Player-Compatible Equilibrium*

Drew Fudenberg† Kevin He‡

First version: September 23, 2017
This version: March 9, 2019

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 some near-optimal heuristics imply our compatibility restrictions in a steady-state setting.

Keywords: non-equilibrium learning, equilibrium refinements, trembling-hand perfect equilibrium, combinatorial bandits, Bayesian upper confidence bounds.

*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.

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

‡Department of Economics, Harvard University. 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 this approach by imposing cross-player restrictions on these trembles in a way that is invariant to the utility representations of players' preferences over game outcomes. The heart of this refinement is the concept of *player compatibility*, which says player i is more *compatible* with strategy s_i^* than player j is with strategy s_j^* if s_i^* is optimal for i against mixed play σ_{-i} of $-i$ whenever s_j^* is optimal for j against any σ_{-j} matching σ_{-i} in terms the play of the third parties, $-ij$. PCE requires that cross-player tremble magnitudes respect compatibility rankings. As we will explain, PCE interprets "trembles" as deliberate experiments to learn how others play, not as mistakes, and derives its cross-player tremble restrictions from an analysis of the relative frequencies of experiments that different players choose to undertake.

As an example, consider the following strategic link-formation game.[1] There are 4 players in the game, divided into two on each side, North and South. Each player simultaneously chooses between **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.) Each player pays a cost for each link she forms and gets a benefit that depends on the counterparties she links with. Thus each player has two parameters, "private linking cost" and "benefit to others." We consider two versions of the game. In the "anti-monotonic" version, low-cost players give high benefits to others. In the "co-monotonic" version, low-cost players give low benefits. We specify the parameters so that there are two pure-strategy Nash outcomes in each version of the game: "all links form" and "no links form."

PCE makes different predictions in these two versions of the link-formation game. Low-cost players are more compatible with **Active**, so PCE says they will "tremble" onto **Active** at least as much. This leads to more high-benefits players choosing **Active** in the anti-monotonic version, but more low-benefits players choosing it in the co-monotonic one. For this reason, the only PCE in the anti-monotonic version is "all **Active**," but both Nash outcomes are PCE outcomes in the co-monotonic version. In contrast, standard equilibrium refinements (such as proper equilibrium, p-dominance, and strategic stability) all make matching predictions in both versions of the game.

Section 2 defines PCE, studies their basic properties, and proves that PCE exist in all finite games. Because showing that player i is more compatible than player j requires considering all possible strategies of j, the compatibility relation is easiest to satisfy when

[1] The game is more fully described in Section 3, Example 2.

i and j are "non-interacting," meaning that their payoffs do not depend on each other's actions. But as we demonstrate in Section 3 using a "restaurant game," PCE can have bite even when all players interact with each other, provided that the interactions are not too strong. In fact, Section 3 uses a series of examples to show that 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, and also notes that PCE satisfies the compatibility criterion of Fudenberg and He (2018) in signaling games.

We then derive the cross-player compatibility restrictions on trembles from models of learning in steady-state environments. Specifically, we consider a learning framework where agents are born into different player roles of a stage game, and believe that they face an unknown, 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. Each agent only learns about others' play through her own payoffs at the end of the game. Because agents expect to play the game many times, they may choose to "experiment" and use myopically sub-optimal strategies for their informational value. The compatibility restriction on trembles then arises from the differences in the attractiveness of various experiments for different players. For example, in the strategic link-formation game, an agent choosing **Inactive** always receives the same payoff regardless of others' play, and hence observes no information about such play. So, they may try playing **Active** even if their prior belief is that the low-benefits counterparty is more likely to play **Active** than the high-benefits one. As is intuitive, we show that the low-cost agent has a stronger incentive to experiment with **Active** than the high-cost one, and will do so more frequently against any mixed play of the counterparties.

To make the analysis more tractable, Section 5 restricts attention to a class of "factorable" games, where repeatedly playing a given strategy s_i would reveal all of the payoff consequences of that strategy and no information about the payoff consequences of any other strategy $s_i' \neq s_i$. This restriction implies that at any strategy profile s, if player i potentially cares about the action taken at some information set H of $-i$, then either H is on the path of s or i can put H onto the path of play via a unilateral deviation. Thus there is no possibility of learning being "blocked" by other players, and no "free riding" by learning from others' experiments. For simplicity we also require that each player moves at most once along any path of play. The strategic link-formation game , signaling games (with different types viewed as different players), and the restaurant game mentioned before all satisfy these restrictions (for generic extensive-form payoffs).

In factorable games, each agents faces a *combinatorial bandit problem* (see Section 5.2). We consider two related models of how agents deal with the trade-off between exploration and exploitation — the classic model of rational Bayesians maximizing discounted expected utility under the belief that the environment (the aggregate strategy distribution in the population) is constant, and the computationally simpler method of Bayesian upper-confidence bounds[2] (Kaufmann, Cappé, and Garivier, 2012). In both of these models, the agent uses an "index policy," meaning that they assign a numerical index to each strategy that depends only on past observations when that strategy was used, and then chooses the strategy with the highest index. We formulate a compatibility condition for index policies, and show that any index policies for i and j satisfying this compatibility condition for strategies s_i^* and s_j^* will lead to i experimenting relatively more with s_i^* than j with s_j^*. To complete the micro-foundation of PCE, we then show that the Bayes optimal policy and the Bayes-UCB heuristic satisfy the compatibility condition for strategies s_i^* and s_j^* whenever i is more compatible s_i^* than player j is with strategy s_j^* and the agents in roles i and j face comparable learning problems (e.g. start with the same patience level, same prior beliefs about the play of third parties, etc).

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

[2]Briefly, upper confidence bound algorithms originated as computationally tractable algorithms for multi-armed bandit problems (Agrawal, 1995; Katehakis and Robbins, 1995). We consider a Bayesian version of the algorithm that keeps track of the learner's posterior beliefs about the payoffs of different strategies, first analyzed by Kaufmann, Cappé, and Garivier (2012). We say more about this procedure in Section 5. See Francetich and Kreps (2018) for a discussion of other heuristics for active learning.

worse payoffs are played less often. I too is the limit of r 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 the central feature of PCE, proper equilibrium and approachable equilibrium do not impose cross-player restrictions on the relative probabilities of various trembles. For this reason, when each types of the sender is viewed as a different player 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). 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 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, when discussing tremble-based refinements in Bayesian games we will always apply them at the interim stage.

Like PCE, *extended proper equilibrium* (Milgrom and Mollner, 2017) 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 her 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 6 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, as shown by example in Online Appendix OA 4.[3]

[3]Extended proper equilibrium is related to *test-set equilibrium* (Milgrom and Mollner, 2018), which is not defined in terms of trembles.

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 (2017), Esponda and Pouzo (2016), and Lehrer (2012). Unlike those papers, we focus on the very patient agents who undertake many "experiments," and characterize the relative rates of experimentation under rational expected-utility maximization and related "near-optimal" heuristics. 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 rule (rational Bayesians) in a restricted set of games (signaling games). In contrast, our analysis applies to a broader class of learning rules — specifically, index policies that satisfy the relative compatibility condition in Definition 18, and to a larger family of games, the factorable games defined in Section 4. We develop new tools to deal with new issues that arise in this more general setting. For instance, Fudenberg and He (2018) compare the Gittins indices of different sender types 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 valid to exchange the stopping times of different players as they may condition on observed play in different parts of the game tree. We deal with this problem by considering how i can nevertheless construct a feasible stopping time that *mimics* an unfeasible one of j. Moreover, when a player faces more than one opponent, their optimal experimentation policy may lead them to see correlated play by their opponents even though the opponents do no actually play correlated strategies. We deal with this complication as well.

In methodology the paper is related to other work on active learning and experimentation, both in single-agent setting such as Doval (2018), Francetich and Kreps (2018), Fryer and Harms (2017), and in multi-agents ones such as Bolton and Harris (1999), Keller et al. (2005), Klein and Rady (2011), Heidhues, Rady, and Strack (2015), Halac, Kartik, and Liu (2016), and Strulovici (2010). Unlike the papers on multi-agent bandit problems, our agents only learn from personal histories, not from the actions or histories of others. But our focus also differs from the existing work on single-agent experimentation. We compare experimentation dynamics under different payoff parameters, which correspond to different players' learning problems. This comparison is central to PCE's refinement based on cross-player tremble restrictions.

2 Player-Compatible Equilibrium

2.1 Compatibility

In this section, we first define the player-compatibility relation and discuss its basic properties. We then introduce PCE, which embodies cross-player tremble restrictions based on this relation.

Consider a strategic-form game with finite number of players $i \in \mathbb{I}$, finite strategy sets $|\mathbb{S}_i| \geq 2$,[4] and utility functions $U_i : \mathbb{S} \to \mathbb{R}$, where $\mathbb{S} := \times_i \mathbb{S}_i$. We assume no player has a strictly dominated strategy, which lets us avoid some complications that would otherwise need to be treated separately. This assumption is consistent with the learning foundation we provide for PCE, because that foundation considers settings where playing one strategy s_i gives no information about the payoff consequences of any other strategy $s_i' \neq s_i$. Thus strictly dominated strategies will never be played, even as experiments, so they may be deleted from the game.

For each i, let $\Delta(\mathbb{S}_i)$ denote the set of mixed strategies and $\Delta^\circ(\mathbb{S}_i)$ the set of strictly mixed strategies, where every pure strategy in $\mathbb{S}_i$ is assigned positive probability. For $K \subseteq \mathbb{I}$, let $\Delta(\mathbb{S}_K)$ represent the set of correlated strategies among players K, i.e. the set of distributions on strategy profiles of players in coalition K, $\times_{i \in K} \mathbb{S}_i$. Let $\Delta^\circ(\mathbb{S}_K)$ represent the interior of $\Delta(\mathbb{S}_K)$, that is the set of full-support correlated strategies on S_K.[5]

We formalize the concept of "compatibility" between players and their strategies in this general setting, which will play a central role in the definition of PCE in determining cross-player restrictions on trembles.

Definition 1. For player $i \neq j$ and strategies $s_i^* \in \mathbb{S}_i$, $s_j^* \in \mathbb{S}_j$, say *i is more compatible with s_i^* than j is with s_j^**, abbreviated as $(s_i^* \mid i) \succsim (s_j^* \mid j)$, if for every correlated strategy $\sigma_{-j} \in \Delta^\circ(\mathbb{S}_{-j})$ such that

$$U_j(s_j^*, \sigma_{-j}) \geq \max_{s_j' \in \mathbb{S}_j \setminus \{s_j^*\}} U_j(s_j', \sigma_{-j}),$$

we have for *every* $\sigma_{-i} \in \Delta^\circ(\mathbb{S}_{-i})$ satisfying $\sigma_{-i}|_{\mathbb{S}_{-ij}} = \sigma_{-j}|_{\mathbb{S}_{-ij}}$,

$$U_i(s_i^*, \sigma_{-i}) > \max_{s_i' \in \mathbb{S}_i \setminus \{s_i^*\}} U_i(s_i', \sigma_{-i}).$$

In words, if s_j^* is weakly optimal for the less-compatible j against the opponents' correlated strategy σ_{-j}, then s_i^* is strictly optimal for the more-compatible i against any correlated

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

[5] Recall that a full-support correlated strategy assigns positive probability to every pure strategy profile.

strategy σ_{-i} of $-i$ that matches σ_{-j} in terms of the play of $-ij$. As this restatement makes clear, the compatibility condition only depends on players' preferences over probability distribution on $\mathbb{S}$, and not on the particular utility representations chosen.

Since $\times_{k\in K}\Delta^\circ(\mathbb{S}_k) \subseteq \Delta^\circ(\mathbb{S}_K)$, 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. [6] We use the more stringent definition so that we can microfound our compatibility-based cross-player restrictions on a broader set of learning models.

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

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

The proof of this result is in the appendix, as are all other omitted proofs.

The next result states that the compatibility relation is asymmetric.[7]

Proposition 2. *If* $(s_i^* \mid i) \succsim (s_j^* \mid j)$*, then* $(s_j^* \mid j) \not\succsim (s_i^* \mid i)$*.*

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.

Proposition 3. *If the game only has two players,* $i \neq j$*, then* $(s_i^* \mid i) \succsim (s_j^* \mid j)$ *never holds, so every trembling-hand perfect equilibrium is a PCE.*

If players i and j care a great deal about each other's strategies, then their best responses are unlikely to be determined only by the play of the third parties. In the other extreme, 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 i's payoff does not depend on the strategy of j and j's payoff does not depend on the strategy of i; (2) they have the same strategy set, $\mathbb{S}_i = \mathbb{S}_j$. As a leading case, every Bayesian game has the multipartite structure when each type is viewed as a different player. In addition, the "link-formation game" in Example 2 is a complete-information game with the multipartite structure. In a game with multipartite structure with $i, j \in \mathbb{I}_c$, we may write $U_i(s_i, s_{-ij})$ without ambiguity, since all augmentations of the strategy profile s_{-ij} with a strategy by player j lead to the same payoff for i. For

[6] Formally, this alternative definition would be "For every $\sigma_{-ij} \in \times_{k\neq i,j}\Delta^\circ(\mathbb{S}_k)$ such that $U_j(s_j^*, \hat{\sigma}_i, \sigma_{-ij}) \geq \max_{s_j' \in \mathbb{S}_j \setminus \{s_j^*\}} U_j(s_j', \hat{\sigma}_i, \sigma_{-ij})$ for some $\hat{\sigma}_i \in \Delta^\circ(\mathbb{S}_i)$, we have for every $\hat{\sigma}_j \in \Delta^\circ(\mathbb{S}_j)$ that $U_i(s_i^*, \hat{\sigma}_j, \sigma_{-ij}) > \max_{s_i' \in \mathbb{S}_i \setminus \{s_i^*\}} U_i(s_i', \hat{\sigma}_j, \sigma_{-ij})$."

[7] If the game can contain strictly dominated strategies or strictly dominant strategies, $(s_i^* \mid i) \sim (s_j^* \mid j)$ if both strategies are strictly dominated for their respective players (so that the "if" clause of the definition is never satisfied) or if both are strictly dominant. .

$s_c^* \in \mathbb{S}_i = \mathbb{S}_j$, the definition of $(s_c^* \mid i) \succsim (s_c^* \mid j)$ reduces to: For every strictly mixed correlated strategy $\sigma_{-ij} \in \Delta^\circ(\mathbb{S}_{-ij})$ such that

$$U_j(s_c^*, \sigma_{-ij}) \geq \max_{s_j' \in \mathbb{S}_j \setminus \{s_c^*\}} U_j(s_j', \sigma_{-ij}),$$

we have

$$U_i(s_c^*, \sigma_{-ij}) > \max_{s_i' \in \mathbb{S}_i \setminus \{s_c^*\}} U_i(s_i', \sigma_{-ij}).$$

While the player-compatibility condition is especially easy to state for non-interacting players, our learning foundation will also justify cross-player tremble restrictions for pairs of players i, j whose payoffs do depend on each others' strategies, as in the players in the "restaurant game" we discuss in Example 1.

2.2 Player-Compatible Trembles and PCE

We now move towards the definition of PCE. PCE is a tremble-based solution concept. It builds on and modifies Selten (1975)'s definition of trembling-hand perfect equilibrium as the limit of equilibria of 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 $\boldsymbol{\epsilon}(s_i \mid i) > 0$ to every player i and pure strategy s_i. Given a tremble profile $\boldsymbol{\epsilon}$, write $\Pi_i^{\boldsymbol{\epsilon}}$ for the set of $\boldsymbol{\epsilon}$-*strategies* of player i, namely:

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

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

$$\sigma_i^\circ \in \arg\max_{\sigma_i \in \Pi_i^{\boldsymbol{\epsilon}}} U_i(\sigma_i, \sigma_{-i}^\circ).$$

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 $\boldsymbol{\epsilon}(s_i^* \mid i) \geq \boldsymbol{\epsilon}(s_j^* \mid j)$ for all i, j, s_i^*, s_j^* such that $(s_i^* \mid i) \succsim (s_j^* \mid j)$. 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,

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

This is a "cross-player tremble restriction," that is, a restriction on the relative probabilities of trembles by different players. Note that it, like the compatibility relation, depends on the players' preferences over distributions on $\mathbb{S}$ but not on the particular utility representation used. This invariance property distinguishes player-compatible trembles from other models of stochastic behavior such as the stochastic terms in logit best responses.

If $(s_i^* \mid i) \succsim (s_j^* \mid j)$, then in every $\boldsymbol{\epsilon}$-PCE either i puts more weight on s_i^* than j puts on s_j^*, or i is putting the maximum possible amount of weight on s_i^*, subject to the minimum required weights on all her other strategies.

Lemma 1. If σ° is an $\boldsymbol{\epsilon}$-PCE and $(s_i^* \mid i) \succsim (s_j^* \mid j)$, then

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

While an $\boldsymbol{\epsilon}$-equilibrium always exists provided $\boldsymbol{\epsilon}$ is close enough to $\mathbf{0}$, these $\boldsymbol{\epsilon}$-equilibria need not satisfy the conclusion of Lemma 1 when the tremble profile $\boldsymbol{\epsilon}$ is not player compatible. We illustrate this in Example 3 of the Online Appendix.

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 $\sigma^{(t)} \to \sigma^*$.

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

Proposition 4. *For any PCE σ^*, player k, and strategy $\bar{s}_k$ such that $\sigma_k^*(\bar{s}_k) > 0$, there exists a sequence of strictly mixed strategy profiles $\sigma_{-k}^{(t)} \to \sigma_{-k}^*$ such that*

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

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

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

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

This result follows from Lemma 1, which shows every $\boldsymbol{\epsilon}$-PCE respects player compatibility up to the "adding up constraint" that probabilities on different actions must sum up to 1 and i must place probability no smaller than $\boldsymbol{\epsilon}(s_i' \mid i)$ on actions $s_i' \neq s_i^*$. 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, 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. *PCE exists in every finite strategic-form game.*

2.3 Some Properties of PCE

A tremble profile $\boldsymbol{\epsilon}$ is *uniform* if for all i and $s_i \in \mathbb{S}_i$, we have $\boldsymbol{\epsilon}(s_i \mid i) = \epsilon$ for the same $\epsilon > 0$. A trembling-hand perfect equilibrium is a *uniform THPE* if it is the limit of $\boldsymbol{\epsilon}$-equilibria where $\boldsymbol{\epsilon} \to 0$ and each tremble profile $\boldsymbol{\epsilon}$ is uniform. The proof of Theorem 1 in fact establishes the existence of uniform THPE, which form a subset of PCE since uniform trembles are always player compatible regardless of the stage game.

One drawback of uniform THPE is that there is no clear microfoundation for uniform trembles. In addition to the cross-player restrictions of the compatibility condition, these uniform trembles impose the same lower bound on the tremble probabilities for all strategies of each given player. PCE and the learning foundation we develop allow for more complicated patterns of experimentation that respect the compatibility structure. (Appendix OA 3.1 exhibits a PCE that is not a uniform THPE.) We study a more permissive refinement than uniform THPE where we can offer a learning story for the tremble restrictions. PCE is a fairly weak solution concept that nevertheless has bite in some cases of interest, as we will discuss in Sections 3.

In Online Appendix OA 2, we show that PCE is invariant to the replication of strategies: We expand the game to include duplicate copies of some of the original strategies. If $(s_i^* \mid i) \succsim (s_j^* \mid j)$ 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^*. We show that the set of PCE in the expanded game is the same as the set of PCE in the original game.

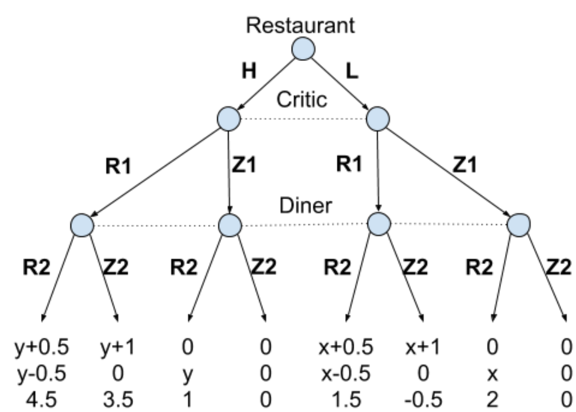
3 Examples of PCE

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 (P1), a regular diner (P2), and a restaurant (P3). Simultaneously, the restaurant decides between ordering high-quality (**H**) or low-quality (**L**) ingredients, while 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. This situation is depicted in the game tree below.



The strategies of the two customers affect each others' payoffs, so P1 and P2 are not non-interacting players. In particular, the critic and the diner cannot be mapped into two types of the same agent in some Bayesian game.

The strategy profile (**Z1**, **Z2**, **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 as she gets paid for writing reviews. It is also an extended proper equilibrium and a test set equilibrium.[8]

It is easy to verify that $(\mathbf{R1}|\text{Critic}) \succsim (\mathbf{R2}|\text{Diner})$. For any σ_{-2} of strictly mixed correlated play by the critic and the restaurant that makes the diner indifferent between **Z2** and **R2**, we must have $U_1(\mathbf{R1}, \sigma_{-1}) \geq 0.5$ for any σ_{-1} that agrees with σ_{-2} in terms of the restaurant's play. This is because the critic's utility from **R1** is minimized when the diner chooses **R2** with probability 1, but even then the critic gets 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 $(\mathbf{R1}|\text{Critic}) \succsim (\mathbf{R2}|\text{Diner})$. Whenever $\sigma_1^{(t)}(\mathbf{R1})/\sigma_2^{(t)}(\mathbf{R2}) > \frac{1}{4}$, the restaurant strictly prefers **H** over **L**. Thus by Proposition 4, there is no PCE where the restaurant plays **L** with positive probability. ♦

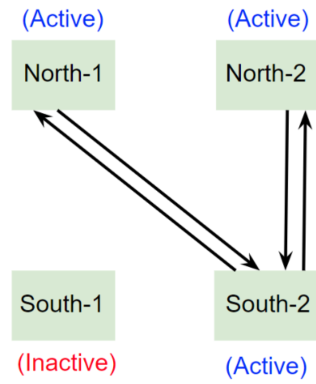
3.2 The Link-Formation Game

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 respectively.

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.

[8] (**Z1**, **Z2**, **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**. To see that it is a test-set equilibrium, note that both the critic and the diner have strict incentives in the equilibrium, so their equilibrium strategies are trivially undominated in the test set. The strict incentives for the customers imply that the only test-set opponent strategy for the restaurant is the customers' equilibrium play, so the restaurant's strategy is also undominated in the test set.



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 she creates. 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 $q_{\text{S2}} - c_{\text{N1}}$ and the payoff to South-2 is $(q_{\text{N1}} - c_{\text{S2}}) + (q_{\text{N2}} - c_{\text{S2}})$. ♦

We consider two versions of this game, shown below. In the anti-monotonic version on the left, players with a higher cost also have a lower quality. In the co-monotonic version on the right, players with a higher cost also 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 because the compatibility structure with respect to own quality is reversed between these two versions of the game. In both versions, (Active | N1) $\succsim$ (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.

By contrast, standard refinements (extended proper equilibrium, proper equilibrium, trembling-hand perfect equilibrium, p-dominance, Pareto efficiency, strategic stability, pairwise stability) all make matching predictions in both versions of the game.

Proposition 5. Each of the following refinements selects the same subset of pure Nash equilibria outcomes 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, strategic stability, and pairwise stability. Moreover the link-formation game is not a potential game.

The following table list the predictions of each of these solution concepts.

Solution Concept	Anti-Monotonic		Co-Monotonic	
	All Links	No Links	All Links	No Links
Trembling-hand perfect	✓	✓	✓	✓
Proper	✓	✓	✓	✓
Extended Proper	✓	✓	✓	✓
p-dominance	✓		✓	
Pareto	✓		✓	
Strategic Stability	✓		✓	
PCE	✓		✓	✓

Table 1: Nash equilibrium outcomes that survive different solution concepts in the two versions of the link-formation game. Only PCE makes distinct predictions in the two versions of the game.

3.3 Signaling Games

Recall that a signaling game is a two-player Bayesian game, where P1 is a sender who knows her 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 her 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 $\sigma_1(\cdot|\theta) \in \Delta(S)$, 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 θ' 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 θ' and s^* is not equilibrium dominated[9] for θ, p satisfies $\frac{p(\theta')}{p(\theta)} \leq \frac{\lambda(\theta')}{\lambda(\theta)}$.

Since every strictly mixed strategy of the receiver is payoff-equivalent to a behavioral strategy for the sender, it is easy to see that type compatibility implies $(s^*|\theta) \succsim (s^*|\theta')$.[10] The next result shows that when specialized to signaling games, all PCE pass the compatibility criterion.

Proposition 6. *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 Factorability and Isomorphic Factoring

This section defines a "factorability" condition that we will use in developing a learning foundation for PCE. Factorability implies that the information gathered from playing one strategy is not at all informative about the payoff consequences of any other strategy. We then define a notion of "isomorphic factoring" for players i and j to formalize the idea that the learning problems faced by these two players are essentially the same. The next section will provide a learning foundation for the compatibility restriction for pairs of players whose learning problems are isomorphic in this way. The examples discussed in Section 3 are factorable and isomorphically factorable for players ranked by compatibility.

[9] Signal s^* is not equilibrium dominated for θ if $\max_{a \in A} u_1(s^*, a; \theta) > u_1(s_1, \sigma_2^*; \theta)$, where s_1 is any on-path signal for type θ, $\sigma_1^*(s_1 \mid \theta) > 0$.

[10] The converse does not hold. Type compatibility requires testing against all receiver strategies and not just the strictly mixed ones, so it is possible that $(s^* \mid \theta) \succsim (s^* \mid \theta')$ but s^* is not more type-compatible with θ than with θ'.

4.1 Definition and Motivation

We begin by introducing some notation. Fix an extensive-form game Γ as the stage game, with players $i \in \mathbb{I}$ along with a player 0 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. So 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. For simplicity, throughout we will maintain the following assumption.

Assumption 1. *Each player moves at most once along any path of play in* Γ.

In addition to any information a player gets in the course of play, we assume that after each play each player observes her own payoff. In general, this need not perfectly reveal other players' actions at all information sets. We now define factorability, which roughly says that playing strategy s_i against any strategy profile of $-i$ identifies all of opponents' actions that can be payoff-relevant for s_i, but does not reveal any information about the payoff consequences of any other strategy $s_i^{'} \neq s_i$.

For an information set H of j with $j \neq i$, write P_H for the partition on $\mathbb{S}_{-i}$ where two strategies $s_{-i}, s_{-i}^{'}$ belong to the same block in P_H if and only if $s_{-i}(H) = s_{-i}^{'}(H)$. Thus partition P_H contains perfect information about play on H, but no other information.

Definition 5. 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* 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 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, $(a_H)_{H \in \mathcal{H}_{-i}} \mapsto U_i(s_i, (a_H)_{H \in \mathcal{H}_{-i}})$ is a one-to-one function of the components $(a_H)_{H \in F_i[s_i]}$, but is not a function of the other components. The substantive restriction in Condition (1) is that i's learning cannot be blocked by another player — by choosing s_i, i can always identify actions on $F_i[s_i]$ regardless of what happens elsewhere in the game tree.[11]

[11] 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.

Condition (2) implies that i does not learn about 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.[12]

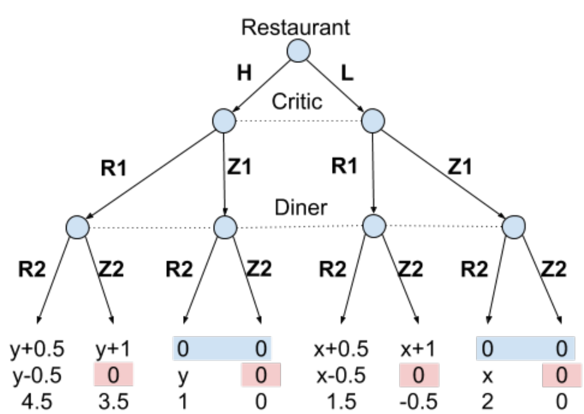
If $F_i[s_i]$ is empty, then s_i is a kind of "opt out" action for i. After choosing s_i, i receives the same utility from every reachable terminal node and gets no information about the payoff consequences of any of her other strategies.

4.2 Examples of Factorable Games

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

4.2.1 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.

[12] Bolton and Harris (1999) and Heidhues, Rady, and Strack (2015) among others analyze strategic experimentation in an equilibrium setting using very specific specifications of the signal structure. There is also an extensive literature on optimal contracts for experimentation in various sorts of agency problems, see e.g. Hörner and Skrzypacz (2016).

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 herself.

4.2.2 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. It is now easy to see that we can let F_i[**Active**] consists of the information sets of the other two agents on the other side of i and let F_i[**Inactive**] be empty. This specification of the s_i-relevant information sets shows the stage game is factorable for every player.

4.2.3 Binary Participation Games

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

Definition 6. 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 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, then let F_i[**In**] be the common collection of $-i$ information sets encountered in paths of play where i chooses **In**. Let F_i[**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[**In**] must lead to different terminal nodes, 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 with 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.

4.2.4 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 $\lambda \in \Delta(\Theta)$, where Θ is a finite set. 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 is because for each type i we have $F_i[s]$ is the set of n_r information sets by the receivers after seeing signal s.

4.3 Examples of Non-Factorable Games

The next result gives a necessary condition for factorability. 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. 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 her information sets.

Proposition 7. *Suppose 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 her information sets that puts H^* onto the path of play.*

This result follows from two lemmas.

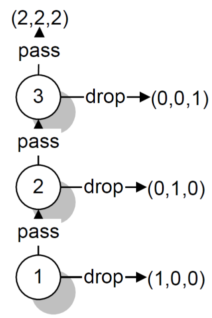
Lemma 2. *For any game factorable for i and any information set H^* for player $j \neq i$ where j has at least two different actions, if $H^* \in F_i[s_i]$ for some extensive-form strategy $s_i \in \mathbb{S}_i$, then H^* is always on the path of play when i chooses s_i.*

Lemma 3. *For any game factorable for i and any information set H^* of player $j \neq i$, suppose i's payoff is not independent of H^*. Then: (i) j has at least two different actions on H^*; (ii) there exists some extensive-form strategy $s_i \in \mathbb{S}_i$ so that $H^* \in F_i[s_i]$.*

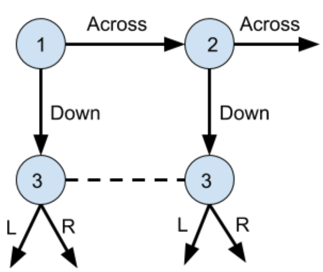
We can combine these two lemmas to prove the proposition.

Proof. By combining Lemmas 2 and 3, there exists some extensive-form strategy $s_i \in \mathbb{S}_i$ so 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 deviation to s_i, and H^* 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. □

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 7 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 outcome (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 7 is satisfied. The information set of 3 must belong to both $F_1[\text{Down}]$ and $F_1[\text{Across}]$, because 3's play can affect 1's payoff even if 1 chooses Across, as 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 $F_2[\text{Down}]$ and $F_2[\text{Across}]$, 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 rules (including rational learning) will not use such strategies, but we do not capture that in the general definition of PCE because our setup there only consider abstract strategy spaces $\mathbb{S}_i$ and not an extensive-form game tree.[13]

4.4 Isomorphic Factoring

Before we turn to compare the learning behavior of agents i and j, we must deal with one final issue. To make sensible comparisons between strategies s_i^* and s_j^* of two different players $i \neq j$ in a learning setting, we must make assumptions on their informational value about the play of others: namely, the information i gets from choosing s_i^* must be essentially the same as the information that j gets from choosing s_j^*. To do this we require that the game be factorable for both i and j, and that t the factoring is "isomorphic" for these two players.

Definition 7. 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 $s_i \in \mathbb{S}_i$.

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 $F_1[\mathbf{In1}] \cap \mathcal{H}_3 = F_2[\text{In2}] \cap \mathcal{H}_3 =$ the singleton set containing the unique information set of the restaurant. As another example, all signaling games (with possibly many receivers as in Section 4.2.4) 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).

5 Learning Foundations

In this section, we provide a learning foundation for PCE's cross-player tremble restrictions. Our main learning result, Theorem 2, studies long-lived agents who get permanently assigned

[13] 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).

into player roles and face a fixed but unknown distribution of opponents' play. We prove that when $(s_i^* \mid i) \succsim (s_j^* \mid j)$ and the game is isomorphically factorable for i and j, agents in the role of i use s_i^* more frequently than agents in the role of j use s_j^*. We obtain this result both for rational agents who maximize discounted expected utility, and for boundedly-rational agents who employ the computationally simpler Bayes-upper confidence bound algorithm. Under either of these behavioral assumptions, "trembles" emerge endogenously during learning as deliberate experiments that seek to learn opponents' play.

5.1 Learning Rules and Learning Environments

We consider an agent born into player role i who maintains this role throughout her life. She has a geometrically distributed lifetime with $0 \leq \gamma < 1$ probability of survival between periods. Each period, the agent plays the stage game Γ, choosing a strategy $s_i \in \mathbb{S}_i$. The agent observes and collects her payoffs at the end of the game. Then, with probability γ, she continues into the next period and plays the stage game again. With complementary probability, she exits the system. Thus each period the agent observes her own payoff. [14] We assume that players have perfect recall, so she also remembers her chosen strategy.

Definition 8. The set of all finite *histories* of all lengths for i is $Y_i := \cup_{t\geq 0}(\mathbb{S}_i \times \mathbb{R})^t$. 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.

When Γ is factorable for i, there is a one-to-one mapping from the set of action profile on the s_i-relevant information sets to the range of $s_{-i} \mapsto U_i(s_i, s_{-i})$, as required by Definition 5, Condition (1). Through this identification, we may think of each one-period history where i plays s_i as an element of $\{s_i\} \times (\times_{H \in F_i[s_i]} A_H)$ instead of an element of $\{s_i\} \times \mathbb{R}$. This convention will make it easier to compare histories of different player roles.

Notation 1. A history y_i will also refer to an element of $\cup_{t\geq 0} \left(\cup_{s_i \in \mathbb{S}_i} \left[\{s_i\} \times (\times_{H \in F_i[s_i]} A_H)\right]\right)^t$.

The agent decides on which strategy to use in each period based on her history so far. This mapping is her *learning rule*.

Definition 9. A *learning rule* $r_i : Y_i \to \mathbb{S}_i$ specifies a pure strategy in the stage game after each history.

Since the agent's play in each period depends on her past observations, the sequence of her plays is a stochastic process whose distribution depends on the distribution of the opponents' play. We assume that there is a fixed objective distribution of opponnent's play,

[14] This is a special case of the terminal-node partitions of Fudenberg and Kamada (2015, 2018) where the elements of each player's terminal node partition are isomorphic to their possible payoffs.

which we call player *i's learning environment.* The leading case of this is when there are multiple populations of learners, one for each player role, and the aggregate system is in a steady state. But, when analyzing the play of a single agent, we remain agnostic about the reason why opponents' play is i.i.d.

Definition 10. A *learning environment for player i* is a probability distribution $\sigma_{-i} \in \prod_{j \neq i} \Delta(S_j)$ over strategies of players $-i$.

The learning environment, together with the agent's learning rule, generate a stochastic process X_i^t describing i's strategy in period t.

Definition 11. Let X_i^t be the $\mathbb{S}_i$-valued random variable representing i's play in period t. Player i's *induced response* of i to σ_{-i} under learning rule r_i is $\phi_i(\cdot; r_i, \sigma_{-i}) : \mathbb{S}_i \to [0, 1]$, where for each $s_i \in \mathbb{S}_i$ we have

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

We can interpret the induced response $\phi_i(\cdot; r_i, \sigma_{-i})$ as a mixed strategy for i representing i's weighted lifetime average play, where the weight on X_i^t, the strategy she uses in period t of her life, is proportional to the probability γ^{t-1} of surviving into that period. The induced response has a population interpretation as well. Suppose there is a continuum of agents in the society, each engaged in their own copy of the learning problem above. In each period, enough new agents are added to the society to exactly balance out the share of agents who exit between periods. Then $\phi_i(\cdot; r_i, \sigma_{-i})$ describes the distribution on $\mathbb{S}_i$ we would find if we sample an individual uniformly at random from the subpopulation for role of i and ask her which $s_i \in \mathbb{S}_i$ they plan on playing today.

Our learning foundation for compatible tremblesinvolves comparing the induces responses of different player roles with the same learning rule and in the same learning environment.

5.2 Two Models of Learning and Experimentation

We will consider two different specifications of the agents' learning rules in factorable games, namely the maximization of expected discounted utility and the Bayes upper confidence bound heuristic. With both rules, agents form a Bayesian belief over opponents' play, independent at different information sets. More precisely, we will assume that each agent i starts with a *regular independent prior:*

Definition 12. Agent i has a *regular independent priors* if her 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 $g_i(\alpha_H) > 0$ for all $\alpha_H \in \Delta^\circ(A_H)$.

Thus, the agent holds a belief about the distribution of actions at each[15] $-i$ information set H, and thinks actions at different information sets are generated independently, whether the information sets belong to the same player or to different ones. Furthermore, the agent holds independent beliefs about the randomizing probabilities at different information sets. [16] The agent updates g_i by applying the Bayes' rule to her 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 response to a given signals s based on the responses received to that signal, and that their beliefs about this response do not depend on the responses they have observed to other signals $s^{'} \neq s$.

If i starts with independent prior beliefs in a stage game factorable for i, the learning problem she faces is a combinatorial bandit problem. A combinatorial bandit consist of a set of *basic arms*, each with an unknown distribution of outcomes, together with a collection of subsets of basic arms called *super arms*. Each period, the agent must choose a super arm, which results in pulling all of the basic arms in that subset and obtaining a utility based on the outcomes of these pulls. To translate into our language, each basic arm corresponds to a $-i$ information set H and the super arms are identified with strategies $s_i \in \mathbb{S}_i$. The subset of basic arms in s_i are the s_i-relevant information sets, $F_i[s_i]$. The collection of outcomes from these basic arms, i.e. the action profile $(a_H)_{H \in F_i[s_i]}$, determine i's payoff, $U_i(s_i, (a_H)_{H \in F_i[s_i]})$.

A special case of combinatorial bandits is additive separability, where the outcome from pulling each basic arm is simply a $\mathbb{R}$-valued reward, and the payoff from choosing a super arm is the sum of these rewards. This corresponds to the stage game being *additively separable* for i.

Definition 13. A factorable game Γ is *additively separable* for i if there is a collection of *auxiliary functions* $u_{i,H} : A_H \to \mathbb{R}$ such that $U_i(s_i, (a_H)_{H \in F_i[s_i]}) = \sum_{H \in F_i[s_i]} u_{i,H}(a_H)$.

The term $u_{i,H}(a_H)$ is the "reward" of action a_H towards i's payoff. The total payoff from s_i is the sum of such rewards over all s_i-relevant information sets. A factorable game is not additively separable for i when the opponents' actions on $F_i[s_i]$ interact in some way to determine i's payoff following s_i. All the examples discussed in Section 3 are additively separable for the players ranked by compatibility.[17] While we provide our learning foundation

[15] 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 the 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 $g_i(\alpha_H) > 0$ for all $\alpha_H \in \Delta^\circ(A_H)$.

[16] As Fudenberg and Kreps (1993) point out, an agent who believes two opponents are randomizing independently may nevertheless have subjective correlation in her uncertainty about the randomizing probabilities of these opponents.

[17] Additive separability is trivially satisfied whenever $|F_i[s_i]| \leq 1$ for each s_i, so that is there is at most one s_i-relevant information set for each strategy s_i of i. So, every signaling game is additively separable for every sender type. It is also satisfied in the link-formation game in Section 4.2.2 even though here

for rational agents in any factorable game, our analysis of the Bayes upper confidence bound algorithm will restrict to such additively separable games.

5.2.1 Expected Discounted Utility and the Gittins Index

Consider a rational agent who maximizes discounted expected utility. In addition to the survival chance $0 \leq \gamma < 1$ between periods, the agent further discount future payoff according to her patience $0 \leq \delta < 1$, so her overall effective discount factor is $0 \leq \delta\gamma < 1$.

Given a belief about the distribution of play at each opponent information set H, we may calculate the Gittins index of each strategy $s_i \in \mathbb{S}_i$, corresponding to a super arm in in the combinatorial bandit problem. We write the solution to the rational agent's dynamic optimization problem as OPT_i, which involves playing the strategy s_i with the highest Gittins index after each history y_i.

The drawback of this learning rule is that the Gittins index is computationally intractable even in simple bandit problems. The combinatorial structure of our bandit problem makes computing the index even more complex, as it needs to consider the evolution of beliefs about each basic arm.

5.2.2 Bayesian Upper-Confidence Bound

The Bayesian upper confidence bound (Bayes-UCB) procedure was first proposed by Kaufmann, Cappé, and Garivier (2012) as a computationally tractable algorithm for dealing with the exploration-exploitation trade-off in bandit problems. In each period t, Bayes-UCB computes the $q(t)$-quantile of the posterior payoff distribution of each arm, then pulls the arm with the highest such quantile value.[18]

We restrict attention to games additively separable for i and adopt a variant of Bayes-UCB. Every $y_{i,H}$ subhistory of play on $H \in F_i[s_i]$ induces a posterior belief $g_i(\cdot|y_{i,H})$ over play on H, so $g_i(\cdot|y_{i,H})$ is an element of $\Delta(\Delta(A_H))$. By an abuse of notation, we use $u_{i,H}(g_i(\cdot|y_{i,H})) \in \Delta(\mathbb{R})$ to mean the distribution over contributions for play distributed according to $g_i(\cdot|y_{i,H})$. As a final bit of notation, when F is a distribution on $\mathbb{R}$, $Q(F;q)$ is the q-quantile of F.

Definition 14. Let prior g_i and quantile-choice function $q : \mathbb{N} \to [0,1]$ be given for i. The

$|F_i[\textbf{Active } i]| = 2$, as each agent computes her payoff by summing her linking costs/benefits with respect to each potential counterparty. Additive separability is also satisfied in the restaurant game in Section 4.2.1 for each customer i. $F_i[R\ i]$ contains two information sets, corresponding to the play of the Restaurant and the other customer. The play of the other customer additively contributes either 0 or -0.5 to i's payoff, depending on whether they choose R or not.

[18] Kaufmann, Cappé, and Garivier (2012) and Kaufmann (2018) show that the appropriate (but not problem-specific) specification of the function q makes it asymptotically optimal for various reward structures.

Bayes-UCB index for s_i after history y_i (relative to g_i and q) is

$$\sum_{H \in F_i[s_i]} Q(\ u_{i,H}(g_i(\cdot|y_{i,H}))\ ;\ q(\#(s_i|y_i))\ ),$$

where $\#(s_i|y_i)$ is the number of times s_i has been used in history y_i.

In words, our Bayes-UCB index computes the q-th quantile of $u_{i,H}(a_H)$ under i's belief about $-i$'s play on H, then sums these quantiles to return an index of the strategy s_i. The *Bayes-UCB policy* UCB_i prescribes choosing the strategy with the highest Bayes-UCB index after every history.

This procedure embodies a kind of wishful thinking for $q \geq 0.5$. The agent optimistically evaluates the payoff consequence of each s_i under the assessment that opponents will play a favorable response to s_i at each of the s_i-relevant information sets, where greater q corresponds to greater optimism in this evaluation procedure. Indeed, if q approaches 1 for every s_i, the Bayes-UCB procedure approaches picking the strategy with the highest potential payoff.

If $F_i[s_i]$ consists of only a single information set of for every s_i, then the procedure we define is the standard Bayes-UCB policy. In general, our procedure differs from the usual Bayes-UCB procedure, which would instead compute

$$Q\left(\sum_{H \in F_i[s_i]} u_{i,H}(g_i(\cdot|y_{i,H})); q(\#(s_i|y_i))\right).$$

Instead, our procedure computes the sum of the quantiles, which is easier than computing the quantile of the sum, a calculation that requires taking the convolution of the associated distributions.

This variant of the Bayesian UCB is analogous to variants of the non-Bayesian UCB algorithm[19] (see e.g. Gai, Krishnamachari, and Jain (2012) and Chen, Wang, and Yuan (2013)) that separately compute an index for each basic arm and choose the super arm maximizing sum of the basic arm indices.[20]

5.2.3 Index policies

The analysis that follows makes heavy use of the fact that the Gittins index and the Bayes-UCB are index policies in the following sense:

[19] The non-Bayesian UCB index of a basic arm is an "optimistic" estimate of its mean reward that combines its empirical mean in the past with a term inversely proportional to the number of times the basic arm has been pulled.

[20] Kveton, Wen, Ashkan, and Szepesvari (2015) have established tight $O(\sqrt{n \log n})$ regret bounds for this kind of algorithm across n periods.

Definition 15. When Γ is factorable for i, a learning rule $r_i : Y_i \to \mathbb{S}_i$ is an *index policy* if there exist 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 $r_i(y_i) \in \underset{s_i \in \mathbb{S}_i}{\arg\max} \{\iota_{s_i}(y_{i,s_i})\}$.

If an agent uses an index policy, we can think of her behavior in the following way. At each history, she computes an index for each strategy $s_i \in \mathbb{S}_i$ based on the subhistory of those periods where she chose s_i,and she then plays a strategy with the highest index with probability 1. [21]

5.3 Induced Responses Respect Player Compatibility

We now analyze how compatibility relations in the stage game translate into restrictions on experimentation frequencies. We aim to demonstrate that if $(s_i^* \mid i) \succsim (s_j^* \mid j)$, then i's induced response plays s_i^* more frequently than j's induced response plays s_j^*. There is little hope of proving a comparative result of this kind if i and j face completely unrelated learning problems. Instead, we will require that i and j use the same learning rule with the same parameters, start with the same prior belief[22] about $-ij$'s play, and face the same distribution of $-ij$'s play. These assumptions are natural when a common population of agents get randomly assigned into player roles, such as in a lab experiment.

Theorem 2 shows that when i and j use the same learning rule and face the same learning environment, we have $\phi_i(s_i^*; r_i, \sigma_{-i}) \geq \phi_j(s_j^*; r_i; \sigma_{-j})$. This provides a microfoundation for the compatibility-based cross-player restrictions on trembles in Definition 3. Throughout, we will fix a stage game Γ that is isomorphically factorable for i and j, with isomorphism $\varphi : \mathbb{S}_i \to \mathbb{S}_j$ between their strategies.

Definition 16. Regular independent priors for i and j are *equivalent* if for each $s_i \in \mathbb{S}_i$ and $H \in F_i[s_i] \cap F_j[\varphi(s_i)]$, $g_i(\alpha) = g_j(\alpha)$ for all $\alpha \in \Delta(A_H)$.

Theorem 2. *Suppose $(s_i^* \mid i) \succsim (s_j^* \mid j)$ with $\varphi(s_i^*) = s_j^*$. Consider two learning agents in the roles of i and j with equivalent independent regular priors.*[23] *For any common survival chance $0 \leq \gamma < 1$ and any mixed strategy profile σ, we have $\phi_i(s_i^*; r_i, \sigma_{-i}) \geq \phi_j(s_j^*; r_j, \sigma_{-j})$ under either of the following conditions:*

[21] 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.

[22] We believe that that our learning foundation for player-compatible trembles continues to hold even when i and j start with different priors under a stronger version of the compatibility condition that converges to the current one as the priors become closer together, but we, we have not been able to prove this.

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

- $r_i = OPT_i$, $r_j = OPT_j$, *and* i *and* j *have the same patience* $0 \le \delta < 1$.

- *The stage game is additively separable for* i *and* j, *at every* $H \in \mathcal{H}_{-ij}$ *the auxiliary functions* $u_{i,H}, u_{j,H}$ *rank* $\alpha \in \Delta(A_H)$ *in the same way,* $r_i = UCB_i$, $r_j = UCB_j$, i *and* j *have the same quantile-choice function* $q_i = q_j$.

This result provides learning foundations for player-compatible trembles in a number of games, including the restaurant game from Section 4.2.1 and the link-formation game from Section 4.2.2, where the additive separability and same-ranking assumptions are satisfied for players ranked by compatibility.

5.4 Proof Outline for Theorem 2

The proof of Theorem 2 follows two steps. In Proposition 8, we abstract away from particular models of experimentation and consider two general index policies r_i, r_j in a stage game that is isomorphically factorable for i and j. Policy r_i is *more compatible* with s_i^* than r_j is with s_j^* if, following i and j's respective histories y_i, y_j that contain the same observations about the play of third parties $-ij$, whenever s_j^* has the highest index under r_j, then no $s_i^{'} \neq s_i^*$ has the highest index under r_i. We prove that for any index policies r_i, r_j where r_i is more compatible with s_i^* than r_j is with s_j^*, we get $\phi_i(s_i^*; r_i, \sigma_{-i}) \ge \phi_j(s_j^*; r_j, \sigma_{-j})$ in any learning environment σ. In Corollaries B.1and B.2, we show that under the conditions of Theorem 2 that relate i and j's learning problems to each other (e.g. i and j have equivalent regular priors, same patience level, etc.), the compatibility relation $(s_i^* \mid i) \succsim (s_j^* \mid j)$ implies OPT_i is more compatible with s_i^* than OPT_j is with s_j^*, and that the same is true for UCB_i and UCB_j.

5.4.1 Comparative Compatibility for Index Policies

We begin by introducing a notion of equivalence between the histories of i and j. Since i could observe j's play and vice versa, this equivalence is only defined in terms of the actions of the $-ij$ third parties.

Definition 17. For Γ isomorphically factorable for i and j with $\varphi(s_i) = s_j$, i's subhistory y_{i,s_i} is *third-party equivalent* to j's subhistory y_{j,s_j}, written as $y_{i,s_i} \sim y_{j,s_j}$, if they contain the same sequence of observations about the actions of $-ij$.

Recall that, by Notation 1, we identify each subhistory y_{i,s_i} with a sequence in $\times_{H \in F_i[s_i]} A_H$ and each subhistory y_{j,s_j} with a sequence in $\times_{H \in F_j[s_j]} A_H$. By isomorphic factorability, $F_i[s_i] \cap \mathcal{H}_{-ij} = F_j[s_j] \cap \mathcal{H}_{-ij}$. Third-party 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 $-ij$'s actions that i encountered from experimenting with s_i are the same as those that j encountered from experimenting with s_j.

As an example, the following histories for the critic and the diner of the restaurant game are third-party equivalent for the strategy R. This is because the subhistories $y_{\text{Critic},R}$ and $y_{\text{Diner},R}$ 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).

y_{Critic} :

period	1	2	3	4	5
own strategy	R	Z	Z	Z	R
others' play	(L, Z)	$\varnothing$	$\varnothing$	$\varnothing$	(H, Z)

y_{Diner} :

period	1	2	3	4
own strategy	Z	R	Z	R
others' play	$\varnothing$	(L, R)	$\varnothing$	(H, Z)

Table 2: The two histories y_{Critic} (with length 5) and y_{Diner} (with length 4) have third-party equivalent subhistories for R. The row "others' play" show what the agent infers about others' play from her payoffs — recall that a customer choosing Z always gets the same payoff and so cannot infer anything about how others play.

We use third-party equivalent histories to define a comparison between two abstract index policies.

Definition 18. Suppose Γ is isomorphically factorable for i and j with $\varphi(s_i^*) = s_j^*$. For two index policies r_i and r_j, we have r_i is *more compatible* with s_i^* than than r_j is with s_j^* if for any histories y_i, y_j and strategy $s_i^{'} \in \mathbb{S}_i$, $s_i^{'} \neq s_i^*$ satisfying

1. $y_{i,s_i^*} \sim y_{j,s_j^*}$ and $y_{i,s_i^{'}} \sim y_{j,\varphi(s_i^{'})}$

2. s_j^* has weakly the highest index for j

$s_i^{'}$ does not have the weakly highest index for i.

Definition 18 is a property of the index policies r_i, r_j, and does not make reference to payoffs in the underlying stage game. The comparison applies to pairs of policies r_i, r_j such that whenever the subhistories of y_i for strategies s_i^* and $s_i^{'} \neq s_i^*$ are third-party equivalent to subhistories of y_j for s_j^* and $\varphi(s_i^{'})$, and s_j^* has the highest r_j-index at history y_j, then $s_i^{'}$ does not have the highest r_i-index under y_i.

We can now state the first intermediary result we need to establish Theorem 2, which is about the relative experimentation frequencies generated by a pair of index policies where the compatibility relation of Definition 18 applies.

Proposition 8. *Suppose Γ is isomorphically factorable for i and j with $\varphi(s_i^*) = s_j^*$, and that index policy r_i is more compatible with s_i^* than index policy 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 $\sigma \in \times_k \Delta(\mathbb{S}_k)$.*

The proof is similar to the coupling argument in the proof of Fudenberg and He (2018)'s Lemma 2, which only applies to the Gittins index in signaling games. Proposition 8 applies to any index policies satisfying the comparative compatibility condition from Definition 18. The proof uses this hypothesis to deduce a general conclusion about the induced responses of these agents in the learning problem, where the two agents typically do not have third-party equivalent histories in any given period.

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* $\mathfrak{A} \in (\times_{H\in\mathcal{H}} A_H)^\infty$. For each such path and learning rule r_i for player i, imagine running the rule against the data-generating process where the k-th time i plays s_i, i observes the action $a_{k,H} \in A_H$ at the information set $H \in F_i[s_i]$. Given a learning rule r_i, each $\mathfrak{A}$ induces a deterministic infinite history of i's strategies $y_i(\mathfrak{A}, r_i) \in (\mathbb{S}_i)^\infty$. We show that under the hypothesis that r_i is more compatible with s_i^* than r_j is with s_j^*, the weighted lifetime frequency of s_i^* in $y_i(\mathfrak{A}, r_i)$ is larger than that of s_j^* in $y_j(\mathfrak{A}, r_j)$ for every $\mathfrak{A}$, where play in different periods of the infinite histories $y_i(\mathfrak{A}, r_i), y_j(\mathfrak{A}, r_j)$ are weighted by the probabilities of surviving into these periods, just as in the definition of induced responses.

Lemma OA.2 in the Appendix shows that when i and j face i.i.d. draws of opponents' play from a fixed learning environment σ, the induced responses are the same as if they each faced a random response path $\mathfrak{A}$ drawn at birth according to the (infinite) product measure over $(\times_{H\in\mathcal{H}} A_H)^\infty$ whose marginal distribution on each copy of $\times_{H\in\mathcal{H}} A_H$ corresponds to σ.

5.4.2 OPT and UCB Satisfy Comparative Compatibility

The second step of our proof is carried out in Appendix B. There, Corollaries B.1 and B.2 show that when the assumptions of Theorem 2 hold and $(s_i^* \mid i) \succsim (s_j^* \mid j)$, both OPT and UCB are more compatible with s_i^* than with s_j^* provided the additional regularity conditions of Theorem 2 hold. This proves the theorem and provides two learning models that microfound PCE's tremble restrictions.[24] Since the compatibility relation is defined in the language of best responses against opponents' strategy profiles in the stage game, the key step in showing that OPT and UCB satisfy the comparative compatibility condition involves reformulating these indices as the expected utility of using each strategy against a certain opponent strategy profile. For the Gittins index, this profile is the "synthetic" opponent strategy profile constructed from the best stopping rule in the auxiliary optimal-stopping problem defining the index. This is similar to the construction of Fudenberg and He (2018), but in the more general setting of this paper the arguments become more subtle, as the induced synthetic strategy may be correlated (even though players have independent

[24]Other natural index rules that we do not analyze explicitly here also serve as microfoundations of our cross-player restrictions on trembles, provided they satisfy Proposition 8 whenever $(s_i^* \mid i) \succsim (s_j^* \mid j)$.

prior beliefs). For Bayes-UCB, under the assumptions of Theorem 2, the agent may rank opponents' mixed strategies on each $H \in F_i[s_i]$ from least favorable to most favorable. The Bayes-UCB index of s_i with quantile q is, roughly speaking, equivalent to the expected utility of s_i when opponents play mixed actions ranked at the q-th quantile in terms of $i's$payoff under i's current belief about opponents' play.

6 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 compatibility relation $(s_i \mid i) \succsim (s_j \mid j)$. As we have shown through examples, this distinguishes 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 the the device of restricted "trembles" can capture some of the implications of non-equilibrium learning. As we saw, PCE's cross-player restrictions arise endogenously in both the standard model of Bayesian agents maximizing their expected discounted lifetime utility, and the computationally tractable heuristics of Bayesian upper confidence bounds. We conjecture that the result that i is more likely to experiment with s_i than j with s_j when $(s_i \mid i) \succsim (s_j \mid j)$ applies in other natural models of learning or dynamic adjustment, such as those considered by Francetich and Kreps (2018), 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 $(s_i \mid i) \succsim (s_j \mid j)$ holds. Definition 1 supposes that player i thinks all mixed strategies of other players are possible, as it considers the set of all strictly 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 $\mathscr{A}_{-i}$ of $\Delta^\circ(\mathbb{S}_{-i})$. As we show in Fudenberg and He (2017), 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.

In this paper we have only provided learning foundations for factorable games, in which player i can only learn about the consequences of strategy s_i by playing it. In more general extensive-form games two complications arise. First, player i may have several actions that lead to the same information set of player j, which makes the optimal learning strategy more complicated. Second, player i may get information about how player j plays at some information sets thanks to an experiment by some other player k, so that player i has an incentive to free ride. We plan to deal with these complications in future work. One particular class of non-factorable games of interest is those with replicated strategies as in Appendix OA 2, where different copies of the same strategy give the same information about others' play. We conjecture that rational learning and heuristics like Bayes UCB can provide a foundation for the cross-player restrictions we impose there on the sums of tremble probabilities across different copies of strategies ranked by compatibility. Moreover, we conjecture that in games where actions have a natural ordering, learning rules based on the idea that nearby strategies induce similar responses can provide learning foundations for refinements in which players tremble more onto nearby actions, as in Simon (1987). More speculatively, the interpretation of trembles as arising from learning may provide learning-theoretic foundations for equilibrium refinements that restrict beliefs at off-path information sets in general extensive-form games, such as perfect Bayesian equilibrium (Fudenberg and Tirole, 1991; Watson, 2017), sequential equilibrium (Kreps and Wilson, 1982) and its extension to games with infinitely many actions (Simon and Stinchcombe, 1995; Myerson and Reny, 2018).

References

Agrawal, R. (1995): "Sample mean based index policies by o (log n) regret for the multi-armed bandit problem," *Advances in Applied Probability*, 27, 1054–1078.

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 (2017): "Learning and Self-confirming Long-Run Biases," *Working Paper*.

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

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

Chen, W., Y. Wang, and Y. Yuan (2013): "Combinatorial Multi-Armed Bandit: General Framework and Applications," in *Proceedings of the 30th International Conference on Machine Learning*, ed. by S. Dasgupta and D. McAllester, Atlanta, Georgia, USA: PMLR, vol. 28 of *Proceedings of Machine Learning Research*, 151–159.

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. M. Kreps (2018): "Choosing a Good Toolkit: Bayes-rule Based Heuristics," *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 (2017): "Learning and Equilibrium Refinements in Signalling Games," *Working Paper*.

——— (2018): "Learning and Type Compatibility in Signaling Games," *Econometrica*, 86, 1215–1255.

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

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

Fudenberg, D. 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. and D. K. Levine (1993): "Steady State Learning and Nash Equilibrium," *Econometrica*, 61, 547–573.

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

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

Gai, Y., B. Krishnamachari, and R. Jain (2012): "Combinatorial Network Optimization With Unknown Variables: Multi-Armed Bandits With Linear Rewards and Individual Observations," *IEEE/ACM Transactions on Networking*, 20, 1466–1478.

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.

Katehakis, M. N. and H. Robbins (1995): "Sequential choice from several populations," *Proceedings of the National Academy of Sciences of the United States of America*, 92, 8584.

Kaufmann, E. (2018): "On Bayesian index policies for sequential resource allocation," *The Annals of Statistics*, 46, 842–865.

Kaufmann, E., O. Cappé, and A. Garivier (2012): "On Bayesian upper confidence bounds for bandit problems," in *Artificial Intelligence and Statistics*, 592–600.

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.

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

Kveton, B., Z. Wen, A. Ashkan, and C. Szepesvari (2015): "Tight Regret Bounds for Stochastic Combinatorial Semi-Bandits," in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, ed. by G. Lebanon and S. V. N. Vishwanathan, San Diego, California, USA: PMLR, vol. 38 of *Proceedings of Machine Learning Research*, 535–543.

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

Milgrom, P. and J. Mollner (2017): "Extended Proper Equilibrium," *Working Paper*.

——— (2018): "Equilibrium Selection in Auctions and High Stakes Games," *Econometrica*, 86, 219–261.

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.

Myerson, R. B. and P. J. Reny (2018): "Perfect Conditional ε-Equilibria of Multi-Stage Games with Infinite Sets of Signals and Actions," *Working Paper*.

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.

Simon, L. K. (1987): "Local perfection," *Journal of Economic Theory*, 43, 134–156.

Simon, L. K. and M. B. Stinchcombe (1995): "Equilibrium refinement for infinite normal-form games," *Econometrica*, 1421–1443.

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

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

Watson, J. (2017): "A General, Practicable Definition of Perfect Bayesian Equilibrium," *Working Paper*.

Appendix

A Omitted Proofs from the Main Text

A.1 Proof of Proposition 1

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

Proof. Suppose there is some $\hat{\sigma}_{-k} \in \Delta^\circ(\mathbb{S}_{-k})$ such that that

$$u_k(s_k^*, \hat{\sigma}_{-k}) \geq \max_{s_k' \in \mathbb{S}_k} u_k(s_k', \hat{\sigma}_{-k}).$$

We show that s_i^* is strictly optimal for i versus any $\hat{\sigma}_{-i} \in \Delta^\circ(\mathbb{S}_{-i})$ with the property that $\hat{\sigma}_{-i}|_{\mathbb{S}_{-ik}} = \hat{\sigma}_{-k}|_{\mathbb{S}_{-ik}}$.

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

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

This profile $\hat{\hat{\sigma}}$ has the property that $\hat{\hat{\sigma}}|_{\mathbb{S}_{-jk}} = \hat{\sigma}_{-k}\ |_{\mathbb{S}_{-jk}}$, because $\hat{\sigma}_{-i}|_{\mathbb{S}_{-ik}} = \hat{\sigma}_{-k}|_{\mathbb{S}_{-ik}}$, $\hat{\hat{\sigma}}$ and $\hat{\sigma}_{-k}$ agree on the $-(ijk)$ marginal. Also, by construction, the conditional distribution of i's action given profile of $(-ijk)$'s actions is the same.

Let $\hat{\sigma}_{-j} = \hat{\hat{\sigma}}|_{\mathbb{S}_{-j}}$. By property just given, $\hat{\sigma}_{-j}|_{\mathbb{S}_{-jk}} = \hat{\sigma}_{-k}\ |_{\mathbb{S}_{-jk}}$. From the hypothesis that $(s_j^* \mid j) \succsim (s_k^* \mid k)$, we get j finds s_j^* strictly optimal against $\hat{\sigma}_{-j}$.

But at the same time, $\hat{\sigma}_{-j}|_{\mathbb{S}_{-ij}} = \hat{\sigma}_{-i}\ |_{\mathbb{S}_{-ij}}$ from the construction of $\hat{\sigma}_{-j}$ as a projection of $\hat{\hat{\sigma}}$, whose projection onto $-i$ gives back $\hat{\sigma}_{-i}$. From $(s_i^* \mid i) \succsim (s_j^* \mid j)$, and the conclusion that j finds s_j^* strictly optimal against $\hat{\sigma}_{-j}$ just obtained, we get i finds s_i^* strictly optimal against $\hat{\sigma}_{-i}$ as desired. $\square$

A.2 Proof of Proposition 2

Proposition 2: *If* $(s_i^* \mid i) \succsim (s_j^* \mid j)$, *then* $(s_j^* \mid j) \not\succsim (s_i^* \mid i)$.

Proof. Assume $(s_i^* \mid i) \succsim (s_j^* \mid j)$ and recall the maintained assumption that the game has no strictly dominated strategy. We show that these assumptions imply $(s_j^* \mid j) \not\succsim (s_i^* \mid i)$.

Partition the set $\Delta^\circ(\mathbb{S}_{-j})$ into three subsets, $\Pi^+ \cup \Pi^0 \cup \Pi^-$, with Π^+consisting of representing$\sigma_{-j} \in \Delta^\circ(\mathbb{S}_{-j})$ that make s_j^* strictly better than the best alternative pure strategy, $\Pi^-$the elements of $\Delta^\circ(\mathbb{S}_{-j})$ that make s_j^* indifferent to the best alternative, and $\Pi^-$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 $(\sigma_{-j}) \in \Pi^0$ such that

$$u_j(s_j^*, \sigma_{-j}) = \max_{s_j^{'} \in \mathbb{S}_j \setminus \{s_j^*\}} u_j(s_j^{'}, \sigma_{-j}).$$

By $(s_i^* \mid i) \succsim (s_j^* \mid j)$, we get

$$u_i(s_i^*, \hat{\sigma}_{-i}) > \max_{s_i^{'} \in \mathbb{S}_i \setminus \{s_i^*\}} u_i(s_i^{'}, \hat{\sigma}_{-i}),$$

for every $\hat{\sigma}_{-i} \in \Delta^\circ(\mathbb{S}_{-i})$ such that $\sigma_{-j}|_{\mathbb{S}_{-ij}} = \sigma_{-i}|_{\mathbb{S}_{-ij}}$, so we do not have $(s_j^* \mid j) \succsim (s_i^* \mid i)$.

Also, if both Π^+ and Π^- are non-empty, then Π^0 is non-empty. This is because both $\sigma_{-j} \mapsto u_j(s_j^*, \sigma_{-j})$ and $\sigma_{-j} \mapsto \max_{s_j^{'} \in \mathbb{S}_j \setminus \{s_j^*\}} u_j(s_j^{'}, \sigma_{-j})$ are continuous functions. If $u_j(s_j^*, \sigma_{-j}) - \max_{s_j^{'} \in \mathbb{S}_j \setminus \{s_j^*\}} u_j(s_j^{'}, \sigma_{-j}) > 0$ and also $u_j(s_j^*, \sigma_{-j}^{'}) - \max_{s_j^{'} \in \mathbb{S}_j \setminus \{s_j^*\}} u_j(s_j^{'}, \sigma_{-j}^{'}) < 0$, then some mixture between σ_{-j} and $\sigma_{-j}^{'}$ 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 $(s_j^* \mid j) \not\succsim (s_i^* \mid i)$.

If only Π^+ is non-empty, then s_j^* is strictly dominant for j. Together with $(s_i^* \mid i) \succsim (s_j^* \mid j)$, this would imply that s_i^* is strictly dominant for i, which would make any other strategy of i strictly dominated, contradiction.

Finally suppose that only Π^- is non-empty, so that for every $\sigma_{-j} \in \Delta^\circ(\mathbb{S}_{-j})$ there exists a strictly better pure response than s_j^* against σ_{-j}, then there exists a mixed strategy σ_j for j that strictly dominates s_j^* against all correlated play in $\Delta^\circ(\mathbb{S}_{-j})$. This shows s_j^* is strictly dominated for j provided $-j$ play a strictly mixed profile. Since the game does not contain strictly dominated strategies, there is a $\sigma_{-i} \in \Delta^\circ(\mathbb{S}_{-i})$ against which s_i^* is a weak best response,so the fact that s_j^* is not a strict best response against any $\sigma_{-j} \in \Delta^\circ(\mathbb{S}_{-j})$ means $(s_j^* \mid j) \not\succsim (s_i^* \mid i)$. □

A.3 Proof of Proposition 3

Proposition 3: *If the game only has two players, $i \neq j$, then $(s_i^* \mid i) \succsim (s_j^* \mid j)$ never holds, so every trembling-hand perfect equilibrium is a PCE.*

Proof. Let any $s_i^* \in \mathbb{S}_i$ and $s_j^* \in \mathbb{S}_j$ be given. Since the game contains no strictly dominated strategies,there exists some σ_i to which s_j^* is a weak best response. Then because t s_i^* cannot be strictly dominant (ore other strategies of i would be strictly dominated), there exists some σ_j where s_i^* is not a strict best response. Since σ_i and σ_j (trivially) match in terms of the play of $-ij$, this shows $(s_i^* \mid i) \not\succsim (s_j^* \mid j)$. □

A.4 Proof of Proposition 6

Proposition 6: *In a signaling game, every PCE σ^* is a Nash equilibrium satisfying the compatibility criterion, as defined in 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 do not best respond to beliefs in the set $P(s, \sigma^*)$ as defined in Fudenberg and He (2018). For any plan assigned positive probability under σ_2^*, by Proposition 4 we may find a sequence of strictly mixed signal profiles $\sigma_1^{(t)}$ of the sender, so that whenever $(s|\theta) \succsim (s|\theta')$ we have $\liminf_{t\to\infty} \sigma_1^{(t)}(s|\theta)/\sigma_1^{(t)}(s|\theta') \geq 1$. Write $q^{(t)}(\cdot|s)$ as the Bayesian posterior belief about sender's type after signal s under $\sigma_1^{(t)}$, which is well defined because each $\sigma_1^{(t)}$ is strictly mixed. Whenever $(s|\theta) \succsim (s|\theta')$, this sequence of posterior beliefs satisfies $\liminf_{t\to\infty} q^{(t)}(\theta|s)/q^{(t)}(\theta'|s) \geq \lambda(\theta)/\lambda(\theta')$, so if the receiver's plan best responds to every element in the sequence, it also best responds to an accumulation point $(q^\infty(\cdot|s))_{s\in S}$ with $q^\infty(\theta|s)/q^\infty(\theta'|s) \geq \lambda(\theta)/\lambda(\theta')$ whenever $(s|\theta) \succsim (s|\theta')$. Since the player compatibility definition used in this paper is slightly stronger 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. $\square$

A.5 Proof of Lemma 1

Lemma 1: *If σ° is an $\boldsymbol{\epsilon}$-PCE and $(s_i^* \mid i) \succsim (s_j^* \mid j)$, then*

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

Proof. Suppose $\boldsymbol{\epsilon}$ is player-compatible and let $\boldsymbol{\epsilon}$-equilibrium σ° be given. For$(s_i^* \mid i) \succsim (s_j^* \mid j)$, suppose $\sigma_j^\circ(s_j^*) = \boldsymbol{\epsilon}(s_j^* \mid j)$. Then we immediately have $\sigma_i^\circ(s_i^*) \geq \boldsymbol{\epsilon}(s_i^* \mid i) \geq \boldsymbol{\epsilon}(s_j^* \mid j) = \sigma_j^\circ(s_j^*)$, where the second inequality comes from $\boldsymbol{\epsilon}$ being player compatible. On the other hand, suppose $\sigma_j^\circ(s_j^*) > \boldsymbol{\epsilon}(s_j^* \mid j)$. 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 $(\sigma_{-ij}, \sigma_i^\circ)$, where both σ_{-ij} and σ_i° are strictly mixed. The definition of $(s_i^* \mid i) \succsim (s_j^* \mid j)$ then implies that s_i^* must be a strict best response for i against (the strictly mixed) $(\sigma_{-ij}, \sigma_j^\circ)$, which has the same marginal on $\mathbb{S}_{-ij}$ as $(\sigma_{-ij}, \sigma_i^\circ)$. In the $\boldsymbol{\epsilon}$-equilibrium, i must assign as much weight to s_i^* as possible, so that $\sigma_i^\circ(s_i^*) = 1 - \sum_{s_i' \neq s_i^*} \boldsymbol{\epsilon}(s_i'|i)$. Combining these two cases establishes the desired result. $\square$

A.6 Proof of Proposition 4

Proposition 4: *For any PCE σ^*, player k, and strategy $\bar{s}_k$ such that $\sigma^*_k(\bar{s}_k) > 0$, there exists a sequence of strictly mixed strategy profiles $\sigma^{(t)}_{-k} \to \sigma^*_{-k}$ such that*

(i) for every pair $i, j \neq k$ with $(s^*_i \mid i) \succsim (s^*_j \mid j)$,

$$\liminf_{t\to\infty} \frac{\sigma^{(t)}_i(s^*_i)}{\sigma^{(t)}_j(s^*_j)} \geq 1;$$

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

Proof. By Lemma 1, for every $\boldsymbol{\epsilon}^{(t)}$-PCE we get

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

This says

$$\inf_{t\geq T}\frac{\sigma^{(t)}_i(s^*_i)}{\sigma^{(t)}_j(s^*_j)} \geq 1 - \sup_{t\geq T}\sum_{s'_i\neq s^*_i}\boldsymbol{\epsilon}^{(t)}(s'_i|i).$$

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

$$\liminf_{t\to\infty}\frac{\sigma^{(t)}_i(s^*_i)}{\sigma^{(t)}_j(s^*_j)} = \lim_{T\to\infty}\left\{\inf_{t\geq T}\frac{\sigma^{(t)}_i(s^*_i)}{\sigma^{(t)}_j(s^*_j)}\right\} \geq 1.$$

This shows that if we fix a PCE $\sigma^*$and consider a sequence of player-compatible trembles $\boldsymbol{\epsilon}^{(t)}$ and $\boldsymbol{\epsilon}^{(t)}$-PCE $\sigma^{(t)} \to \sigma^*$, then each $\sigma^{(t)}_{-k}$ satisfies $\liminf_{t\to\infty}\sigma^{(t)}_i(s^*_i)/\sigma^{(t)}_j(s^*_j) \geq 1$ whenever $i, j \neq k$ and $(s^*_i \mid i) \succsim (s^*_j \mid j)$. Furthermore, from $\sigma^*_k(\bar{s}_k) > 0$ and $\sigma^{(t)}_k \to \sigma^*_k$, we know there is some $T_1 \in \mathbb{N}$ so that $\sigma^{(t)}_k(\bar{s}_k) > \sigma^*_k(\bar{s}_k)/2$ 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 $t \geq \max(T_1, T_2)$, $\sigma^{(t)}_k$ 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 $\sigma^{(t)}_{-k}$. Now the subsequence of opponent play $(\sigma^{(t)}_{-k})_{t\geq\max(T_1,T_2)}$ satisfies the requirement of this proposition. $\square$

A.7 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 action, 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. □

A.8 Proof of Proposition 5

Proposition 5: *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, strategic stability, and pairwise stability. Moreover the link-formation game is not a potential game.*

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.

Proof. 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 ϵ', 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 her 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 parametrization. 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 parametrization does not have a potential either. $\square$

A.9 Proof of Lemma 2

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 $(s_i, (a_H)_{H \in \mathcal{H}_{-i}}) = (s_i, a_{H^*}, (a_H)_{H \in \mathcal{H}_{-i} \setminus H^*})$. Find a different action of j on H^*, $a'_{H^*} \neq a_{H^*}$. Since H^* is off the path of play, both $(s_i, a_{H^*}, (a_H)_{H \in \mathcal{H}_{-i} \setminus H^*})$ and $(s_i, a'_{H^*}, (a_H)_{H \in \mathcal{H}_{-i} \setminus H^*})$ 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. $\square$

A.10 Proof of Lemma 3

Proof. First, there must be at least two different actions for j on H^*, else i's payoff would be trivially independent of H^*.

So, 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

$$U_i(s_i, a_{H^*}, (a_H)_{H \in \mathcal{H}_{-i} \setminus H^*}) \neq U_i(s_i, a'_{H^*}, (a_H)_{H \in \mathcal{H}_{-i} \setminus H^*}),$$

where $(a_H)_{H \in \mathcal{H}_{-i} \setminus H^*}$ 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, $(a_{H^*}, (a_H)_{H \in \mathcal{H}_{-i} \setminus H^*})$ and $(a'_{H^*}, (a_H)_{H \in \mathcal{H}_{-i} \setminus H^*})$ 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$\Pi_i[s_i]$. $\square$

A.11 Proof of Proposition 8

Proof. Let $0 \leq \delta, \gamma < 1$ and the learning environment σ be fixed. Consider the product distribution η on the space of response paths, $(\times_{H\in\mathcal{H}} A_H)^\infty$, whose marginal on each copy of $\times_{H\in\mathcal{H}} A_H$ is the action distribution of σ.

Lemma OA.2 in the Online Appendix shows that i's induced response against i.i.d. play drawn from σ_{-i} is the same as as playing against a response path drawn from η at the start of i's life. That is, for any player i, any rule r_i, and any strategy $s_i \in \mathbb{S}_i$, we have

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

where $y_i^t(\mathfrak{A}, r_i)$ refers to the t-th period strategy in the infinite history of strategies $y_i(\mathfrak{A}, r_i)$.

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

Denote the period where s_i^* appears in $y_i(\mathfrak{A}, r_i)$ for the k-th time as $T_i^{(k)}$, the period where s_j^* appears in $y_j(\mathfrak{A}, r_j)$ 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 $y_i(\mathfrak{A}, r_i)$ 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 $y_j(\mathfrak{A}, r_j)$ 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 $y_i(\mathfrak{A}, r_i)$. that leads to $s_i^{'}$ being played for the $(\#(\varphi(s_i^{'}), 1) + 1)$-th time, and find the subhistory y_j of $y_j(\mathfrak{A}, r_j)$ 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, $y_{i,s_i^{'}} \sim y_{j,s_j^{'}}$ since i has played $s_i^{'}$ for $\#(\varphi(s_i^{'}), 1)$ times and j has played $s_j^{'}$ for the same number of times, while the definition of response sequence implies they would have seen the same history of play on the common information sets of $-ij$, $F_i[s_i^{'}] \cap F_j[s_j^{'}]$. This satisfies the definition of third-party equivalence of histories.

Since $r_j(y_j) = s_j^*$ and r_j is an index rule, 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 $y_i(\mathfrak{A}, r_i)$. 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 we get $T_i^{(K)} \neq \infty$ 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 $y_j(\mathfrak{A}, r_j)$ 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 rule, 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. □

B Compatibility for Index Rules and the Compatibility Relation

In this section, we show that under the conditions of Theorem 2, the Gittins index and the UCB index satisfy the comparative compatibility condition from Definition 18. Omitted proofs from this section can be found in the Online Appendix.

B.1 The Gittins Index

Let survival chance $\gamma \in [0,1)$ and patience $\delta \in [0,1)$ be fixed. 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 s_i-relevant information sets. 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 $-i$ play on the s_i-relevant information sets the t-th time that i uses s_i — by factorability, only these actions and not actions elsewhere in the game tree determine i's payoff from playing s_i. 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 times, observing the first T elements $(a_H(t))_{t=1}^{T}$, then stopping.

Write $V(\tau; s_i, \nu_{s_i})$ for the value of the above auxiliary problem under the (not necessarily optimal) stopping time τ. 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 $\alpha(\nu_{s_i}, \tau) \in \Delta^\circ(\times_{H\in\mathcal{H}[s_i]} A_H)$, 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 B.1. 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 $\times_{\in F_i[s_i]} A_H \cup \{\varnothing\}$, 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 Y_i. The process over internal histories is i's prediction about what would happen in the auxiliary problem (which is an artificial device for computing the Gittins index) if he were to use τ.

Enumerate all possible profiles of moves at information sets $F_i[s_i]$ as $\times_{H\in F_i[s_i]} A_H = \{\boldsymbol{a}_{(-i)}^{(1)}, ..., \boldsymbol{a}_{(-i)}^{(K)}\}$, let $p_{t,k} := \mathbb{P}_{(-i)}[\tilde{\boldsymbol{a}}_{(-i),t} = \boldsymbol{a}_{(-i)}^{(k)}]$ for $1 \leq k \leq K$ be the probability under $\nu_{s_i}$of seeing the profile of actions $\boldsymbol{a}_{(-i)}^{(k)}$ 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 B.2. 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 $\boldsymbol{a}_{(-i)}^{(k)}$. Denote this profile by $\alpha(\nu_{s_i}, \tau)$.

Note that the synthetic correlated profile depends on the belief ν_{a_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 strictly 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 B.1.

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

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

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 the payoff of playing s_i against a independent randomizations by the opponents.[25]

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 B.2 will establish that the mimicking stopping time generates a synthetic correlated profile that matches the corresponding profile of τ_j on $F_i[s_i] \cap F_j[s_j]$.

The key issue in this construction is that τ_j maps j's internal histories to stopping decisions, which does not live in the same space as i's internal histories. In particular, τ_j makes use of i's play to decide whether to stop. For i 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 Γ isomorphically factorable for i and j with $\varphi(s_i) = s_j$, we may write $F_i[s_i] = F^C \cup F^j$ with $F^C \subseteq \mathcal{H}_{-ij}$ and $F^j \subseteq \mathcal{H}_j$. Similarly, we may write $F_j[s_j] = F^C \cup F^i$ with $F^i \subseteq \mathcal{H}_i$. (So, F^C is the **c**ommon information sets that are payoff-relevant for both s_i and s_j.) Whenever j plays s_j, he observes some $(\boldsymbol{a}_{(C)}, \boldsymbol{a}_{(i)}) \in (\times_{H \in F^C} A_H) \times (\times_{H \in F^i} A_H)$, where $\boldsymbol{a}_{(C)}$ is a profile of actions at information sets in F^C and $\boldsymbol{a}_{(i)}$ is a profile of actions at information sets in F^i. So, a pure-strategy stopping rule in the auxiliary problem defining j's Gittins index for s_j is a function $\tau_j : \cup_{t \geq 1} [(\times_{H \in F^C} A_H) \times (\times_{H \in F^i} A_H)]^t \to \{0, 1\}$ that maps finite histories of observations to stopping decisions, where "0" means continue and "1" means stop.

Definition B.3. Player i's *mimicking stopping rule for* τ_j draws $\alpha^i \in \times_{H \in F^i} \Delta(A_H)$ from j's belief ν_{s_j} on F^i, and then draws $(\boldsymbol{a}_{(i),\ell})_{\ell \geq 1}$ by independently generating $\boldsymbol{a}_{(i),\ell}$ from α^i each period. Conditional on $(\boldsymbol{a}_{(i),\ell})$, i stops according to the rule $(\tau_i | (\boldsymbol{a}_{(i),\ell}))((\boldsymbol{a}_{(C),\ell}, \boldsymbol{a}_{(j),\ell})_{\ell=1}^t) := \tau_j((\boldsymbol{a}_{(C),\ell}, \boldsymbol{a}_{(i),\ell})_{\ell=1}^t)$.[26]

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

[25] Forexample, suppose $F_i[s_i]$ consists of two information sets, one for each of two players $k_1 \neq k_2$, whose choose between **Heads** and **Tails**. Agent i's prior belief is that each of k_1, k_2 is either always playing **Heads** or always playing **Tails**, with each of the 4 possible combinations of strategies given 25% prior probability. Now consider the stopping rule where i stops if k_1 and k_2 play differently in the first period, but continues for 100 more periods if they play the same action in the first period. Then the procedure defined above generates a distribution over pairs of **Heads** and **Tails** that is mostly given by play in periods 2 through 100, which is either (**Heads**, **Heads**) or (**Tails**, **Tails**), each with 50% probability. Thus the stopping rule τ creates correlation in the observed play of the opponents.

[26] Here $(\boldsymbol{a}_{(-j),\ell})_{\ell=1}^t = (\boldsymbol{a}_{(C),\ell}, \boldsymbol{a}_{(i),\ell}))_{\ell=1}^t$. Note this is a valid (stochastic) stopping time, as the event $\{\tau_i \leq T\}$ is independent of any $\boldsymbol{a}_I(t)$ for $t > T$.

of i's play using this drawn behavior strategy and follows the pure-strategy stopping rule $\tau_i | (\boldsymbol{a}_{(i),\ell})_{\ell=1}^{\infty}$.

As in Definition B.1, the mimicking strategy and i's belief ν_{s_i} generates a stochastic process $(\tilde{\boldsymbol{a}}_{(j),t}, \tilde{\boldsymbol{a}}_{(C),t})_{t\geq 1}$ of internal histories for i (representing actions on $F_i[s_i]$ that i anticipates seeing when he plays s_i). It also induces a stochastic process $(\tilde{\boldsymbol{e}}_{(i),t}, \tilde{\boldsymbol{e}}_{(C),t})_{t\geq 1}$ of "external histories" defined in the following way:

Definition B.4. The stochastic process of *external histories* $(\tilde{\boldsymbol{e}}_{(i),t}, \tilde{\boldsymbol{e}}_{(C),t})_{t\geq 1}$ is defined from the process of internal histories $(\tilde{\boldsymbol{a}}_{(j),t}, \tilde{\boldsymbol{a}}_{(C),t})_{t\geq 1}$ that τ_i generates and given by: (i) if $\tau_i < t$, then $(\tilde{\boldsymbol{e}}_{(i),t}, \tilde{\boldsymbol{e}}_{(C),t}) = \varnothing$; (ii) otherwise, $\tilde{\boldsymbol{e}}_{(C),t} = \tilde{\boldsymbol{a}}_{(C),t}$, and $\tilde{\boldsymbol{e}}_{(i),t}$ is the t-th element of the infinite sequence $(\boldsymbol{a}_{(i),\ell})_{\ell=1}^{\infty}$ 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}.

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

To see the distribution of the external history in more detail, note that for any $\bar{\boldsymbol{a}}_{(C)} \in \times_{H\in F^C} A_H$ and $\bar{\boldsymbol{a}}_{(i)} \in \times_{H\in F^i} A_H$, the probability that the first period of i's external history is $(\bar{\boldsymbol{a}}_{(i)}, \bar{\boldsymbol{a}}_{(C)})$ is

$$\mathbb{P}_e[(\tilde{\boldsymbol{e}}_{(i),1}, \tilde{\boldsymbol{e}}_{(C),1}) = (\bar{\boldsymbol{a}}_{(i)}, \bar{\boldsymbol{a}}_{(C)})] = \int \alpha_i(\bar{\boldsymbol{a}}_{(i)}) d\nu_{s_j}(\alpha_i) \cdot \int \alpha_{-i}(\bar{\boldsymbol{a}}_{(C)}) d\nu_{s_i}(\alpha_{-i}),$$

where the multiplication comes from the fact that i's simulation of $(\boldsymbol{a}_\ell^i)_{\ell=1}^{\infty}$ happens before the auxiliary problem starts. (This expression depends on ν_{s_j} because the mimicking stopping rule simulates $(\boldsymbol{a}_\ell^i)_{\ell=1}^{\infty}$ based on j's belief.)

Also, for any $\bar{\boldsymbol{a}}'_{(C)} \in \times_{H\in F^C} A_H$ and $\bar{\boldsymbol{a}}'_{(i)} \in \times_{H\in F^i} A_H$,

$$\begin{aligned}
&\mathbb{P}_e[(\tilde{\boldsymbol{e}}_{(i),1}, \tilde{\boldsymbol{e}}_{(C),1}, \tilde{\boldsymbol{e}}_{(i),2}, \tilde{\boldsymbol{e}}_{(C),2}) = (\bar{\boldsymbol{a}}_{(i)}, \bar{\boldsymbol{a}}_{(C)}, \bar{\boldsymbol{a}}'_{(i)}, \bar{\boldsymbol{a}}'_{(C)})] \\
=&\mathbb{P}_e[\tau_i \geq 2 \mid (\tilde{\boldsymbol{e}}_{(i),1}, \tilde{\boldsymbol{e}}_{(C),1}) = (\bar{\boldsymbol{a}}_{(i)}, \bar{\boldsymbol{a}}_{(C)})] \cdot \int \alpha_i(\bar{\boldsymbol{a}}_{(i)}) \cdot \alpha_i(\bar{\boldsymbol{a}}'_{(i)}) d\nu_{s_j}(\alpha_i) \cdot \int \alpha_{-i}(\bar{\boldsymbol{a}}_{(C)}) \cdot \alpha_{-i}(\bar{\boldsymbol{a}}'_{(C)}) d\nu_{s_i}(\alpha_{-i}) \\
=&(1 - \tau_j(\bar{\boldsymbol{a}}_{(i)}, \bar{\boldsymbol{a}}_{(C)})) \cdot \int \alpha_i(\bar{\boldsymbol{a}}_{(i)}) \cdot \alpha_i(\bar{\boldsymbol{a}}'_{(i)}) d\nu_{s_j}(\alpha_i) \cdot \int \alpha_{-i}(\bar{\boldsymbol{a}}_{(C)}) \cdot \alpha_{-i}(\bar{\boldsymbol{a}}'_{(C)}) d\nu_{s_i}(\alpha_{-i}).
\end{aligned}$$

If $\tilde{\boldsymbol{e}}_{(i),1} = \bar{\boldsymbol{a}}_{(i)}$, then i must be using a pure-strategy stopping rule $(\tau_i | (\boldsymbol{a}_\ell^i))$ that stops after period 1 if and only if $\tau_j(\bar{\boldsymbol{a}}_{(i)}, \tilde{\boldsymbol{e}}_{(C),1}) = 1$. So the conditional probability term on the second line is either 0 or 1; for the second-period external history not to be $\varnothing$, we must have $\tau_j(\bar{\boldsymbol{a}}_{(i)}, \bar{\boldsymbol{a}}_{(C)}) = 0$. In the final line, the multiplication in the first integrand reflects the fact that elements of the sequence $(\boldsymbol{a}^i_\ell)$ are independently generated from the simulated $\alpha_i \sim \nu_{s_j}$, and the multiplication in the second integrand comes from i's knowledge that plays on the information sets F^C are independently generated across periods from some fixed α_{-i}.

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 that are both s_i-relevant and s_j-relevant. This is formalized in the next lemma.

Lemma B.2. *For Γ isomorphically factorable for i and j with $\varphi(s_i) = s_j$, 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 B.1. *Suppose Γ isomorphically factorable for i and j with $\varphi(s_i^*) = s_j^*$, $\varphi(s_i^{'}) = s_j^{'}$, where $s_i^* \neq s_i^{'}$ and $(s_i^* \mid i) \succsim (s_j^* \mid j)$. 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 profile of strictly mixed correlated actions at information sets $\cup_{s_j\in\mathbb{S}_j} F_j[s_j] \subseteq \mathcal{H}_{-j}$, namely a collection of strictly mixed correlated profiles $(\alpha_{F_j[s_j]})_{s_j\in\mathbb{S}_j}$ where $\alpha_{F_j[s_j]} \in \Delta^\circ(\times_{H\in F_j[s_j]}A_H)$. For each $s_j \neq s_j^{'}$ the profile $\alpha_{F_j[s_j]}$ is the synthetic correlated profile $\alpha(\nu_{s_j}, \tau_{s_j}^*)$ as given by Definition B.2, 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 $\alpha_{F_j[s_j^{'}]}$ is instead the synthetic correlated profile associated with the mimicking stopping rule for $\tau_{s_i^{'}}^*$, i.e. agent i's pure-strategy optimal stopping time in i's auxiliary problem for $s_i^{'}$.

Next, define a profile of strictly mixed correlated actions at information sets $\cup_{s_i\in\mathbb{S}_i} F_i[s_i] \subseteq \mathcal{H}_{-i}$ for i's opponents. For each $s_i \notin \{s_i^*, s_i^{'}\}$, just use the marginal distribution of $\alpha_{F_j[\varphi(s_i)]}$ constructed before on $F_i[s_i] \cap \mathcal{H}_{-ij}$, then arbitrarily specify play at j's information sets contained in $F_i[s_i]$, 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 $\alpha_{F_i[s_i^*]}$ is the synthetic correlated profile associated with the mimicking stopping rule for $\tau_{s_j^*}^*$.

From Lemma B.2, these two profiles of correlated actions agree when marginalized to information sets of $-ij$. Therefore, they can be completed into strictly mixed correlated strategies, σ_{-i} and σ_{-j} respectively, such that $\sigma_{-i}|_{\mathbb{S}_{-ij}} = \sigma_{-j}|_{\mathbb{S}_{-ij}}$. For each $s_j \neq s_j^{'}$, the Gittins index of s_j for j is $U_j(s_j, \sigma_{-j})$. Also, since $\alpha_{F_j[s_j^{'}]}$ is the mixed profile associated with the suboptimal mimicking stopping time, $U_j(s_j^{'}, \sigma_{-j})$ 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, we get

$$U_j(s_j^*, \sigma_{-j}) \geq \max_{s_j \neq s_j^*} U_j(s_j, \sigma_{-j}).$$

By the definition of $(s_i^* \mid i) \succsim (s_j^* \mid j)$ we must also have

$$U_i(s_i^*, \sigma_{-i}) > \max_{s_i \neq s_i^*} U_i(s_i, \sigma_{-i}),$$

so in particular $U_i(s_i^*, \sigma_{-i}) > U_i(s_i^{'} \sigma_{-i})$. But $U_i(s_i^*, \sigma_{-i})$ is no larger than the Gittins index of s_i^*, for $\alpha_{F_i[s_i^*]}$ is the synthetic strategy associated with a suboptimal mimicking stopping time. As $U_i(s_i^{'}, \sigma_{-i})$ is equal to the Gittins index of $s_i^{'}$, this shows $s_i^{'}$ cannot have even weakly the highest Gittins index at history y_i, for s_i^* already has a strictly higher Gittins index than $s_i^{'}$ does. $\square$

The following corollary of Proposition B.1, combined with Proposition 8, establishes the first statement of Theorem 2.

Corollary B.1. *When $(s_i^* \mid i) \succsim (s_j^* \mid j)$, i and j have the same patience δ, survival chance γ, and equivalent independent regular priors, OPT_i is more compatible with s_i^* OPT_j is with s_j^*.*

Proof. Equivalent regular priors require that priors are independent and that i and j share the same prior beliefs over play on $F^* := F_i[s_i^*] \cap F_j[s_j^*]$ and over play on $F^{'} := F_i[s_i^{'}] \cap F_j[s_j^{'}]$. Thus after histories y_i, y_j such that $y_{i,s_i^*} \sim y_{j,s_j^*}$ and $y_{i,s_i^{'}} \sim y_{j,s_j^{'}}$ $\nu_{s_i^*}|_{F^*} = \nu_{s_j^*}|_{F^*}$ and $\nu_{s_i^{'}}|_{F^{'}} = \nu_{s_j^{'}}|_{F^{'}}$, so the hypotheses of Proposition B.1 are satisfied. $\square$

B.2 Bayes-UCB

We start with a lemma that shows the Bayes-UCB index for a strategy s_i is equal to i's payoff from playing s_i against a certain profile of mixed actions on $F_i[s_i]$, where this profile depends on i's belief about actions on $F_i[s_i]$, the quantile q, and how $v_{s_i,H}$ ranks mixed actions in $\Delta(A_H)$ for each $H \in F_i[s_i]$.

Lemma B.3. *Let n_{s_i} be the number of times i has played s_i in history y_i and let $q_{s_i} = q(n_{s_i}) \in (0, 1)$. Then the Bayes-UCB index for s_i and given quantile-choice function q after* *history y_i is equal to $U_i(s_i, (\bar{\alpha}_H)_{H\in F_i[s_i]})$ for some profile of mixed actions where $\bar{\alpha}_H \in \Delta^\circ(A_H)$ for each H. Furthermore, $\bar{\alpha}_H$ only depends on q_{s_i}, $g_i(\cdot|y_{i,H})$ i's posterior belief about play on H, and how $u_{s_i,H}$ ranks mixed strategies in $\Delta(A_H)$.*

Proof. For each $H \in F_i[s_i]$, the random variable $\tilde{u}_{s_i,H}(y_{i,H})$ only depends on $y_{i,H}$ through the posterior $g_i(\cdot|y_{i,H})$. Furthermore, $Q(\tilde{u}_{s_i,H}(y_{i,H}); q_{s_i})$ is strictly between the highest and lowest possible values of $u_{s_i,H}(\cdot)$, each of which can be attained by some pure action on A_H, so there is a strictly mixed $\bar{\alpha}_H \in \Delta^\circ(A_H)$ so that $Q(\tilde{u}_{s_i,H}(y_{i,H}); q_{s_i}) = u_{s_i,H}(\bar{\alpha}_H)$. Moreover, if $u_{s_i,H}$ and $u^{'}_{s_i,H}$ rank mixed strategies on $\Delta(A_H)$ in the same way, there are $a \in \mathbb{R}$ and $b > 0$ so that $u^{'}_{s_i,H} = a + bu_{s_i,H}$. Then $Q(\tilde{u}^{'}_{s_i,H}(y_{i,H}); q_{s_i}) = a + bQ(\tilde{u}_{s_i,H}(y_{i,H}); q_{s_i})$, so $\bar{\alpha}_H$ still works for $u^{'}_{s_i,H}$. $\square$

The second statement of Theorem 2 follows as a corollary.

Corollary B.2. *If $(s_i^* \mid i) \succsim (s_j^* \mid j)$, and the hypotheses of Theorem 2 are satisfied, then UCB_i is more compatible with s_i^* than UCB_j is with s_j^*.*

The idea is that when i and j have matching beliefs, by Lemma B.3 we may calculate their Bayes-UCB indices for different strategies as their myopic expected payoff of using these strategies against some common opponents' play. This is similar to the argument for the Gittins index, up to replacing the synthetic correlated profile with the mixed actions defined in Lemma B.3. Applying the definition of compatibility, we can deduce that when $(s_i^* \mid i) \succsim (s_j^* \mid j)$ and $\varphi(s_i^*) = s_j^*$, if s_j^* has the highest Bayes-UCB index for j then s_i^* must have the highest Bayes-UCB index for i.

Online Appendix

OA 1 Proofs Omitted from the Appendix

OA 1.1 Proof of Lemma B.2

Lemma B.2: *For Γ isomorphically factorable for i and j with $\varphi(s_i) = s_j$, 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.*

Proof. Let $(\tilde{\boldsymbol{a}}_{(i),t}, \tilde{\boldsymbol{a}}_{(C),t})_{t\geq 1}$ and $(\tilde{\boldsymbol{e}}_{(i),t}, \tilde{\boldsymbol{e}}_{(C),t})_{t\geq 1}$ 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 F^j as $\times_{H\in F^j} A_H = \{\boldsymbol{a}_{(j)}^{(1)}, ..., \boldsymbol{a}_{(j)}^{(K_j)}\}$, and possible profiles of actions on F^i as $\times_{H\in F^i} A_H = \{\boldsymbol{a}_{(i)}^{(1)}, ..., \boldsymbol{a}_{(i)}^{(K_i)}\}$.

Write $p_{t,(k_j,k_C)} := \mathbb{P}_{-i}[(\tilde{\boldsymbol{a}}_{(i),t}, \tilde{\boldsymbol{a}}_{(C),t}) = (\boldsymbol{a}_{(j)}^{(k_j)}, \boldsymbol{a}_{(C)}^{(k_C)})]$ for $k_j \in \{1, ..., K_j\}$ and $k_C \in \{1, ..., K_C\}$. Also write $q_{t,(k_i,k_C)} := \mathbb{P}_e[(\tilde{\boldsymbol{e}}_{(i),t}, \tilde{\boldsymbol{e}}_{(C),t}) = (\boldsymbol{a}_{(i)}^{(k_i)}, \boldsymbol{a}_{(C)}^{(k_C)})]$ for $k_i \in \{1, ..., K_i\}$ and $k_C \in \{1, ..., K_C\}$. 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}. This is because of the hypothesis $\nu_{s_i}|_{F^C} = \nu_{s_j}|_{F^C}$, together with fact that i simulates the data-generating process on F^i by drawing a mixed action α^i according to j's belief $\nu_{s_j}|_{F^i}$. This means for every $k_i \in \{1, ..., K_i\}$ and every $k_C \in \{1, ..., K_C\}$,

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

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

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

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

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

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

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

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

Since the choice of $\boldsymbol{a}_{(C)}^{(\bar{k}_C)} \in \times_{I\in F^C} A_I$ 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. □

OA 1.2 Proof of Corollary B.2

Corollary B.2: *The Bayes-UCB rule $r_{i,UCB}$ and $r_{j,UCB}$ satisfy the hypotheses of Proposition 8 when $(s_i^* \mid i) \succsim (s_j^* \mid j)$, provided the hypotheses of Theorem 2 are satisfied.*

Proof. Consider histories y_i, y_j with $y_{i,s_i^*} \sim y_{j,s_j^*}$ and $y_{i,s_i^{'}} \sim y_{j,s_j^{'}}$. By Lemma B.3, there exist $\bar{\alpha}_H^{-i} \in \Delta^\circ(A_H)$ for every $H \in \cup_{s_i\in\mathbb{S}_i} F_i[s_i]$ and $\bar{\alpha}_H^{-j} \in \Delta^\circ(A_H)$ for every $H \in \cup_{s_i\in\mathbb{S}_i} F_j[s_j]$ so that $\iota_{i,s_i}(y_i) = U_i(s_i,(\alpha_H^{-i})_{H\in F_i[s_i]})$ and $\iota_{j,s_j}(y_j) = U_j(s_j,(\alpha_H^{-j})_{H\in F_j[s_j]})$ for all s_i, s_j, where $\iota_{i,s_i}(y_i)$ is the Bayes-UCB index for s_i after history y_i and $\iota_{j,s_j}(y_j)$ is the Bayes-UCB index for s_j after history y_j.

Because $y_{i,s_i^*} \sim y_{j,s_j^*}$ and $y_{i,s_i^{'}} \sim y_{j,s_j^{'}}$, y_i contains the same number of s_i^* experiments as y_j contains s_j^*, and y_i contains the same number of $s_i^{'}$ experiments as y_j contains $s_j^{'}$. Also by third-party equivalence and the fact that i and j start with the same beliefs on common relevant information sets, they have the same posterior beliefs $g_i(\cdot|y_{i,I})$, $g_j(\cdot|y_{j,I})$ for any $H \in F_i[s_i^*] \cap F_j[s_j^*]$ and $H \in F_i[s_i^{'}] \cap F_j[s_j^{'}]$. Finally, the hypotheses of Theorem 2 say that on any $H \in F_i[s_i^*] \cap F_j[s_j^*]$, $u_{s_i^*,H}$ and $u_{s_j^*,H}$ have the same ranking of mixed actions, while on any $H \in F_i[s_i^{'}] \cap F_j[s_j^{'}]$, $u_{s_i^{'},H}$ and $u_{s_j^{'},H}$ have the same ranking of mixed actions. So, by Lemma B.3, we may take $\bar{\alpha}_H^{-i} = \bar{\alpha}_H^{-j}$ for all $H \in F_i[s_i^*] \cap F_j[s_j^*]$ and $H \in F_i[s_i^{'}] \cap F_j[s_j^{'}]$.

Find some $\sigma_{-j} = (\sigma_{-ij},\sigma_i) \in \times_{k\neq j}\Delta^\circ(\mathbb{S}_k)$ so that σ_{-j} generates the random actions $(\bar{\alpha}_H^{-j})$ on every $H \in \cup_{s_j\in\mathbb{S}_j} F_j[s_j]$. Then we have $\iota_{j,s_j}(y_j) = U_j(s_j,\sigma_{-j})$ for every $s_j \in \mathbb{S}_j$. The fact that s_j^* has weakly the highest index means s_j^* is weakly optimal against σ_{-j}. Now take $\sigma_{-i} = (\sigma_{-ij},\sigma_j)$ where $\sigma_j \in \Delta^\circ(\mathbb{S}_j)$ is such that it generates the random actions

$(\bar{\alpha}_H^{-i})$ on $F_i[s_i^*] \cap \mathcal{H}_j$ and $F_i[s_i^{'}] \cap \mathcal{H}_j$. But since $\bar{\alpha}_H^{-i} = \bar{\alpha}_H^{-j}$ for all $H \in F_i[s_i^*] \cap F_j[s_j^*]$ and $H \in F_i[s_i^{'}] \cap F_j[s_j^{'}]$, σ_{-i} generates the random actions $(\bar{\alpha}_H^{-i})$ on all of $F_i[s_i^*]$ and $F_i[s_i^{'}]$, meaning $\iota_{i,s_i^*}(y_i) = U_i(s_i^*, \sigma_{-i})$ and $\iota_{i,s_i^{'}}(y_i) = U_i(s_i^{'}, \sigma_{-i})$. The definition of compatibility implies $U_i(s_i^*, \sigma_{-i}) > U_i(s_i^{'}, \sigma_{-i})$, so $\iota_{i,s_i^*}(y_i) > \iota_{i,s_i^{'}}(y_i)$. This shows $s_i^{'}$ does not have weakly the highest Bayes-UCB index, since s_i^* has a strictly higher one. $\square$

OA 2 Replication Invariance of PCE

Fix a *base game* where each player i has finite strategy set $\mathbb{S}_i$ and utility function $U_i : \mathbb{S} \to \mathbb{R}$. An *extended game with duplicates* is a game where i's strategy set $\bar{\mathbb{S}}_i$ is a finite subset of $\mathbb{S}_i \times \mathbb{N}$. A generic extended-game strategy $(s_i, n_i) \in \bar{\mathbb{S}}_i$ for i can be thought of as a copy of her strategy $s_i \in \mathbb{S}_i$ in the base game. We require that for every $s_i \in \mathbb{S}_i$, there exists some $n_i \in \mathbb{N}$ so that $(s_i, n_i) \in \bar{\mathbb{S}}_i$. So, i has a non-zero but finite number of copies of every strategy she had in the base game, and could have different numbers of copies of different strategies.

The payoff functions $(\bar{U}_i)_{i \in \mathbb{I}}$ in the extended game formalize the idea that duplicate strategies lead to the same payoffs as the original strategies in the base game. For each i and for a strategy profile $(s, n) \in \bar{\mathbb{S}}$ in the extended game,

$$\bar{U}_i((s_k, n_k)_{k \in \mathbb{I}}) = U_i((s_k)_{k \in \mathbb{I}}).$$

A *tremble profile of the extended game* $\bar{\boldsymbol{\epsilon}}$ assigns a positive number $\bar{\boldsymbol{\epsilon}}((s_i, n_i) \mid i) > 0$ to every player i and 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 OA.5. Tremble profile $\bar{\boldsymbol{\epsilon}}$ is *player compatible in the extended game* if $\sum_{n_i} \bar{\boldsymbol{\epsilon}}((s_i^*, n_i) \mid i) \geq \sum_{n_j} \bar{\boldsymbol{\epsilon}}((s_j^*, n_j) \mid j)$ for all i, j, s_i^*, s_j^* such that $(s_i^* \mid i) \succsim (s_j^* \mid j)$. 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 $\mathscr{C}(\bar{\boldsymbol{\epsilon}})(s_i \mid i) := \sum_{n_i} \bar{\boldsymbol{\epsilon}}((s_i, n_i) \mid i)$.

Analogously, 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 $\mathscr{E}(\boldsymbol{\epsilon})((s_i, n_i) \mid i) := \boldsymbol{\epsilon}(s_i \mid i)/N(s_i)$.

Lemma OA.1. *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.

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, $\mathscr{C}(\bar{\sigma})_i(s_i^{'}) > \mathscr{C}(\bar{\boldsymbol{\epsilon}})(s_i^{'} \mid i)$ means $\sum_{n_i} \bar{\sigma}_i(s_i^{'}, n_i) > \sum_{n_i} \bar{\boldsymbol{\epsilon}}((s_i^{'}, n_i) \mid i)$. So for at least one $n_i^{'}$, $\bar{\sigma}_i(s_i^{'}, n_i^{'}) > \bar{\boldsymbol{\epsilon}}((s_i^{'}, n_i^{'}) \mid i)$, 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 $\mathscr{E}(\sigma)_i(s_i^{'}, n_i^{'}) > \mathscr{E}(\boldsymbol{\epsilon})((s_i^{'}, n_i^{'}) \mid i)$, then also $\mathscr{E}(\sigma)_i(s_i^{'}, n_i) > \mathscr{E}(\boldsymbol{\epsilon})((s_i^{'}, n_i) \mid i)$ for every n_i such that $(s_i^{'}, n_i) \in \bar{\mathbb{S}}_i$. Therefore, we also have $\sigma_i(s_i^{'}) > \boldsymbol{\epsilon}(s_i^{'} \mid i)$, so σ is not an $\boldsymbol{\epsilon}$-equilibrium in the base game as desired. □

PCE is defined as usual in the extended game.

Definition OA.6. 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 OA.2. *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.

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 OA.1), 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. $\square$

In the next section, we will provide a learning foundation for the cross-player restriction on the sum of trembles, $\sum_{n_i} \bar{\boldsymbol{\epsilon}}((s_i^*, n_i) \mid i) \geq \sum_{n_j} \bar{\boldsymbol{\epsilon}}((s_j^*, n_j) \mid j)$ when $(s_i^* \mid i) \succsim (s_j^* \mid j)$ in the base game.

OA 3 Additional Results and Examples

OA 3.1 A PCE that is Not a Uniform THPE

There are 4 players, each with two actions, L and R. The following matrices give 1 and 2's payoffs if 3 plays L and R, respectively.

	L	R
L	$1,1$	$1,0$
R	$0,1$	$0,0$

	L	R
L	$1,1$	$1,1+\delta$
R	$10,1$	$10,+\delta$

Here $\delta > 0$ is a small positive number. If 1 and 2 are sufficiently certain that 3 will play R, then they want to play R. Otherwise, they want to play L. For δ small enough, we have $(R \mid 1) \succsim (R \mid 2)$.

Player 4 gets 0 from playing L. If player 4 chooses R, he gets a positive payoff from matching player 1, a negative payoff from matching player 2. So, $u_4(a_1, a_2, a_3, L) = 0$, $u_4(a_1, a_2, a_3, R) = (1) \cdot \mathbf{1}(a_1 = R) + (-2) \cdot \mathbf{1}(a_2 = R)$. Finally, player 3 wants to mismatch player 4's play, so $u_1(a_1, a_2, a_4, a_4)$ is 1 if $a_3 \neq a_4$, -1 otherwise.

The profile (L, L, L, R) is a PCE but not a uniform THPE. Consider a tremble profile where all actions except 1's R has tremble probability $\epsilon > 0$, but 1's R has probability 10ϵ. For small ϵ, it is a constrained equilibrium for everyone to place only the required trembles on the strategies not in (L, L, L, R). In particular, 4 wants to play R because the expected payoff is $1 \cdot (10\epsilon) + (-2) \cdot \epsilon > 0$.

Yet, in any constrained equilibrium with uniform trembles where 3 is placing high enough probability on L, 1 and 2 each puts probability ϵ on R, which leads to 4 playing L since $1 \cdot \epsilon + (-2) \cdot \epsilon < 0$. But this means it cannot be optimal for 3 to play L.

OA 3.2 An ϵ-equilibrium that Violates the Conclusion of Proposition 4

Lemma 1 gives a condition satisfied by all $\boldsymbol{\epsilon}$-equilibria where $\boldsymbol{\epsilon}$ satisfies player compatibility. The following example shows that when $\boldsymbol{\epsilon}$ violates player compatibility, the conclusion of Lemma 1 need not hold.

Example 3. There are 3 players: player 1 chooses a row, player 2 chooses a column, and player 3 chooses a matrix.

Matrix A	Left	Right
Top	1,1,0	1,4,0
Bottom	3,1,0	3,4,0

Matrix B	Left	Right
Top	3,3,1	3,2,1
Bottom	1,3,1	1,2,1

When player 3 chooses **Matrix A**, player 1 gets 1 from **Top** and 3 from **Bottom**. When player 3 chooses **Matrix B**, player 1 gets 3 from **Top** and 1 from **Bottom**. Player 2 always gets the same utility from **Left** as player 1 does from **Top**, but player 2 gets 1 more utility from **Right** than player 1 gets from **Bottom**. Therefore, (**Right** | P2) $\succsim$ (**Bottom** | P1). Suppose $\boldsymbol{\epsilon}$(**Bottom** | P1)=$2h$ and all other minimum probabilities are h for any $0 < h < \frac{1}{2}$. In any $\boldsymbol{\epsilon}-$equilibrium σ°, σ_3°(**Matrix B**) $> \frac{1}{2}$, so player 1 puts the minimum probability $2h$ on **Bottom**, and player 2 puts the minimum probability h on **Right**. This violates the conclusion of lemma 1. ♦

OA 3.3 Response Paths

The next result shows that playing against opponent strategies drawn i.i.d. from σ_{-i} each period generates the same experimentation frequencies as playing against a response paths drawn according to a certain distribution η at the start of the learner's life, then held fixed forever. The η is the same for all agents and does not depend on their (possibly stochastic) learning rules. In the main text, this result is used to couple together the learning problems of two agents $i \neq j$ and compare their experimentation frequencies.

Lemma OA.2. *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 rule $r_i : Y_i \to \Delta(\mathbb{S}_i)$, and any strategy $s_i \in \mathbb{S}_i$, we have*

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

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

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 rule r_i, and every t.

Think of each response path $\mathfrak{A}$ as a two-dimensional array, $\mathfrak{A} = (a_{t,H})_{t\in\mathbb{N},H\in\mathcal{H}}$. For non-negative integers $(N_H)_{H\in\mathcal{H}}$, each profile of sequences of actions $((a_{n_H,H})_{n_H=1}^{N_H})_{H\in\mathcal{H}}$ where $a_{n_H,H} \in A_H$ defines a "cylinder set" of response paths with the form:

$$\{\mathfrak{A} : a_{t,H} = a_{n_H,H} \text{ for each } H \in \mathcal{H}, 1 \leq n_H \leq N_H\}.$$

That is, the cylinder set consists of those response paths whose first N_H elements for information set H match a given sequence, $(a_{n_H,H})_{n_H=1}^{N_H}$. (If $N_H = 0$, then there is no restriction on $a_{t,H}$ for any t.) We specify the distribution η by specifying the probability it assigns to these cylinder sets:

$$\eta\left\{((a_{n_H,H})_{n_H=1}^{N_H})_{H\in\mathcal{H}}\right\} = \prod_{H\in\mathcal{H}}\prod_{n_H=1}^{N_H}\sigma(s : s(H) = a_{n_H,H}),$$

where we have abused notation to write $((a_{n_H,H})_{n_H=1}^{N_H})_{H\in\mathcal{H}}$ for the cylinder set satisfying this profile of sequences, and we have used the convention that the empty product is defined to be 1. Recall that a strategy profile s in the extensive-form game specifies an action $s(H) \in A_H$ for every information set H in the game tree. The probability that η assigns to the cylinder set involves multiplying the probabilities that the given mixed strategy σ leads to such a pure-strategy profile s so that $a_{n_H,H}$ is to be played at information set H, across all such $a_{n_H,H}$ restrictions defining the cylinder set.

We establish the claim by induction on t for period-t history. In the base case of $t = 1$, we show playing against a response path drawn according to η and playing against a pure strategy[27] drawn from $\sigma_{-i} \in \times_{k\neq i}\Delta(\mathbb{S}_k)$ generate the same period-1 history. Fixing a learning rule $r_i : Y_i \to \Delta(\mathbb{S}_i)$ of i, the probability of i having the period-1 history $(s_i^{(1)}, (a_H^{(1)})_{H\in F_i[s_i^{(1)}]}) \in Y_i[1]$ in the random-matching model is $r_i(\emptyset)(s_i^{(1)}) \cdot \sigma(s : s(I) = a_H^{(1)} \text{ for all } H \in F_i[s_i^{(1)}])$. That is, i's rule must play $s_i^{(1)}$ in the first period of i's life, which happens with probability $r_i(\emptyset)(s_i^{(1)})$. Then, i must encounter such a pure strategy that generates the required profile of moves $(a_H^{(1)})_{H\in F_i[s_i^{(1)}]}$ on the $s_i^{(1)}$-relevant information sets, which has probability $\sigma(s : s(H) = a_H^{(1)} \text{ for all } H \in F_i[s_i^{(1)}])$. The probability of this happening against a response path drawn

[27] 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.

from η is

$$\begin{aligned}
&r_i(\emptyset)(s_i^{(1)}) \cdot \eta(\mathfrak{A} : a_{1,H} = a_H^{(1)} \text{ for all } H \in F_i[s_i^{(1)}]) \\
=&r_i(\emptyset)(s_i^{(1)}) \cdot \prod_{H \in F_i[s_i^{(1)}]} \sigma(s : s(H) = a_H^{(1)}) \\
=&r_i(\emptyset)(s_i^{(1)}) \cdot \sigma(s : s(H) = a_H^{(1)} \text{ for all } H \in F_i[s_i^{(1)}]),
\end{aligned}$$

where the second line comes from the probability η assigns to cylinder sets, and the third line comes from the fact that $\sigma \in \times_k \Delta(\mathbb{S}_k)$ involves independent mixing of pure strategies across different players.

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 $(y_i[T], (s_i^{(T+1)}, (a_H^{(T+1)})_{H \in F_i[s_i^{(T+1)}]}))$, where $y_i[T]$ is a length-T history and $(s_i^{(T+1)}, (a_H^{(T+1)})_{H \in F_i[s_i^{(T+1)}]})$ is a one-period history corresponding to what happens in period $T+1$. Therefore, we may write for each $y_i[T+1]$,

$$\phi_{i,T+1}^{RM}(y_i[T+1]) = \phi_{i,T}^{RM}(y_i[T]) \cdot \phi_{i,T+1|T}^{RM}((s_i^{(T+1)}, (a_H^{(T+1)})_{H \in F_i[s_i^{(T+1)}]})|y_i[T]),$$

and

$$\phi_{i,T+1}^{\eta}(y_i[T+1]) = \phi_{i,T}^{\eta}(y_i[T]) \cdot \phi_{i,T+1|T}^{\eta}((s_i^{(T+1)}, (a_H^{(T+1)})_{H \in F_i[s_i^{(T+1)}]})|y_i[T]),$$

where $\phi_{i,T+1|T}^{RM}$ and $\phi_{i,T+1|T}^{\eta}$ are the conditional probabilities of the form "having history $(s_i^{(T+1)}, (a_H^{(T+1)})_{H \in F_i[s_i^{(T+1)}]})$ 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 $\phi_{i,T}^{RM} = \phi_{i,T}^{\eta}$, we can conclude $\phi_{i,T+1}^{RM} = \phi_{i,T+1}^{\eta}$.

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

$$\phi_{i,T+1|T}^{RM}((s_i^{(T+1)}, (a_H^{(T+1)}))|y_i[T]) = r_i(y_i(T))(s_i^{(T+1)}) \cdot \sigma(s : s(H) = a_H^{(T+1)} \text{ for all } H \in F_i[s_i^{(T+1)}]),$$

since the matching is independent across periods.

But 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 $\mathfrak{A}$ is such that the response $(a_H^{(T+1)})_{H \in F_i[s_i^{(T+1)}]}$ 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 $H \in \mathcal{H}_{-i}$, let the non-negative integer N_H represent the number of times i has observed play on the information set H in the history $y_i[T]$. For each H, let $(a_{n_H,H})_{n_H=1}^{N_H}$ represent the sequence of opponent actions observed on H in chronological order. The history $y_i[T]$ so far shows i is facing a response sequence in the cylinder set consistent with $((a_{n_H,H})_{n_H=1}^{N_H})_{H\in\mathcal{H}}$. If $\mathfrak{A}$ is to respond to i's next play of $s_i^{(T+1)}$ with $a_H^{(T+1)}$ on the $s_i^{(T+1)}$-relevant information sets, then $\mathfrak{A}$ must belong to a more restrictive cylinder set, satisfying the restrictions:

$$((a_{n_H,H})_{n_H=1}^{N_H})_{H\in\mathcal{H}\setminus F_i[s_i^{(T+1)}]}, ((a_{n_H,H})_{n_H=1}^{N_H+1})_{H\in F_i[s_i^{(T+1)}]},$$

where for each $H \in F_i[s_i^{(T+1)}]$, $a_{N_H+1,H} = a_I^{(T+1)}$. The conditional probability is then given by the ratio of η-probabilities of these two cylinder sets, which from the definition of η must be $\prod_{H\in F_i[s_i^{(T+1)}]} \sigma(s : s(H) = a_H^{(T+1)})$. As before, the independence of σ across players means this is equal to $\sigma(s : s(H) = a_H^{(T+1)}$ for all $H \in F_i[s_i^{(T+1)}])$. □

OA 4 PCE and Extended Proper Equilibrium

We have already seen that extended proper equilibria need not satisfy player compatibility. Here we give an example of the converse: an equilibrium that satisfies player compatibility and is not extended proper.

Example 4. Consider a three-player game where Row chooses a row, Column chooses a column, and Geo chooses a matrix. The payoff to Geo is always 0. The payoffs to Row and Column are listed in the tables below.

West	**Left**	**Right**
Up	1,1	1,1
Down	0,1	0,0

East	**Left**	**Right**
Up	1,0	1,1
Down	0,0	0,0

The strategy profile (**Up**, **Left**, **West**) is not an extended proper equilibrium, because Column would deviate to **Right** against a tremble where Row's costly deviation to **Down** is much rarer than Geo's costless deviation to **East**. However, it is a PCE. For example, take a sequence of trembles $\boldsymbol{\epsilon}^{(t)}$ where the minimum probability on each action is $1/t$. Such uniform trembles always satisfy player compatibility, so every $\boldsymbol{\epsilon}^{(t)}$-equilibrium in this sequence is an $\boldsymbol{\epsilon}^{(t)}$-PCE, and we see that Column is indifferent between **Left** and **Right** if Row deviates to **Down** exactly as much as Geo deviates to **East**. ♦