphi(s_i)] \cap \mathcal{H}_{-ij}$ by isomorphic factoring.

9.1 Weighted Fictitious Play

To see that WFP satisfies index compatibility for s_i^* and s_j^* under the conditions of Theorem 2, let histories y_i, y_j and strategy $s_i^{'} \neq s_i^*$ be given with $y_{i,s_i^*} \sim y_{j,s_j^*}$, $y_{i,s_i^{'}} \sim y_{j,\varphi(s_i^{'})}$, and s_j^* having weakly the highest index for j. Construct two totally mixed, independent behavior strategy profile, $\beta, \tilde{\beta}$ as follows. For each $s_j \in \mathbb{S}_j$, $\beta(h) := \alpha_h(\cdot; y_j)$ for all $h \in F_j[s_j]$. (This is well-defined by Condition (2) of factorability, as $F_j[s_j] \cap F_j[s_j'] = \varnothing$ if $s_j \neq s_j^{'}$.) For those $h \in \mathcal{H} \backslash \cup_{s_j \in \mathbb{S}_j} F_j[s_j]$, arbitrarily specify a strictly mixed action $\alpha_h \in \Delta(A_h)$ for $\beta(h)$. Having constructed β we turn to $\tilde{\beta}$. For each $s_i \in \{s_i^*, s_i^{'}\}$, $\tilde{\beta}(h) := \alpha_h(\cdot; y_i)$ for all $h \in F_i[s_i]$. For all other $h \in \mathcal{H}$, let $\tilde{\beta}(h) := \beta(h)$.

From the definition of $y_{i,s_i^*} \sim y_{j,s_j^*}$, $\tilde{\beta}(h) = \beta(h)$ for all $h \in F_i[s_i] \cap \mathcal{H}_{-ij}$. From the definition of $y_{i,s_i^{'}} \sim y_{j,\varphi(s_i^{'})}$, $\tilde{\beta}(h) = \beta(h)$ for all $h \in F_i[s_i'] \cap \mathcal{H}_{-ij}$. Also, $\tilde{\beta}(h) = \beta(h)$ for all other $h \in \mathcal{H}_{-ij}$ by construction. So, $\tilde{\beta}$ and β are totally mixed behavior strategy profiles that match on the $-ij$ marginal, and they can be represented by $\tilde{\sigma}, \sigma$ totally mixed strategy profiles (over $\mathbb{S}$) that match on the $-ij$ marginal.

Since j's payoff from each s_j only depends on $-j$'s play on $F_j[s_j]$ by Condition (1) of factorability, $\sum_{s \in \mathbb{S}} u_j(s_j, s_{-j}) \cdot \sigma(s)$ equals to the index that the weighted fictitious play agent assigns to s_j after history y_j. Since s_j^* has the weakly highest index, $\sum_{s \in \mathbb{S}} u_j(s_j^*, s_{-j}) \cdot \sigma(s) = \max_{s_j^{'} \in \mathbb{S}_j} \sum_{s \in \mathbb{S}} u_j(s_j^{'}, s_{-j}) \cdot \sigma(s)$. From the definition of player compatibility, s_i^* is strictly optimal against $\tilde{\sigma}$, which in particular means $\sum_{s \in \mathbb{S}} u_i(s_i^*, s_{-i}) \cdot \tilde{\sigma}(s) > \sum_{s \in \mathbb{S}} u_i(s_i^{'}, s_{-i}) \cdot \tilde{\sigma}(s)$. The RHS is i's index for s_i' after y_i, since $\tilde{\sigma}$ marginalized to every $h \in F_i[s_i']$ is $\alpha_h(\cdot; y_i)$ by construction. This says $s_i^{'}$ does not have the weakly highest index for i after y_i.

Thus, WFP satisfies index compatibility for s_i^* and s_j^*.

9.2 The Gittins Index

Write $V(\tau; s_i, \nu_{s_i})$ for the value of the above auxiliary problem under the (not necessarily optimal) stopping time τ in the definition of the Gittins index. The Gittins index of s_i is $\sup_{\tau>0} V(\tau; s_i, \nu_{s_i})$. We begin by linking $V(\tau; s_i, \tau_{s_i})$ to i's stage-game payoff from playing s_i. From belief ν_{s_i} and stopping time τ, we will construct the correlated profile $\alpha(\nu_{s_i}, \tau) \in \Delta^\circ(\times_{h\in F_i[s_i]} A_h)$, so that $V(\tau; s_i, \nu_{s_i})$ is equal to i's expected payoff when playing s_i while opponents play according to this correlated profile on the s_i-relevant information sets.

Definition 19. A full-support belief $\nu_{s_i} \in \times_{h\in F_i[s_i]}\Delta(\Delta(A_h))$ for player i together with a (possibly random) stopping rule $\tau > 0$ together induce a stochastic process $(\tilde{\boldsymbol{a}}_{(-i),t})_{t\geq 1}$ over the space $\times_{h\in F_i[s_i]} A_h \cup \{\varnothing\}$, where $\tilde{\boldsymbol{a}}_{(-i),t} \in \times_{h\in F_i[s_i]} A_h$ represents the opponents' actions observed in period t if $\tau \geq t$, and $\tilde{\boldsymbol{a}}_{(-i),t} = \varnothing$ if $\tau < t$. We call $\tilde{\boldsymbol{a}}_{(-i),t}$ player i's *internal history* at period t and write $\mathbb{P}_{(-i)}$ for the distribution over internal histories that the stochastic process induces.

Internal histories live in the same space as player i's actual experience in the learning problem, represented as a history in $\mathbb{O}_i$. The process over internal histories is i's prediction about what would happen in the auxiliary problem if they were to use τ.

Enumerate all possible profiles of moves at information sets $F_i[s_i]$ as $\times_{h\in F_i[s_i]} A_h = \{\boldsymbol{a}^{(1)}_{(-i)}, ..., \boldsymbol{a}^{(K)}_{(-i)}\}$, let $p_{t,k} := \mathbb{P}_{(-i)}[\tilde{\boldsymbol{a}}_{(-i),t} = \boldsymbol{a}^{(k)}_{(-i)}]$ for $1 \leq k \leq K$ be the probability under $\nu_{s_i}$of seeing the profile of actions $\boldsymbol{a}^{(k)}_{(-i)}$ in period t of the stochastic process over internal histories, $(\tilde{\boldsymbol{a}}_{(-i),t})_{t\geq 0}$, and let $p_{t,0} := \mathbb{P}_{(-i)}[\tilde{\boldsymbol{a}}_{(-i),t} = \varnothing]$ be the probability of having stopped before period t.

Definition 20. The *synthetic correlated profile* at information sets in $F_i[s_i]$ is the element of $\Delta^\circ(\times_{h\in F_i[s_i]} A_h)$ (i.e. a correlated random action) that assigns probability $\frac{\sum_{t=1}^{\infty} \beta^{t-1} p_{t,k}}{\sum_{t=1}^{\infty} \beta^{t-1}(1-p_{t,0})}$ to the profile of actions $\boldsymbol{a}^{(k)}_{(-i)}$. Denote this profile by $\alpha(\nu_{s_i}, \tau)$.

Note that the synthetic correlated profile depends on the belief ν_{s_i} stopping rule τ, and effective discount factor β. Since the belief ν_{s_i} has full support, there is always a positive probability assigned to observing every possible profile of actions on $F_i[s_i]$ in the first period, so the synthetic correlated profile is totally mixed. The significance of the synthetic correlated profile is that it gives an alternative expression for the value of the auxiliary problem under stopping rule τ.

Lemma 6.

$$V(\tau; s_i, \nu_{s_i}) = u_i(s_i, \alpha(\nu_{s_i}, \tau))$$

The proof is the same as in Fudenberg and He (2018) and is omitted.[22]

Consider now the situation where i and j share the same beliefs about play of $-ij$ on the common information sets $F_i[s_i] \cap F_j[s_j] \subseteq \mathcal{H}_{-ij}$. For any pure-strategy stopping time τ_j of j, we define a random stopping rule of i, the *mimicking stopping time* for τ_j. Lemma 7 will establish that the mimicking stopping time generates a synthetic correlated profile that matches the corresponding profile of τ_j on $F_i[s_i] \cap F_j[s_j]$.

Note that τ_j maps j's internal histories to stopping decisions, which do not live in the same space as i's internal histories. In particular, τ_j could make use of i's play to decide whether to stop. To mimic such a rule, i makes use of external histories, which include both the common component of i's internal history on $F_i[s_i] \cap F_j[s_j]$, as well as simulated histories on $F_j[s_j] \backslash (F_i[s_i] \cap F_j[s_j])$.

For a given isomorphism φ between i and j with $\varphi(s_i) = s_j$ and F_i, F_j, we may write $F_i[s_i] = F^C \cup \bar{F}^{-i}$ with $F^C \subseteq \mathcal{H}_{-ij}$ and $\bar{F}^{-i} \subseteq \mathcal{H}_{-i}$. Similarly, we may write $F_j[s_j] = F^C \cup \bar{F}^{-j}$ with $\bar{F}^{-j} \subseteq \mathcal{H}_{-j}$. (So, F^C is the **c**ommon information sets that are observed after both s_i and s_j.) Whenever j plays s_j, they observes some $(\boldsymbol{a}_{(C)}, \boldsymbol{a}_{(-j)}) \in (\times_{h \in F^C} A_h) \times (\times_{h \in \bar{F}^{-j}} A_h)$, where $\boldsymbol{a}_{(C)}$ is a profile of actions at information sets in F^C and $\boldsymbol{a}_{(-j)}$ is a profile of actions at information sets in $\bar{F}^{-j}$. So a pure-strategy stopping rule in the auxiliary problem defining j's Gittins index for s_j is a function $\tau_j : \cup_{t \geq 1}[(\times_{h \in F^C} A_h) \times (\times_{h \in \bar{F}^{-j}} A_h)]^t \to \{0, 1\}$ that maps finite histories in $\mathbb{O}_j$to stopping decisions, where "0" means continue and "1" means stop.

Definition 21. Player i's *mimicking stopping rule for* τ_j draws $\alpha^{-j} \in \times_{h \in \bar{F}^{-j}} \Delta(A_h)$ from j's belief ν_{s_j} on $\bar{F}^{-j}$, and then draws $(\boldsymbol{a}_{(-j),\ell})_{\ell \geq 1}$ by independently generating $\boldsymbol{a}_{(-j),\ell}$ from α^{-j} each period. Conditional on $(\boldsymbol{a}_{(-j),\ell})$, i stops according to the rule $(\tau_i | (\boldsymbol{a}_{(-j),\ell}))((\boldsymbol{a}_{(C),\ell}, \boldsymbol{a}_{(-i),\ell})_{\ell=1}^t) := \tau_j((\boldsymbol{a}_{(C),\ell}, \boldsymbol{a}_{(-j),\ell})_{\ell=1}^t)$.[23]

That is, the mimicking stopping rule involves ex-ante randomization across a family of pure-strategy stopping rules $\tau_i | (\boldsymbol{a}_{(-j),\ell})_{\ell=1}^\infty$, indexed by $(\boldsymbol{a}_{(-j),\ell})_{\ell=1}^\infty$. First, i draws a behavior strategy on the information sets $\bar{F}^{-j}$ according to j's belief about $-j$'s play there. Then, i simulates an infinite sequence $(\boldsymbol{a}_{(-j),\ell})_{\ell=1}^\infty$ of i's play using this drawn behavior strategy and follows the pure-strategy stopping rule $\tau_i | (\boldsymbol{a}_{(-j),\ell})_{\ell=1}^\infty$.

As in the definition of internal histories, the mimicking strategy and i's belief ν_{s_i} generates a stochastic process $(\tilde{\boldsymbol{a}}_{(-i),t}, \tilde{\boldsymbol{a}}_{(C),t})_{t \geq 1}$ of internal histories for i (representing actions

[22]Notice that even though i starts with the belief that opponents randomize independently at different information sets, and also holds an independent prior belief, $V(\tau; s_i, \nu_{s_i})$ may not be the payoff of playing s_i against a independent randomizations by the opponent because of the endogenous correlation that we discussed in the text.

[23]Note this is a valid (stochastic) stopping time, as the event $\{\tau_i \leq T\}$ only depends on i's observations in $\mathbb{O}_i$ in the first T periods, plus some private randomizations of i.

on $F_i[s_i]$ that i anticipates seeing when they plays s_i). It also induces a stochastic process $(\tilde{\boldsymbol{e}}_{(-j),t}, \tilde{\boldsymbol{e}}_{(C),t})_{t\geq 1}$ of "external histories" defined in the following way:

Definition 22. The stochastic process of *external histories* $(\tilde{\boldsymbol{e}}_{(-j),t}, \tilde{\boldsymbol{e}}_{(C),t})_{t\geq 1}$ is defined from the process of internal histories $(\tilde{\boldsymbol{a}}_{(-i),t}, \tilde{\boldsymbol{a}}_{(C),t})_{t\geq 1}$ that τ_i generates and given by: (i) if $\tau_i < t$, then $(\tilde{\boldsymbol{e}}_{(-j),t}, \tilde{\boldsymbol{e}}_{(C),t}) = \varnothing$; (ii) otherwise, $\tilde{\boldsymbol{e}}_{(C),t} = \tilde{\boldsymbol{a}}_{(C),t}$, and $\tilde{\boldsymbol{e}}_{(-j),t}$ is the t-th element of the infinite sequence $(\boldsymbol{a}_{(-j),\ell})_{\ell=1}^{\infty}$ that i simulated before the first period of the auxiliary problem.

Write $\mathbb{P}_e$ for the distribution over the sequence of of external histories generated by i's mimicking stopping time for τ_j, which is a function of τ_j, ν_{s_j}, and ν_{s_i}.[24]

When using the mimicking stopping time for τ_j in the auxiliary problem, i expects to see the same distribution of $-ij$'s play before stopping as j does when using τ_j, on the information sets in $F_i[s_i] \cap F_j[s_j]$. This is formalized in the next lemma.

Lemma 7. *Suppose the game is isomorphcially factorable for i and j with $\varphi(s_i) = s_j$, and suppose i holds belief ν_{s_i} over play in $F_i[s_i]$ and j holds belief ν_{s_j} over play in $F_j[s_j]$, such that $\nu_{s_i}|_{F_i[s_i]\cap F_j[s_j]} = \nu_{s_j}|_{F_i[s_i]\cap F_j[s_j]}$, that is the two sets of beliefs match when marginalized to the common information sets in $\mathcal{H}_{-ij}$. Let τ_i be i's mimicking stopping time for τ_j. Then, the synthetic correlated profile $\alpha(\nu_{s_j}, \tau_j)$ marginalized to the information sets of $-ij$ is the same as $\alpha(\nu_{s_i}, \tau_i)$ marginalized to the same information sets.*

Proposition 9. *Suppose the game is isomorphcially factorable for i and j with $\varphi(s_i) = s_j$, $\varphi(s_i^{'}) = s_j^{'}$, where $s_i^* \neq s_i^{'}$. Suppose i is more player-compatible with s_i^* than j is with s_j^*. Suppose i holds belief $\nu_{s_i} \in \times_{h\in F_i[s_i]}\Delta(\Delta(A_h))$ about opponents' play after each s_i and j holds belief $\nu_{s_j} \in \times_{h\in F_j[s_j]}\Delta(\Delta(A_h))$ about opponents' play after each s_j, such that $\nu_{s_i^*}|_{F_i[s_i^*]\cap F_j[s_j^*]} = \nu_{s_j^*}|_{F_i[s_i^*]\cap F_j[s_j^*]}$ and $\nu_{s_i^{'}}|_{F_i[s_i^{'}]\cap F_j[s_j^{'}]} = \nu_{s_j^{'}}|_{F_i[s_i^{'}]\cap F_j[s_j^{'}]}$. If s_j^* has the weakly highest Gittins index for j under effective discount factor $0 \leq \delta\gamma < 1$, then $s_i^{'}$ does not have the weakly highest Gittins index for i under the same effective discount factor.*

Proof. We begin by defining a collection of totally mixed correlated profiles $(\alpha_{[s_j]})_{s_j\in\mathbb{S}_j}$ where $\alpha_{[s_j]} \in \Delta^{\circ}(\times_{h\in F_j[s_j]}A_h)$. For each $s_j \neq s_j^{'}$ the profile $\alpha_{[s_j]}$ is the synthetic correlated profile $\alpha(\nu_{s_j}, \tau_{s_j}^*)$, where $\tau_{s_j}^*$ is an optimal pure-strategy stopping time in j's auxiliary stopping problem involving s_j. For $s_j = s_j^{'}$, the correlated profile $\alpha_{[s_j^{'}]}$ is instead the synthetic correlated

[24]To understand the distinction between internal and external histories, note that the probability of i's first-period internal history satisfying $(\tilde{\boldsymbol{a}}_{(-i),1}, \tilde{\boldsymbol{a}}_{(C),1}) = (\bar{\boldsymbol{a}}_{(-i)}, \bar{\boldsymbol{a}}_{(C)})$ for some fixed values $(\bar{\boldsymbol{a}}_{(-i)}, \bar{\boldsymbol{a}}_{(C)}) \in \times_{h\in F_i[s_i]}A_h$ is given by the probability that a mixed play α_{-i} on $F_i[s_i]$, drawn according to i's belief ν_{s_i}, would generate the profile of actions $(\bar{\boldsymbol{a}}_{(-i)}, \bar{\boldsymbol{a}}_{(C)})$. On the other hand, the probability of i's first-period external history satisfying $(\tilde{\boldsymbol{e}}_{(-j),1}, \tilde{\boldsymbol{e}}_{(C),1}) = (\bar{\boldsymbol{a}}_{(-j)}, \bar{\boldsymbol{a}}_{(C)})$ for some fixed values $(\bar{\boldsymbol{a}}_{(-j)}, \bar{\boldsymbol{a}}_{(C)}) \in \times_{h\in F_j[s_j]}A_h$ also depends on j's belief ν_{s_j}, for this belief determines the distribution over $(\boldsymbol{a}_{(-j),\ell})_{\ell=1}^{\infty}$ drawn before the start of the auxiliary problem.

profile associated with the mimicking stopping rule for $\tau^*_{s_i'}$, i.e. mimicking agent i's pure-strategy optimal stopping time in i's auxiliary problem for s_i'.

Next, define a profile of totally mixed correlated actions $(\alpha_{[s_i]})_{s_i \in \mathbb{S}_i}$ for i's opponents on information sets $(F_i[s_i])_{s_i \in \mathbb{S}_i}$. For each $s_i \notin \{s_i^*, s_i'\}$, just use the marginal distribution of $\alpha_{[\varphi(s_i)]}$ constructed before on $F_i[s_i] \cap F_j[\varphi(s_i)]$, then arbitrarily specify play in $F_i[s_i] \backslash F_j[\varphi(s_i)]$, if any. For s_i' the correlated profile is $\alpha(\nu_{s_i'}, \tau^*_{s_i'})$, i.e. the synthetic move associated with i's optimal stopping rule for s_i'. Finally, for s_i^*, the correlated profile $\alpha_{[s_i^*]}$ is the synthetic correlated profile associated with the mimicking stopping rule for $\tau^*_{s_j^*}$.

From Lemma 7, for every s_i, the profiles of correlated actions $\alpha_{[s_i]}$ and $\alpha_{[\varphi(s_i)]}$ agree when marginalized to the information sets $F_i[s_i] \cap F_j[\varphi(s_i)]$. Therefore, $(\alpha_{[s_i]})_{s_i \in \mathbb{S}_i}$ and $(\alpha_{[s_j]})_{s_j \in \mathbb{S}_j}$ can be completed into two totally mixed correlated strategy profiles, $\tilde{\sigma}$ and σ (over $\mathbb{S}$), such that $\tilde{\sigma}|_{F_i[s_i] \cap F_j[\varphi(s_i)]} = \sigma|_{F_i[s_i] \cap F_j[\varphi(s_i)]}$ for every s_i. For each $s_j \neq s_j'$, the Gittins index of s_j for j is $u_j(s_j, \sigma_{s_j})$. Also, since $\alpha_{[s_j']}$ is the mixed profile associated with the suboptimal mimicking stopping time, $u_j(s_j', \sigma_{s_j'})$ is no larger than the Gittins index of s_j' for j. By the hypothesis that s_j^* has the weakly highest Gittins index for j,$u_j(s_j^*, \sigma_{s_j^*}) \geq \max_{s_j \neq s_j^*} u_j(s_j, \sigma_{s_j})$. By the definition of strong player compatibility, we must also have $u_i(s_i^*, \sigma_{s_i^*}) > \max_{s_i \neq s_i^*} u_i(s_i, \sigma_{s_i})$, so in particular $u_i(s_i^*, \sigma) > u_i(s_i' \sigma_{s_i'})$. But $u_i(s_i^*, \sigma_{s_i^*})$ is no larger than the Gittins index of s_i^*, for $\alpha_{[s_i^*]}$ is the synthetic strategy associated with a suboptimal mimicking stopping time. As $u_i(s_i' \sigma_{s_i'})$ is equal to the Gittins index of s_i' this shows s_i' cannot have even weakly the highest Gittins index at this belief,for s_i^* already has a strictly higher Gittins index than s_i' does. $\square$

To see that OPT is index compatible for s_i^*, s_j^* under the conditions of Theorem 2, let histories y_i, y_j and strategy $s_i' \neq s_i^*$ be given with $y_{i,s_i^*} \sim y_{j,s_j^*}$, $y_{i,s_i'} \sim y_{j,\varphi(s_i')}$. Since g_i, g_j are equivalent priors, i, j's posterior beliefs match on every $F \in F_i[s_i] \cap F_j[\varphi(s_i)]$, for $s_i \in \{s_i^*, s_i'\}$. After such histories, if s_j^* has weakly the highest Gittins index for j, we use the hypothesis of player compatibility and Proposition 9 to see that s_i' does not have the weakly highest Gittins index for i.

9.3 Proof of Lemma 7

Proof. Let $(\tilde{\boldsymbol{a}}_{(-i),t}, \tilde{\boldsymbol{a}}_{(C),t})_{t \geq 1}$ and $(\tilde{\boldsymbol{e}}_{(-j),t}, \tilde{\boldsymbol{e}}_{(C),t})_{t \geq 1}$ be the stochastic processes of internal and external histories for τ_i, with distributions $\mathbb{P}_{-i}$ and $\mathbb{P}_e$. Enumerate possible profiles of actions on F^C as $\times_{h \in F^C} A_h = \{\boldsymbol{a}_{(C)}^{(1)}, ..., \boldsymbol{a}_{(C)}^{(K_C)}\}$, possible profiles of actions on $\bar{F}^{-i}$ as $\times_{h \in \bar{F}^{-i}} A_h = \{\boldsymbol{a}_{(-i)}^{(1)}, ..., \boldsymbol{a}_{(-i)}^{(K_{-i})}\}$, and possible profiles of actions on $\bar{F}^{-j}$ as $\times_{h \in \bar{F}^{-j}} A_h = \{\boldsymbol{a}_{(-j)}^{(1)}, ..., \boldsymbol{a}_{(-j)}^{(K_{-j})}\}$.

Write $p_{t,(k_{-i},k_C)} := \mathbb{P}_{-i}[(\tilde{\boldsymbol{a}}_{(-i),t}, \tilde{\boldsymbol{a}}_{(C),t}) = (\boldsymbol{a}_{(-i)}^{(k_{-i})}, \boldsymbol{a}_{(C)}^{(k_C)})]$ for $k_{-i} \in \{1, ..., K_{-i}\}$ and $k_C \in \{1, ..., K_C\}$. Also write $q_{t,(k_{-j},k_C)} := \mathbb{P}_e[(\tilde{\boldsymbol{e}}_{(-j),t}, \tilde{\boldsymbol{e}}_{(C),t}) = (\boldsymbol{a}_{(-j)}^{(k_{-j})}, \boldsymbol{a}_{(C)}^{(k_C)})]$ for $k_{-j} \in \{1, ..., K_{-j}\}$

and $k_C \in \{1, ..., K_C\}$. Let $p_{t,(0,0)} = q_{t,(0,0)} := \mathbb{P}_{-i}[\tau_i < t] = \mathbb{P}_e[\tau_i < t]$ be the probability of having stopped before period t.

The distribution of external histories that i expects to observe before stopping under belief ν_{s_i} when using the mimicking stopping rule τ_i is the same as the distribution of internal histories that j expects to observe when using stopping rule τ_j under belief ν_{s_j}, because i simulates the data-generating process on $\bar{F}^{-j}$ by drawing a mixed action α^{-j} according to j's belief $\nu_{s_j}|_{\bar{F}^{-j}}$ and $\nu_{s_i}|_{F^C} = \nu_{s_j}|_{F^C}$. Thus for every $k_{-j} \in \{1, ..., K_{-j}\}$ and every $k_C \in \{1, ..., K_C\}$,

$$\frac{\sum_{t=1}^{\infty} (\delta\gamma)^{t-1} q_{t,(k_{-j},k_C)}}{\sum_{t=1}^{\infty} (\delta\gamma)^{t-1} (1 - q_{t,(0,0)})} = \alpha(\nu_{s_j}, \tau_j)(\boldsymbol{a}_{(-j)}^{(k_{-j})}, \boldsymbol{a}_{(C)}^{(k_C)}).$$

For a fixed $\bar{k}_C \in \{1, ..., K_C\}$, summing across k_{-j} gives

$$\frac{\sum_{t=1}^{\infty} (\delta\gamma)^{t-1} \sum_{k_{-j}=1}^{K_{-j}} q_{t,(k_{-j},\bar{k}_C)}}{\sum_{t=1}^{\infty} (\delta\gamma)^{t-1} (1 - q_{t,(0,0)})} = \alpha(\nu_{s_j}, \tau_j)(\boldsymbol{a}_{(C)}^{(\bar{k}_C)}).$$

By definition, the processes $(\tilde{\boldsymbol{a}}_{(-i),t}, \tilde{\boldsymbol{a}}_{(C),t})_{t \geq 0}$ and $(\tilde{\boldsymbol{e}}_{(-j),t}, \tilde{\boldsymbol{e}}_{(C),t})_{t \geq 0}$ have the same marginal distribution on the second dimension:

$$\sum_{k_{-j}=1}^{K_{-j}} q_{t,(k_{-j},\bar{k}_C)} = \mathbb{P}_{-i}[\tilde{\boldsymbol{a}}_{(C),t} = \boldsymbol{a}_{(C)}^{(\bar{k}_C)}] = \sum_{k_{-i}=1}^{K_{-i}} p_{t,(k_{-i},\bar{k}_C)}.$$

Making this substitution and using the fact that $p_{t,(0,0)} = q_{t,(0,0)}$,

$$\frac{\sum_{t=1}^{\infty} (\delta\gamma)^{t-1} \sum_{k_{-i}=1}^{K_{-i}} p_{t,(k_{-i},\bar{k}_C)}}{\sum_{t=1}^{\infty} (\delta\gamma)^{t-1} (1 - p_{t,(0,0)})} = \alpha(\nu_{s_j}, \tau_j)(\boldsymbol{a}_{(C)}^{(\bar{k}_C)}).$$

But by the definition of synthetic correlated profile, the LHS is $\sum_{k_{-i}=1}^{K_{-i}} \alpha(\nu_{s_i}, \tau_i)(\boldsymbol{a}_{(-i)}^{(k_{-i})}, \boldsymbol{a}_{(C)}^{(\bar{k}_C)}) = \alpha(\nu_{s_i}, \tau_i)(\boldsymbol{a}_{(C)}^{(\bar{k}_C)})$.

Since the choice of $\boldsymbol{a}_{(C)}^{(\bar{k}_C)} \in \times_{h \in F^C} A_h$ was arbitrary, we have shown that the synthetic profile $\alpha(\nu_{s_j}, \tau_j)$ of the original stopping rule τ_j and the one associated with the mimicking strategy of i, $\alpha(\nu_{s_i}, \tau_i)$, coincide on F^C. $\square$

Online Appendix

10 Proofs Omitted from the Appendix

10.1 Proof of Proposition 1

Proof. Suppose s_k^* is weakly optimal for k against some totally mixed correlated profile $\sigma^{(k)}$. We show that s_i^* is strictly optimal for i against any totally mixed and correlated $\sigma^{(i)}$ with the property that $\text{marg}_{-ik}(\sigma^{(k)}) = \text{marg}_{-ik}(\sigma^{(i)})$.

To do this, we first modify $\sigma^{(i)}$ into a new totally profile by copying how the action of i correlates with the actions of $-(ik)$ in $\sigma^{(k)}$. For each $s_{-ik} \in \mathbb{S}_{-ik}$ and $s_i \in \mathbb{S}_i$, $\sigma^{(k)}(s_i, s_{-ik}) > 0$ since $\text{marg}_{-k}(\sigma^{(k)}) \in \Delta^\circ(\mathbb{S}_{-k})$. So write $p(s_i \mid s_{-ik}) := \frac{\sigma^{(k)}(s_i, s_{-ik})}{\sum_{s_i' \in \mathbb{S}_i} \sigma^{(k)}(s_i' s_{-ik})} > 0$ as the conditional probability that i plays s_i given $-ik$ play s_{-ik}, in the profile $\sigma^{(k)}$. Now construct the profile $\hat{\hat{\sigma}} \in \Delta^\circ(\mathbb{S})$, where

$$\hat{\hat{\sigma}}(s_i, s_{-ik}, s_k) := p(s_i \mid s_{-ik}) \cdot \sigma^{(i)}(s_{-ik}, s_k).$$

Profile $\hat{\hat{\sigma}}$ has the property that $\text{marg}_{-jk}(\hat{\hat{\sigma}}) = \text{marg}_{-jk}(\sigma^{(k)})$. To see this, note first that because $\hat{\hat{\sigma}}$ and $\sigma^{(k)}$ agree on the $-(ijk)$ marginal $\text{marg}_{-ik}(\sigma^{(k)}) = \text{marg}_{-ik}(\sigma^{(i)})$. Also, by construction, the conditional distribution of i's action given profile of $(-ijk)$'s actions is the same.

From the hypothesis that $s_j^* \succsim s_k^*$, we get j finds s_j^* strictly optimal against $\hat{\hat{\sigma}}$.

But at the same time, $\text{marg}_{-i}(\hat{\hat{\sigma}}) = \text{marg}_{-i}(\sigma^{(i)})$ by construction, so this implies also $\text{marg}_{-ij}(\hat{\hat{\sigma}}) = \text{marg}_{-ij}(\sigma^{(i)})$. From $s_i^* \succsim s_j^*$, and the conclusion that j finds s_j^* strictly optimal against $\hat{\hat{\sigma}}$ just obtained, we get i finds s_i^* strictly optimal against $\sigma^{(i)}$ as desired. $\square$

10.2 Proof of Proposition 2

Proof. Suppose that $s_i^* \succsim s_j^*$ and that neither (ii) nor (iii) holds. We show that these assumptions imply $s_j^* \not\succsim s_i^*$.

Partition the set $\Delta^\circ(\mathbb{S})$ into three subsets, $\Pi^+ \cup \Pi^0 \cup \Pi^-$, with Π^+ consisting of $\sigma \in \Delta^\circ(\mathbb{S})$ that make s_j^* strictly better than the best alternative pure strategy, Π^0 the elements of $\Delta^\circ(\mathbb{S})$ that make s_j^* indifferent to the best alternative, and Π^- the elements that make s_j^* strictly worse. (These sets are well defined because $|\mathbb{S}_j| \geq 2$, so j has at least one alternative pure strategy to s_j^*.) If Π^0 is non-empty, then there is some $\sigma \in \Pi^0$ such that $\sum_{s \in \mathbb{S}} u_j(s_j^*, s_{-j})\sigma(s) = \max_{s_j' \in \mathbb{S}_j} \sum_{s \in \mathbb{S}} u_j(s_j', s_{-j})\sigma(s)$. Because $s_i^* \succsim s_j^*$, $\sum_{s \in \mathbb{S}} u_i(s_i^*, s_{-i})\hat{\sigma}(s) > \max_{s_i' \in \mathbb{S}_i \setminus \{s_i^*\}} \sum_{s \in \mathbb{S}} u_i(s_i' s_{-i})\hat{\sigma}(s)$ for every $\hat{\sigma} \in \Delta^\circ(\mathbb{S})$ such that $\text{marg}_{-ij}(\sigma) = \text{marg}_{-ij}(\hat{\sigma})$. Since at least one such $\hat{\sigma}$ exists, we do not have $s_j^* \succsim s_i^*$.

Also, if both Π^+ and Π^- are non-empty, then Π^0 is non-empty. This is because both $\sigma \mapsto \sum_{s\in\mathbb{S}} u_j(s_j^*, s_{-j})\sigma(s)$ and $\sigma \mapsto \max_{s_j^{'}\in\mathbb{S}_j\setminus\{s_j^*\}} \sum_{s\in\mathbb{S}} u_j(s_j^{'}, s_{-j})\sigma(s)$ are continuous functions. If $\sum_{s\in\mathbb{S}} u_j(s_j^*, s_{-j})\sigma(s) - \max_{s_j^{'}\in\mathbb{S}_j\setminus\{s_j^*\}} \sum_{s\in\mathbb{S}} u_j(s_j^{'}, s_{-j})\sigma(s) > 0$ and also $\sum_{s\in\mathbb{S}} u_j(s_j^*, s_{-j})\tilde{\sigma}(s) - \max_{s_j^{'}\in\mathbb{S}_j\setminus\{s_j^*\}} \sum_{s\in\mathbb{S}} u_j(s_j^{'}, s_{-j})\tilde{\sigma}(s) < 0$, then some mixture between σ and $\tilde{\sigma}$ must belong to Π^0.

So we have shown that if either Π^0 is non-empty or both Π^+ and Π^- are non-empty, then $s_j^* \not\succsim s_i^*$.

If only Π^+ is non-empty, then s_j^* is strictly interior dominant for j. Together with $s_i^* \succsim s_j^*$, this would imply that s_i^* is strictly interior dominant for i, contradicting the assumption that (iii) does not hold.

Finally suppose that only Π^- is non-empty, so that for every $\sigma \in \Delta^\circ(\mathbb{S})$ there exists a strictly better pure response than s_j^* against σ_{-j}. Then, from Lemma 4 of Pearce (1984), there is s a mixed strategy σ_j for j that weakly dominates s_j^* against all correlated strategy distributions. This σ_j strictly dominates s_j^* against strategy distributions in $\Delta^\circ(\mathbb{S}_{-j})$, so s_j^* is strictly interior dominated for j. Since (ii) does not hold, there is a $\sigma_{-i} \in \Delta^\circ(\mathbb{S}_{-i})$ against which s_i^* is a weak best response. Then, the fact that s_j^* is not a strict best response against any $\sigma_{-j} \in \Delta^\circ(\mathbb{S}_{-j})$ means $s_j^* \not\succsim s_i^*$. □

10.3 Proof of Lemma 3

Proof. We prove the first statement by contraposition. If $\mathscr{C}(\bar{\sigma})$ is not an $\mathscr{C}(\bar{\boldsymbol{\epsilon}})$-equilibrium in the base game, then some i assigns more than the required weight to some $s_i^{'} \in \mathbb{S}_i$ that does not best respond to $\mathscr{C}(\bar{\sigma})_{-i}$. This means no $(s_i', n_i) \in \bar{\mathbb{S}}_i$ best responds to $\bar{\sigma}_{-i}$, since all copies of a strategy are payoff equivalent. Since $\mathscr{C}(\bar{\sigma})$ and $\mathscr{C}(\bar{\boldsymbol{\epsilon}})$ are defined by adding up the respective extended-game probabilities, $\mathscr{C}(\bar{\sigma})_i(s_i^{'}) > \mathscr{C}(\bar{\boldsymbol{\epsilon}})(s_i^{'})$ means $\sum_{n_i} \bar{\sigma}_i(s_i', n_i) > \sum_{n_i} \bar{\boldsymbol{\epsilon}}(s_i', n_i)$. So for at least one $n_i^{'}$, $\bar{\sigma}_i(s_i', n_i^{'}) > \bar{\boldsymbol{\epsilon}}(s_i', n_i^{'})$, that is $\bar{\sigma}_i$ assigns more than required weight to the non best response $(s_i', n_i^{'}) \in \bar{\mathbb{S}}_i$. We conclude $\bar{\sigma}$ is not an $\bar{\boldsymbol{\epsilon}}$-equilibrium, as desired.

Again by contraposition, suppose $\mathscr{E}(\sigma)$ is not an $\mathscr{E}(\boldsymbol{\epsilon})$-equilibrium in the extended game. This means some i assigns more than the required weight to some $(s_i', n_i^{'}) \in \bar{\mathbb{S}}_i$ that does not best respond to $\mathscr{E}(\sigma)_{-i}$. This implies $s_i^{'}$ does not best respond to σ_{-i}. By the definition of $\mathscr{E}(\boldsymbol{\epsilon})$ and $\mathscr{E}(\sigma)$, if $\mathscr{E}(\sigma)_i(s_i', n_i^{'}) > \mathscr{E}(\boldsymbol{\epsilon})(s_i', n_i^{'})$, then also $\mathscr{E}(\sigma)_i(s_i', n_i) > \mathscr{E}(\boldsymbol{\epsilon})(s_i', n_i)$ for every n_i such that $(s_i', n_i) \in \bar{\mathbb{S}}_i$. Therefore, we also have $\sigma_i(s_i^{'}) > \boldsymbol{\epsilon}(s_i^{'})$, so σ is not an $\boldsymbol{\epsilon}$-equilibrium in the base game as desired. □

10.4 Proof of Proposition 7

Proof. Suppose $\bar{\sigma}^*$ is a PCE in the extended game. So, we have $\bar{\sigma}^{(t)} \to \bar{\sigma}^*$ where each $\bar{\sigma}^{(t)}$ is an $\bar{\boldsymbol{\epsilon}}^{(t)}$-PCE, and each $\bar{\boldsymbol{\epsilon}}^{(t)}$ is player-compatible (in the extended game sense). This means each $\mathscr{C}(\bar{\boldsymbol{\epsilon}}^{(t)})$ is player compatible in the base game sense, and furthermore each $\mathscr{C}(\bar{\sigma}^{(t)})$ is an $\mathscr{C}(\bar{\boldsymbol{\epsilon}}^{(t)})$-equilibrium (by Lemma 3), hence an $\mathscr{C}(\bar{\boldsymbol{\epsilon}}^{(t)})$-PCE. Since $\bar{\boldsymbol{\epsilon}}^{(t)} \to \mathbf{0}$, $\mathscr{C}(\bar{\boldsymbol{\epsilon}}^{(t)}) \to \mathbf{0}$ as well. Since $\bar{\sigma}^{(t)} \to \bar{\sigma}^*$, $\mathscr{C}(\bar{\sigma}^{(t)}) \to \mathscr{C}(\bar{\sigma}^*)$. We have shown $\mathscr{C}(\bar{\sigma}^*)$ is a PCE in the base game.

The proof of the other statement is exactly analogous. $\square$

10.5 Proof of Proposition 8

Proof. We have $r_i = \mathrm{OPT}_i$ if and only if for every $\tilde{y}_i \in \tilde{Y}_i$, $r_i(\psi(\tilde{y}_i))$ has the (weakly) higher Gittins index. Since $r_i, \tilde{r}_i$ are equivalent up to duplicates, this means for any $\tilde{y}_i \in \tilde{Y}_i$, $\tilde{r}_i(\tilde{y}_i)$ either puts probability 1 on **Out** or probability 1 on **In** and **In-d**. Since **In** and **In-d** can be viewed as two identical ways of pulling the risky arm in a two-armed bandit with one safe arm and one risky arm, $\tilde{r}_i$ is optimal if and only if $\tilde{r}_i(\tilde{y}_i)$ assigns positive probability 1 to **In** and **In-d** when the risky arm has a (weakly) higher Gittins index than the safe one. These two statements are equivalent when $\tilde{r}_i, r_i$ are equivalent up to duplicates, since the Gittins index of the risky arm is the same under $\tilde{y}_i$ and $\psi(\tilde{y}_i)$. Similarly, $r_i = \mathrm{WFP}_i$ if and only if for every $\tilde{y}_i \in \tilde{Y}_i$, $r_i(\psi(\tilde{y}_i))$ has the (weakly) higher "WFP" index, defined as the one-period expected payoff of playing a certain strategy against the weighted fictitious play conjecture of $-i$'s play. These indices are the same after history $\tilde{y}_i$ in the extended stage game and after $\psi(\tilde{y}_i)$ in the original stage game.

Finally, let X_i^t be the random variable representing i's play in period t in the base game under policy r_i and social distribution σ_{-i}. Let $\tilde{X}_i^t$ be the random variable representing i's play in period t in the extended game under policy $\tilde{r}_i$ and social distribution $\tilde{\sigma}_{-i}$. Because $r_i, \tilde{r}_i$ are equivalent up to duplicates to the empty history, $\mathbb{P}_{r_i,\sigma_{-i}}[X_i^1 = \mathbf{Out}] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\tilde{X}_i^1 = \mathbf{Out}]$. Since σ_{-i} and $\tilde{\sigma}_{-i}$ are $-i$ equivalent, (r_i, σ_{-i}) and $(\tilde{r}_i, \tilde{\sigma}_{-i})$ generate the same distribution over length-1 histories (up to duplicates), i.e. $\mathbb{P}_{r_i,\sigma_{-i}}[y_i] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i)]$ for all $y_i \in (\{\mathbf{In}, \mathbf{Out}\} \times \mathbb{R})$. By induction suppose $\mathbb{P}_{r_i,\sigma_{-i}}[y_i] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i)]$ for all $y_i \in (\{\mathbf{In}, \mathbf{Out}\} \times \mathbb{R})^t$, for some $t \geq 1$. If $r_i(y_i) = \mathbf{Out}$, then using the fact that $r_i, \tilde{r}_i$ are equivalent up to duplicates, $\tilde{r}_i(\tilde{y}_i)(\mathbf{Out}) = 1$ for all $\tilde{y}_i \in \psi^{-1}(y_i)$. Thus, for all $x \in \mathbb{R}$, by the inductive hypothesis $\mathbb{P}_{r_i,\sigma_{-i}}[(y_i, \mathbf{Out}, x)] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i) \times (\mathbf{Out}, x)]$, and $\mathbb{P}_{r_i,\sigma_{-i}}[(y_i, \mathbf{In}, x)] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i) \times (\mathbf{In}, x)] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i) \times (\textbf{In-d}, x)] = 0$. On the other hand, if $r_i(y_i) = \mathbf{In}$, then using the fact that $r_i, \tilde{r}_i$ are equivalent up to duplicates, $\tilde{r}_i(\tilde{y}_i)(\mathbf{In}) + \tilde{r}_i(\tilde{y}_i)(\textbf{In-d}) = 1$ for all $\tilde{y}_i \in \psi^{-1}(y_i)$. Thus, for all $x \in \mathbb{R}$, by the inductive hypothesis, $\mathbb{P}_{r_i,\sigma_{-i}}[(y_i, \mathbf{Out}, x)] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i) \times (\mathbf{Out}, x)] = 0$, and $\mathbb{P}_{r_i,\sigma_{-i}}[(y_i, \mathbf{In}, x)] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i) \times (\mathbf{In}, x)] + \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i) \times (\textbf{In-d}, x)]$.

In either case, we get $\mathbb{P}_{r_i,\sigma_{-i}}[y_i] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i)]$ for all $y_i \in (\{\mathbf{In}, \mathbf{Out}\} \times \mathbb{R})^{t+1}$, and also $\mathbb{P}_{r_i,\sigma_{-i}}[X_i^t = \mathbf{Out}] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\tilde{X}_i^t = \mathbf{Out}]$. By induction we get $\mathbb{P}_{r_i,\sigma_{-i}}[X_i^t = \mathbf{Out}] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\tilde{X}_i^t = \mathbf{Out}]$ for every $t \geq 1$, thus $\phi_i(\mathbf{In}; r_i, \sigma_{-i}) = \phi_i(\mathbf{In}; \tilde{r}_i, \tilde{\sigma}_{-i}) + \phi_i(\mathbf{In\text{-}d}; \tilde{r}_i, \tilde{\sigma}_{-i})$. $\square$

11 Refinements in the Link-Formation Game

Proposition 10. *Each of the following refinements selects the same subset of pure Nash equilibria when applied to the anti-monotonic and co-monotonic versions of the link-formation game: extended proper equilibrium, proper equilibrium, trembling-hand perfect equilibrium, p-dominance, Pareto efficiency, and strategic stability. Pairwise stability does not apply to the link-formation game. Finally, the link-formation game is not a potential game.*

Proof. **Step 1. Extended proper equilibrium, proper equilibrium, and trembling-hand perfect equilibrium allow the "no links" equilibrium in both versions of the game.** For (q_i) anti-monotonic with (c_i), for each $\epsilon > 0$ let N1 and S1 play **Active** with probability ϵ^2, N2 and S2 play **Active** with probability ϵ. For small enough ϵ, the expected payoff of **Active** for player i is approximately $(10 - c_i)\epsilon$ since terms with higher order ϵ are negligible. It is clear that this payoff is negative for small ϵ for every player i, and that under the utility re-scalings $\beta_{N1} = \beta_{S1} = 10$, $\beta_{N2} = \beta_{S2} = 1$, the loss to playing **Active** smaller for N2 and S2 than for N1 and S1. So this strategy profile is a $(\boldsymbol{\beta}, \epsilon)$-extended proper equilibrium. Taking $\epsilon \to 0$, we arrive at the equilibrium where each player chooses **Inactive** with probability 1.

For the version with (q_i) co-monotonic with (c_i), consider the same strategies without re-scalings, i.e. $\boldsymbol{\beta} = \mathbf{1}$. Then already the loss to playing **Active** smaller for N2 and S2 than for N1 and S1, making the strategy profile a $(\mathbf{1}, \epsilon)$-extended proper equilibrium.

These arguments show that the "no links" equilibrium is an extended proper equilibrium in both versions of the game. Every extended proper equilibrium is also proper and trembling-hand perfect, which completes the step.

Step 2. $p-$dominance eliminates the "no links" equilibrium in both versions of the game. Regardless of whether (q_i) are co-monotonic or anti-monotonic with (c_i), under the belief that all other players choose **Active** with probability p for $p \in (0, 1)$, the expected payoff of playing **Active** (due to additivity across links) is $(1-p) \cdot 0 + p \cdot (10 + 30 - 2c_i) > 0$ for any $c_i \in \{14, 19\}$.

Step 3. Pareto eliminates the "no links" equilibrium in both versions of the game. It is immediate that the no-links equilibrium outcome is Pareto dominated by the all-links equilibrium outcome under both parameter specifications, so Pareto efficiency would rule it out whether (c_i) is anti-monotonic or co-monotonic with (q_i).

Step 4. Strategic stability (Kohlberg and Mertens, 1986) **eliminates the "no links" equilibrium in both versions of the game.** First suppose the (c_i) are anti-monotonic with (q_i). Let $\eta = 1/100$ and let $\epsilon^{'} > 0$ be given. Define $\epsilon_{N1}(\textbf{Active}) = \epsilon_{S1}(\textbf{Active}) = 2\epsilon^{'}$, $\epsilon_{N2}(\textbf{Active}) = \epsilon_{S2}(\textbf{Active}) = \epsilon^{'}$ and $\epsilon_i(\textbf{Inactive}) = \epsilon^{'}$ for all players i. When each i is constrained to play s_i with probability at least $\epsilon_i(s_i)$, the only Nash equilibrium is for each player to choose **Active** with probability $1 - \epsilon^{'}$. (To see this, consider N2's play in any such equilibrium σ. If N2 weakly prefers **Active**, then N1 must strictly prefer it, so $\sigma_{N1}(\textbf{Active}) = 1 - \epsilon^{'} \geq \sigma_{N2}(\textbf{Active})$. On the other hand, if N2 strictly prefers **Inactive**, then $\sigma_{N2}(\textbf{Active}) = \epsilon^{'} < 2\epsilon^{'} \leq \sigma_{N1}(\textbf{Active})$. In either case, $\sigma_{N1}(\textbf{Active}) \geq \sigma_{N2}(\textbf{Active})$.) When both North players choose **Active** with probability $1 - \epsilon^{'}$, each South player has **Active** as their strict best response, so $\sigma_{S1}(\textbf{Active}) = \sigma_{S2}(\textbf{Active}) = 1 - \epsilon^{'}$. Against such a profile of South players, each North player has **Active** as their strict best response, so $\sigma_{N1}(\textbf{Active}) = \sigma_{N2}(\textbf{Active}) = 1 - \epsilon^{'}$.

Now suppose the (c_i) are co-monotonic with (q_i). Again let $\eta = 1/100$ and let $\epsilon^{'} > 0$ be given. Define $\epsilon_{N1}(\textbf{Active}) = \epsilon_{S1}(\textbf{Active}) = \epsilon^{'}$, $\epsilon_{N2}(\textbf{Active}) = \epsilon^{'}/1000$, $\epsilon_{S2}(\textbf{Active}) = \epsilon^{'}$ and $\epsilon_i(\textbf{Inactive}) = \epsilon^{'}$ for all players i. Suppose by way of contradiction there is a Nash equilibrium σ of the constrained game which is η-close to the **Inactive** equilibrium. In such an equilibrium, N2 must strictly prefer **Inactive**, otherwise N1 strictly prefers **Active** so σ could not be η-close to the **Inactive** equilibrium. Similar argument shows that S2 must strictly prefer **Inactive**. This shows N2 and S2 must play **Active** with the minimum possible probability, that is $\sigma_{N2}(\textbf{Active}) = \epsilon^{'}/1000$ and $\sigma_{S2}(\textbf{Active}) = \epsilon^{'}$. This implies that, even if $\sigma_{N1}(\textbf{Active})$ were at its minimum possible level of $\epsilon^{'}$, S1 would still strictly prefer playing **Inactive** because S1 is 1000 times as likely to link with the low-quality opponent as the high-quality opponent. This shows $\sigma_{S1}(\textbf{Active}) = \epsilon^{'}$. But when $\sigma_{S1}(\textbf{Active}) = \sigma_{S2}(\textbf{Active}) = \epsilon^{'}$, $N1$ strictly prefers playing **Active**, so $\sigma_{N1}(\textbf{Active}) = 1 - \epsilon^{'}$. This contradicts σ being η-close to the no-links equilibrium.

Step 5. Pairwise stability (Jackson and Wolinsky, 1996) **does not apply to this game**. This is because each player chooses between either linking with every player on the opposite side who plays **Active**, or linking with no one. A player cannot selectively cut off one of their links while preserving the other.

Step 6. The game does not have an ordinal potential, so refinements of potential games (Monderer and Shapley, 1996) **do not apply**. To see that this is not a potential game, consider the anti-monotonic parameterization. Suppose a potential P of the form $P(a_{N1}, a_{N2}, a_{S1}, a_{S2})$ exists, where $a_i = 1$ corresponds to i choosing **Active**, $a_i = 0$

corresponds to i choosing **Inactive**. We must have

$$P(0,0,0,0) = P(1,0,0,0) = P(0,0,0,1),$$

since a unilateral deviation by one player from the **Inactive** equilibrium does not change any player's payoffs. But notice that $u_{N1}(1,0,0,1) - u_{N1}(0,0,0,1) = 10 - 14 = -4$, while $u_{S2}(1,0,0,1) - u_{S2}(1,0,0,0) = 30 - 19 = 11$. If the game has an ordinal potential, then both of these expressions must have the same sign as $P(1,0,0,1) - P(1,0,0,0) = P(1,0,0,1) - P(0,0,0,1)$, which is not true. A similar argument shows the co-monotonic parameterization does not have a potential either. $\square$