Online Appendix for "Player-Compatible Learning and Player-Compatible Equilibrium"

Drew Fudenberg and Kevin He

OA1 Proofs Omitted from the Appendix

OA1.1 Proof of Proposition 1

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

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

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

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

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

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

OA1.2 Proof of Proposition 2

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

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

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

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

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

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

OA2 Refinements in the Link-Formation Game

Proposition OA.1. *Each of the following refinements selects the same subset of pure Nash equilibria when applied to the anti-monotonic and co-monotonic versions of the link-formation game: extended proper equilibrium, proper equilibrium, trembling-hand perfect equilibrium, p-dominance, Pareto efficiency, and strategic stability. Pairwise stability does not apply to the link-formation game. Finally, the link-formation game is not a potential game.*

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

For the version with (q_i) co-monotonic with (c_i), consider the same strategies without re-scalings, i.e. $\boldsymbol{\beta} = \mathbf{1}$. Then already the loss to playing **Active** is 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 $(1 - p) \cdot 0 + p \cdot (10 - c_i) + (1 - p) \cdot 0 + p \cdot (30 - c_i) > 0$ 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 $1 - \epsilon'$. In particular, if $\epsilon' < 1/100$, then the Nash equilibrium in the ϵ-constrained game is not η-close to the "no links" equilibrium. 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 $0 < \epsilon' < 1/100$ 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 their links while preserving the other.

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

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

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

OA3 Replication Invariance of PCE

This section argues that PCE is invariant to adding duplicate copies of strategies to the game. Fix a *base game* with the strategic form $(\mathbb{I}, (\mathbb{S}_i, u_i)_{i \in \mathbb{I}})$ where $\mathbb{I}$ is the set of players, each player i has a finite strategy set $\mathbb{S}_i$ and utility function $u_i : \mathbb{S} \to \mathbb{R}$.

Definition OA.1. An *extended game with duplicates* is any game with the strategic form $(\mathbb{I}, (\bar{\mathbb{S}}_i, \bar{u}_i)_{i \in \mathbb{I}})$ such that, for every $i \in \mathbb{I}$, $\bar{\mathbb{S}}_i \subseteq \mathbb{S}_i \times \mathbb{N}$ is a finite set with $\text{proj}_{\mathbb{S}_i}(\bar{\mathbb{S}}_i) = \mathbb{S}_i$ and

$\bar{u}_i((s_j, n_j)_{j \in \mathbb{I}}) = u_i(s)$ for all $s \in \mathbb{S}$ and $(n_j)_{j \in \mathbb{I}} \in \mathbb{N}^{\mathbb{I}}$ with $(s_j, n_j)_{j \in \mathbb{I}} \in \bar{\mathbb{S}}$.

The interpretation is that each player i can have multiple copies of every strategy they had in the base game, and could have different numbers of copies of different strategies, where duplicate copies of the same strategy have the same payoff consequences. Mapping back to the learning framework, we think of different strategies of i in the extended game as different learning opportunities about $-i$'s play. Copies of different strategies are learning opportunities that provide orthogonal information, while copies of the same strategy provide the same information. As an example, suppose that in the Restaurant Game the critic can arrive at the restaurant by taking the red bus or the blue bus, and the color of the bus is not observed by other players, does not change anyone's payoffs, and does not change what the critic observes. We can then replace $\mathbf{R}_c$ with two actions $\mathbf{R}_c^{red}, \mathbf{R}_c^{blue}$ at the critic's information set and expand the game tree, letting $\mathbf{R}_c^{red}$ and $\mathbf{R}_c^{blue}$ both have the same payoff consequences as $\mathbf{R}_c$ in the original game. This modified game is an extended game with duplicates for the original game.

Subsection OA3.1 defines player-compatible trembles and PCE in extended games with duplicates. Using the compatibility relation $\succsim$ from the base game, a tremble profile in the extended game with duplicates is player compatible if the *sum* of tremble probabilities assigned to all copies of s_i^* exceeds the sum assigned to all copies of s_j^*, whenever $s_i^* \succsim s_j^*$. PCE is then defined using this restriction on trembles. We show that the set of PCE in the base game coincides with the set of PCE in the extended game with duplicates.

This definition of player-compatible trembles in extended games with duplicates fits with our interpretation of trembles as experimentation frequencies and an analysis of how learning dynamics in the extended game compare with those in the base game. The idea is that if all copies of a strategy s_i give i the same information about others' play, then i should be exactly indifferent between all such copies after all histories in the learning process. Holding fixed initial beliefs and the social distribution, i's weighted lifetime average play of s_i in the base game should then equal the sum of their weighted lifetime average plays of all copies of s_i in the extended game with duplicates. Thus, any comparisons that hold between the "tremble" probabilities of i onto s_i^* and j onto s_j^* in the base game must also hold between the sum of "tremble" probabilities of i onto the copies of s_i^* and j onto the copies of s_j^* in the extended game. We formalize this intuition in binary participation games in Subsection OA3.2 for rational learning and weighted fictitious play.

OA3.1 PCE in Extended Games with Duplicates

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

Definition OA.2. Tremble profile $\bar{\boldsymbol{\epsilon}}$ is *player compatible in the extended game* if $\sum_{n_i} \bar{\boldsymbol{\epsilon}}(s_i^*, n_i) \geq \sum_{n_j} \bar{\boldsymbol{\epsilon}}(s_j^*, n_j)$ for all $i, j \in \mathbb{I}$, $s_i^* \in \mathbb{S}_i$, $s_j^* \in \mathbb{S}_j$ such that $s_i^* \succsim s_j^*$, where $\succsim$ is the player-compatibility relation from the base game. An $\bar{\boldsymbol{\epsilon}}$-constrained equilibrium where $\bar{\boldsymbol{\epsilon}}$ is player compatible is called a *player-compatible* $\bar{\boldsymbol{\epsilon}}$*-constrained equilibrium* (or $\bar{\boldsymbol{\epsilon}}$*-PCE*).

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

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

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

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

Lemma OA.1. *If $\bar{\sigma}$ is an $\bar{\boldsymbol{\epsilon}}$-constrained equilibrium in the extended game, then $\mathscr{C}(\bar{\sigma})$ is a $\mathscr{C}(\bar{\boldsymbol{\epsilon}})$-constrained equilibrium in the base game. If σ is an $\boldsymbol{\epsilon}$-constrained equilibrium in the base game, then $\mathscr{E}(\sigma)$ is an $\mathscr{E}(\boldsymbol{\epsilon})$-constrained equilibrium in the extended game.*

PCE is defined as usual in the extended game.

Definition OA.3. 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.

OA3.2 Learning and Trembles in Binary Participation Games with Duplicates

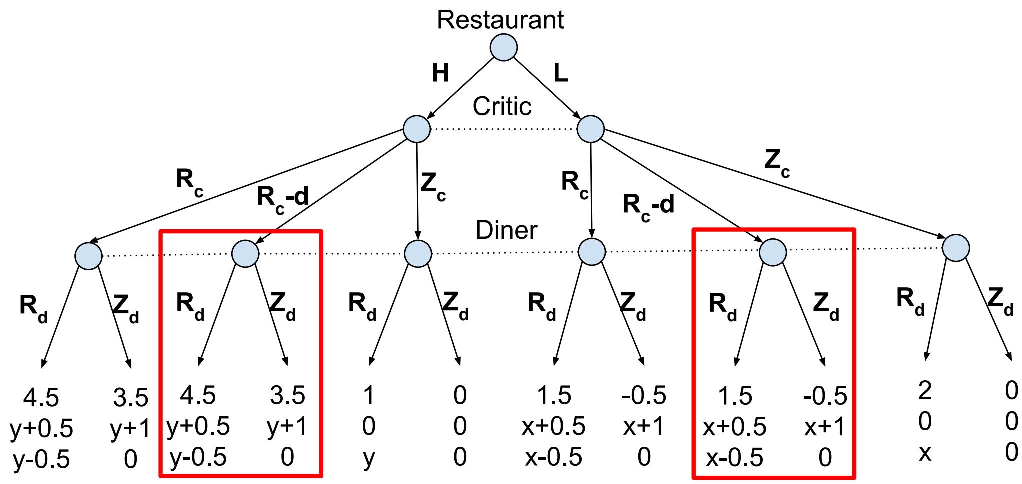
We give the simplest illustration of how learning dynamics in extended games with duplicates relate to those in the base game, using binary participation games. These results can also be developed for other factorable games, but at the cost of more complicated notation.

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

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

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

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



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

Definition OA.4. Let $\zeta : \tilde{Y}_i \to Y_i$ be such that for $\tilde{y}_i \in \tilde{Y}_i$, $\zeta(\tilde{y}_i) \in Y_i$ replaces every instance of **In-d** with **In**. Learning policies $\tilde{r}_i : \tilde{Y}_i \to \Delta(\{\mathbf{In}, \mathbf{In\text{-}d}, \mathbf{Out}\}\}$ and $r_i : Y_i \to \{\mathbf{In}, \mathbf{Out}\}$ are *equivalent up to duplicates* if for every $\tilde{y}_i \in \tilde{Y}_i$, if $r_i(\zeta(\tilde{y}_i)) = \mathbf{Out}$, then also $\tilde{r}_i(\tilde{y}_i)(\mathbf{Out}) = 1$. If $r_i(\zeta(\tilde{y}_i)) = \mathbf{In}$, then $\tilde{r}_i(\tilde{y}_i)(\mathbf{In}) + \tilde{r}_i(\tilde{y}_i)(\mathbf{In\text{-}d}) = 1$.

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

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

Proposition OA.3. *Suppose stochastic learning policy $\tilde{r}_i$ in the extended game is equivalent up to duplicates with the learning policy r_i in the base game.*

- *For a fixed patience parameter* $0 \leq \delta < 1$ *and regular prior* g_i *over others' play,*[1] r_i *is* OPT_i *if and only if* $\tilde{r}_i$ *is an optimal learning policy with the extended game.*

- *For a fixed decay parameter* $0 \leq \rho < 1$ *and initial counts* $N_h^{a_h}(\varnothing)$, r_i *is* WFP_i *if and only if after every* $\tilde{y}_i \in \tilde{Y}_i$, $\tilde{r}_i(\tilde{y}_i)$ *is supported on strategies that maximize payoffs under the weighted fictitious play conjecture of* $-i$*'s play.*

- *For* $-i$ *equivalent social distributions* $\sigma, \tilde{\sigma}$ *for the base game and extended games,* $\phi_i(\boldsymbol{In}; r_i, \sigma_{-i}) = \phi_i(\boldsymbol{In}; \tilde{r}_i, \tilde{\sigma}_{-i}) + \phi_i(\boldsymbol{In\text{-}d}; \tilde{r}_i, \tilde{\sigma}_{-i})$.

Theorem 2 shows that in the baseline binary participation game, $\phi_i(\mathbf{In}_i; r_i, \sigma_{-i}) \geq \phi_j(\mathbf{In}_j; r_j, \sigma_{-j})$ for every social distribution σ whenever $\mathbf{In}_i \succsim \mathbf{In}_j$ and r_i, r_j are either OPT or WFP under the same "initial conditions," where $\mathbf{In}_i$ and $\mathbf{In}_j$ refer to i and j's copies of $\mathbf{In}$. Combining this result with the above proposition, we find a motivation for player-compatible trembles in the extended game. If $\tilde{r}_i, \tilde{r}_j$ are either OPT with the same δ and same prior beliefs about $-ij$'s play, or WFP with the same initial counts on $-ij$'s information sets, then $\phi_i(\mathbf{In}_i; \tilde{r}_i, \tilde{\sigma}_{-i}) + \phi_i(\mathbf{In\text{-}d}_i; \tilde{r}_i, \tilde{\sigma}_{-i}) \geq \phi_j(\mathbf{In}_j; \tilde{r}_j, \tilde{\sigma}_{-j}) + \phi_j(\mathbf{In\text{-}d}_j; \tilde{r}_j, \tilde{\sigma}_{-j})$ for any social distribution $\tilde{\sigma}$ in the extended game, where $\mathbf{In\text{-}d}_i$ and $\mathbf{In\text{-}d}_j$ refer to i and j's copies of $\mathbf{In\text{-}d}$.

OA3.3 Proofs

OA3.3.1 Proof of Lemma OA.1

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

Again by contraposition, suppose $\mathscr{E}(\sigma)$ is not an $\mathscr{E}(\boldsymbol{\epsilon})$-constrained equilibrium in the extended game. This means some i assigns more than the required weight to some $(s_i', n_i') \in \bar{\mathbb{S}}_i$ that does not best respond to $\mathscr{E}(\sigma)_{-i}$. This implies s_i' does not best respond to σ_{-i}. By the definition of $\mathscr{E}(\boldsymbol{\epsilon})$ and $\mathscr{E}(\sigma)$, if $\mathscr{E}(\sigma)_i(s_i', n_i') > \mathscr{E}(\boldsymbol{\epsilon})(s_i', n_i')$, then also $\mathscr{E}(\sigma)_i(s_i', n_i) >$

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

$\mathscr{E}(\boldsymbol{\epsilon})(s_i', n_i)$ for every n_i such that $(s_i', n_i) \in \bar{\mathbb{S}}_i$. Therefore, we also have $\sigma_i(s_i') > \boldsymbol{\epsilon}(s_i')$, so σ is not an $\boldsymbol{\epsilon}$-constrained equilibrium in the base game as desired. ☐

OA3.3.2 Proof of Proposition OA.2

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 $\mathscr{C}(\bar{\boldsymbol{\epsilon}}^{(t)})$-constrained 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. ☐

OA3.3.3 Proof of Proposition OA.3

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

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

$(\mathbf{In}, x)] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i) \times (\mathbf{In\text{-}d}, x)] = 0$. On the other hand, if $r_i(y_i) = \mathbf{In}$, then using the fact that $r_i, \tilde{r}_i$ are equivalent up to duplicates, $\tilde{r}_i(\tilde{y}_i)(\mathbf{In}) + \tilde{r}_i(\tilde{y}_i)(\mathbf{In\text{-}d}) = 1$ for all $\tilde{y}_i \in \psi^{-1}(y_i)$. Thus, for all $x \in \mathbb{R}$, by the inductive hypothesis, $\mathbb{P}_{r_i,\sigma_{-i}}[(y_i, \mathbf{Out}, x)] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i) \times (\mathbf{Out}, x)] = 0$, and $\mathbb{P}_{r_i,\sigma_{-i}}[(y_i, \mathbf{In}, x)] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i) \times (\mathbf{In}, x)] + \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i) \times (\mathbf{In\text{-}d}, x)]$. In either case, we get $\mathbb{P}_{r_i,\sigma_{-i}}[y_i] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\psi^{-1}(y_i)]$ for all $y_i \in (\{\mathbf{In}, \mathbf{Out}\} \times \mathbb{R})^{t+1}$, and also $\mathbb{P}_{r_i,\sigma_{-i}}[X_i^t = \mathbf{Out}] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\tilde{X}_i^t = \mathbf{Out}]$. By induction we get $\mathbb{P}_{r_i,\sigma_{-i}}[X_i^t = \mathbf{Out}] = \mathbb{P}_{\tilde{r}_i,\tilde{\sigma}_{-i}}[\tilde{X}_i^t = \mathbf{Out}]$ for every $t \geq 1$, thus $\phi_i(\mathbf{In}; r_i, \sigma_{-i}) = \phi_i(\mathbf{In}; \tilde{r}_i, \tilde{\sigma}_{-i}) + \phi_i(\mathbf{In\text{-}d}; \tilde{r}_i, \tilde{\sigma}_{-i})$. □

References

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

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

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

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