MONITORING, MARKET PRIMITIVES, AND THE STABILITY OF ALGORITHMIC COLLUSION SUPPLEMENTARY APPENDIX

CLEMENS POSSNIG

I. The Algorithm Class

The following assumptions are sufficient for the results stated in Section 3 of the main text to go through, upon minor extensions to known results from stochastic approximation theory, to be found in **benaim1999dynamics**, Borkar (2009), Benaïm and Faure (2012). The setting considered here generalizes the one studied in the main text by allowing for a non-vanishing bias term, a robustification of the results discussed below.

This robustification means it is sufficient for practitioners to verify smoothness and bound a possible asymptotic bias, without needing to know the specific functional form of the bias.

In keeping with the main text, we have $N > 1$ learning agents, and consider state variables taking $K > 0$ different values. Define $\overline{\mathbf{Y}} \subseteq \mathbb{R}^{NK}$ as the space of actions, stacked over i. The following constructs the family of bias functions that the results extend to.

Definition 1. *For $\gamma > 0$, let $\mathcal{B}^k_\gamma$ be the set of $\mathcal{C}^k$ functions with bounded derivatives :*

$$\mathcal{B}^k_\gamma = \left\{ g : \overline{\mathbf{Y}} \to \mathbb{R}^{NK} \mid \sup_{x \in \overline{\mathbf{Y}}} \|g(x)\| + \sum_{j=1}^{k} \sup_{x \in \overline{\mathbf{Y}}} \|D^j g(x)\| \leq \gamma \right\}, \tag{1}$$

where $D^j g$ represents the j'th derivative.

For all results to follow, state variables will be fixed. I identify time periods by n in order to distinguish the continuous timescale t used in the associated continuous time systems. Allowing for non-vanishing bias g, the algorithms can jointly (stacked over i) be written as

$$\boldsymbol{\rho}_{n+1} = \boldsymbol{\rho}_n + \alpha_t \boldsymbol{A} \left[\Psi(\boldsymbol{\rho}_n) + g(\boldsymbol{\rho}_n) + \boldsymbol{\delta}_n + \boldsymbol{M}_{n+1} \right], \tag{2}$$

where now $g(\boldsymbol{\rho}_n) + \boldsymbol{\delta}_n + \boldsymbol{M}_{n+1} = \hat{\Psi}_n - \Psi(\boldsymbol{\rho}_n)$ is the representation of period-n errors in the critic estimation, as outlined in Section 2 of the main text, where in addition $g \in \mathcal{B}^2_\gamma$ for some

$\gamma \geq 0$ represents a non-vanishing bias term. As in the main text, $\alpha_t = \alpha_t^1$ is agent 1's stepsize schedule, $\boldsymbol{A}$ is a diagonal matrix with i'th diagonal entry equal to $\lim_{t\to\infty} \frac{\alpha_t^i}{\alpha_t^1} \in (0, \infty)$, which will weight the limiting ODE by limiting relative updating speeds. Fixing $g(\boldsymbol{\rho})$, we often write $\Psi_g(\boldsymbol{\rho}) = \Psi(\boldsymbol{\rho}) + g(\boldsymbol{\rho})$ to save space.

I re-state here the version of Assumption 6 of Appendix A in the main text.

Assumption 1. *Let $\mathcal{F}_n$ be the σ-field generated by $\{\boldsymbol{\rho}_n, \boldsymbol{\delta}_n, \boldsymbol{M}_n, \boldsymbol{\rho}_{n-1}, \boldsymbol{\delta}_{n-1}, \boldsymbol{M}_{n-1} \ldots, \boldsymbol{\rho}_0, \boldsymbol{\delta}_0, \boldsymbol{M}_0\}$, i.e. all the information available to the updating rule at a given period n.*

(1) Stepsizes α_n^i satisfy, for all i, to be square-summable, but not summable.

(2) For all i, j, $\lim_{n\to\infty} \frac{\alpha_n^i}{\alpha_n^j}$ exists and lies in (c, ∞), for some $c > 0$.

(3) Ψ is admissible.

(4) $\boldsymbol{M}_{n+1}$ is a Martingale-difference noise. There is $0 < \bar{M} < \infty$ and $q \geq 2$ such that for all n

$$\mathbb{E}[\boldsymbol{M}_{n+1} \,|\mathcal{F}_n] = 0; \quad \mathbb{E}\left[\|\boldsymbol{M}_{n+1}\|^q \mid \mathcal{F}_n\right] < \bar{M} \;\; \mathcal{F}_0 - \textit{almost surely.}$$

(5) There exists a continuous function

$$\Omega : \overline{Y} \mapsto \mathcal{J}(\overline{Y}),$$

where $\mathcal{J}(\overline{Y})$ is the space of positive definite matrices given vectors in $\overline{Y}$, such that for all n

$$\mathbb{E}[\boldsymbol{M}_{n+1}\boldsymbol{M}'_{n+1} \,|\mathcal{F}_n] = \Omega(\boldsymbol{\rho}_n),$$

for all $\boldsymbol{\rho}_n \in \overline{Y}$.

(6)

$$\mathbb{E}\big[\|\boldsymbol{\delta}_n\| \mid \mathcal{F}_n\big] = o(b_n),$$

where $b_n \to 0$ satisfies $\max_i \lim_{n\to\infty} \frac{\alpha_n^i}{b_n} = 0$.

(7)

$$\sup_{n\geq 0} \mathbb{E}\left[\|\boldsymbol{\delta}_n\|^2\right] < \infty,$$

Point (1) is known as the Robbins-Monro condition (Robbins and Monro 1951) on stepsizes. It ensures that stepsizes converge slowly enough so that the whole real line can be mapped (as a continuous-time interval), while converging not too slowly in order for error terms to be averaged out. (2) ensures that all stepsizes lie within the same order of magnitude. Point (3) ensures global integrability and uniqueness of solutions to $\dot{\boldsymbol{\rho}} = \Psi(\boldsymbol{\rho})$. In the example of ACQ, it is an assumption on payoffs W^i, and that best responses can't grow too quickly. Point (4) implies that given current information in period n, new errors due to $n+1$'s estimator of Ψ are well-behaved. It is a common assumption in stochastic approximation theory. Point (5) ensures that some variance in error terms remains for all n; this is satisfied e.g. if the estimation of Ψ involves exploratory noise, or stochasticity during the estimation as is true under randomized Bellman-iteration schemes. This assumption will be the main driver that pushes iterations away from unstable equilibria. Point (6) ensures that the bias term vanishes faster than stepsizes. Point (7) is a further regularity condition on the bias term.

II. Proofs for the general Algorithm Class

Recall the following definitions

Definition 2. *Define*

(1) $E_S \subset \overline{Y}$ to be the set of Nash equilibria in policy profiles based on payoff functions W^i. In other words, E_S is the set of profiles $\boldsymbol{\rho}^*$ s.t. $\boldsymbol{\rho}^* \in \bar{B}_S(\boldsymbol{\rho}^*)$.

(2) $\boldsymbol{\rho}^* \in E_S$ as 'differential Nash equilibrium' if $\boldsymbol{\rho}^*$ is interior, first order conditions hold for each agent at $\boldsymbol{\rho}^*$, and the Hessian of each agent's optimization problem at $\boldsymbol{\rho}^*$ is negative definite. Define the subset of such $\boldsymbol{\rho}^*$ as $E_S^* \subseteq E_S$.

Given these definitions regarding the underlying payoff environment, assume:

Assumption 2 (Equilibrium existence and differentiability)**.**

(1) Given state variable S, E_S^* is nonempty.

Definition 3. *Given some ODE $\dot{\rho} = f(\rho)$, let ρ^* be a rest point of $f(\rho)$. Let $\Lambda = eigv[Df(\rho^*)]$ the set of eigenvalues of the linearization of f at ρ^*. For a complex number z, let $\mathbf{Re}[z] \in \mathbb{R}$ be the real part. ρ^* is*

- *Hyperbolic if* $\mathbf{Re}[\lambda] \neq 0$ *holds for all* $\lambda \in \Lambda$.
- *Asymptotically stable if* $\mathbf{Re}[\lambda] < 0$ *holds for all* $\lambda \in \Lambda$.
- *Linearly unstable if* $\mathbf{Re}[\lambda] > 0$ *holds for at least one* $\lambda \in \Lambda$.

Also define the limit set as

$$L_{S,g} = \bigcap_{n \geq 0} cl\left(\{\rho_\ell \,|\ell \geq n\}\right),$$

the set of limits of convergent subsequences ρ_{t_k}, keeping track of the existence of the bias function g.

Theorem 1

The first result extends Theorem 1 in the main text:

Theorem 1. *Let* $\boldsymbol{\rho}^* \in E_S^*$ *be asymptotically stable for* Ψ*. Then for all* γ *small enough and all* $g \in \mathcal{B}_\gamma^1$ *there is a profile* $\boldsymbol{\rho}^g$ *such that*

(1) $\sup_{g \in \mathcal{B}_\gamma^1} |\boldsymbol{\rho}^g - \boldsymbol{\rho}^*| \to 0$ *as* $\gamma \to 0$.

(2) $\mathbb{P}\left[L_{S,g} = \{\boldsymbol{\rho}^g\}\right] > 0$.

Proof. Notice that accordingly, rest point ρ^g may not be an exact Nash equilibrium of the underlying game, but an ε-equilibrium:

Definition 4. *A profile* $\boldsymbol{\rho}$ *is an* ε*-equilibrium if for all players* i *all individual profiles* $\boldsymbol{\rho}' \in \overline{\mathbf{Y}}$ *and states* $s \in \mathbf{S}$

$$W^i(\boldsymbol{\rho}, s) \geq W^i(\boldsymbol{\rho}', \boldsymbol{\rho}^{\text{-}i}, s) - \varepsilon.$$

The implied statement in a game as e.g. outlined in section 4 of the main text would then be:

Corollary 1. *Let* $\boldsymbol{\rho}^* \in E$ *be asymptotically stable for* Ψ*. Then for all* $\varepsilon > 0$ *there is* $\gamma \geq 0$ *small enough such that for all* $g \in \mathcal{B}_\gamma^1$ *there is a profile* $\boldsymbol{\rho}^g$ *such that*

(1) $\boldsymbol{\rho}^g$ *is an* ε*-equilibrium.*

(2) $\mathbb{P}[L_{S,g} = \{\boldsymbol{\rho}^g\}] > 0$.

Now to the proof: Taking Ψ_g that satisfies Assumption 1, a solution to the differential equation $\dot{\rho} = \Psi_g(\rho)$ can be defined as a flow $\phi : \mathbb{R} \times \overline{\mathbf{Y}} \to \overline{\mathbf{Y}}$. The following definition can be found in Mertikopoulos, Hsieh, and Cevher (2024, Section 4.2):

Definition 5. *Take a flow $\phi : \mathbb{R} \times \mathcal{M} \to \mathcal{M}$ given some metric space $\mathcal{M}$, and a nonempty compact subset $\mathcal{S} \subseteq \mathcal{M}$. We say*

(1) $\mathcal{S}$ is invariant under ϕ if $\phi_t(\mathcal{S}) = \mathcal{S}$ for all $t \in \mathbb{R}$.

(2) $\mathcal{S}$ is an attractor of ϕ if it admits a neighborhood $\mathcal{W} \subseteq \mathcal{M}$ such that $d\left(\phi_t(w), \mathcal{S}\right) \to 0$ uniformly in $w \in \mathcal{W}$ as $t \to \infty$.

(3) $\mathcal{S}$ is internally chain transitive (ICT) if it is invariant and $\phi\big|_{\mathcal{S}}$ has no attractors except $\mathcal{S}$.

Point (3) is the main object of interest to algorithmic learners. Indeed, one can think of ICT sets as a generalization to periodic orbits of an ordinary differential equation, where solutions to the ODE are allowed to take on arbitrarily small jumps. This generalization turns out to be very useful in the description of long run behavior of discrete-time stochastic systems. Importantly, ICT sets include rest points and limit cycles (if they exist). Consider Papadimitriou and Piliouras (2018) for an intuitive discussion. The following result shows why these sets are of importance in our analysis:

Proposition 1. *Impose Assumption 1. Almost surely, $L_{S,g}$ is an ICT set of the differential equation*

$$\dot{\boldsymbol{\rho}} = \boldsymbol{A}\Psi_g(\boldsymbol{\rho}(t)),$$

where A is a diagonal matrix where all diagonal entries are strictly positive, representing the limiting relative stepsizes of all algorithms.

Proof. The proof is a slight generalization of Borkar (2009, Theorem 2). The approach is to construct a linear interpolation of (2), and show that this will shadow solutions to $\dot{\boldsymbol{\rho}} = \boldsymbol{A}\Psi_g(\boldsymbol{\rho}(t))$ asymptotically, for large enough t. To deal with potentially asymptotically differing stepsizes, take e.g. 1's stepsize schedule α_n^1. By Assumption 1 (ii), we can multiply and divide each algorithm's iteration and write for each i

$$\begin{aligned}\boldsymbol{\rho}^i_{n+1} &= \boldsymbol{\rho}^i_n + \alpha^1_n \frac{\alpha^i_n}{\alpha^1_n}\left[\Psi^i(\boldsymbol{\rho}_n) + g^i(\boldsymbol{\rho}_n) + \boldsymbol{\delta}^i_n + M^i_{n+1}\right] \\ &= \alpha^1_n \bar{\alpha}^i \left[\Psi^i_g(\rho_n) + \boldsymbol{\delta}^i_n + \boldsymbol{M}^i_{n+1}\right] + \alpha^1_n o_P(1),\end{aligned}$$

where $g^i(\boldsymbol{\rho}_n) \in \mathbb{R}^K$ is i's bias subvector of $g(\boldsymbol{\rho}_n)$, and the last term is due to the vanishing error $\left(\frac{\alpha^i_n}{\alpha^1_n} - \bar{\alpha}^i\right)$, which can be handled analogously to the error terms $\boldsymbol{\delta}_n$, which are discussed below. $\bar{\alpha}^i$ is then the ith diagonal element of scaling matrix $\boldsymbol{A}$ in the statement of this proposition. For notational ease, in the following we then write $\alpha_n = \alpha^1_n$, and proceed with the proof.

Following the notation in Borkar (2009), introduce:

$$\tau_0 = 0; \quad \tau_n = \sum_{i=1}^{n} \alpha_i; \quad m(t) = \sup\{k \geq 0 : \tau_k \leq t\}.$$

Then, construct the interpolation as

$$X(\tau_n + s) = \boldsymbol{\rho}_n + s\frac{\boldsymbol{\rho}_{n+1} - \boldsymbol{\rho}_n}{\alpha_{n+1}}, \quad s \in [0, \alpha_{n+1}]. \tag{3}$$

Following the proof of Borkar (2009, Theorem 2), we only need to take care of the additional term $\boldsymbol{\delta}_n$ present in iteration (2). We will consider the accumulated $\boldsymbol{\delta}_n, \boldsymbol{M}_{n+1}$ error terms. First, note that

$$\begin{aligned}&\sup\left\{\left\|\sum_{\ell=n}^{k-1} \alpha_{\ell+1}\Big(\boldsymbol{\delta}_{\ell+1} + \boldsymbol{M}_{\ell+2}\Big)\right\| : \ k = n+1, \ldots, m(\tau_n + T)\right\} \\ &\leq \sup_{n \leq k \leq m(\tau_n+T)-1} \left\|\sum_{\ell=n}^{k} \alpha_{\ell+1}\Big(\boldsymbol{M}_{\ell+2}\Big)\right\| + \sup_{n \leq k \leq m(\tau_n+T)-1} \left\|\sum_{\ell=n}^{k} \alpha_{\ell+1}\Big(\boldsymbol{\delta}_{\ell+1}\Big)\right\| \\ &= R_n + \sup_{n \leq k \leq m(\tau_n+T)-1} H^k_n.\end{aligned}$$

By Assumption 1, R_n is a standard error term in stochastic approximation theory, satisfying the usual assumptions of Robbins-Monro algorithms with martingale difference noise. It is a standard result that R_n converges almost surely to zero.[1] We need to take care of the

[1] See e.g. Faure and Roth (2010, Proposition 2.16).

additional term δ_n present in iteration (2). It suffices to show that, for all $T > 0$

$$\sup_{n \le k \le m(\tau_n+T)-1} H_n^k \to 0, \tag{4}$$

almost surely as $n \to \infty$. First, note that

(5)

$$H_n^k \le \sup_{n \le k \le m(\tau_n+T)-1} \left\| \sum_{\ell=n}^{k} \alpha_{\ell+1} \Big(\|\boldsymbol{\delta}_{\ell+1}\| - \mathbb{E}\big[\|\boldsymbol{\delta}_{\ell+1}\| \mid \mathcal{F}_{\ell+1}\big] \Big) \right\| + \sum_{\ell=n}^{m(\tau_n+T)-1} \alpha_{\ell+1} \mathbb{E}\big[\|\boldsymbol{\delta}_{\ell+1}\| \mid \mathcal{F}_{\ell+1}\big]$$

(6)

$$= R_{2,n} + H_{2,n},$$

where $\mathcal{F}_\ell$ is the filtration defined in Assumption 1. Now, by Assumption 1 $(6)-(7)$, $R_{2,n}$ is the supremum on another martingale difference noise term with bounded variance, just as R_n. Thus, again for $R_{2,n}$ we have almost sure convergence to zero. As for $H_{2,n}$, recall from Assumption 1 6 that $\mathbb{E}\big[\|\boldsymbol{\delta}_{\ell+1}\| \mid \mathcal{F}_{\ell+1}\big] = o(b_\ell)$. Hence, there exists some $C_H > 0$ such that for all n large enough,

$$H_{2,n} \le C_H \sum_{\ell=n}^{m(\tau_n+T)-1} \alpha_{\ell+1} b_{\ell+1} \le \sum_{\ell=n}^{m(\tau_n+T)-1} \alpha_{\ell+1}^2,$$

by assumption that $\max_i \lim_{n\to\infty} \frac{b_n}{\alpha_n^i} = 0$. Thus, by square summability of α_i, the sum above must converge to zero in n, and therefore $H_{2,n} \to 0$ as well, and the result (4) follows.

Hence, the arguments in Borkar (2009, Lemma 1, Theorem 2) extend to this case as well, which concludes the proof. □

As $\boldsymbol{\rho}^*$ is a differential equilibrium by assumption, point (1) of Theorem 1 follows by an application of the inverse function theorem as shown below. We will prove something more general: as long as ρ^* is hyperbolic (c.f. Definition 3), point (2) holds.

This follows because when $\boldsymbol{\rho}^*$ is hyperbolic, there is a neighborhood U around 0 such that Ψ has a differentiable inverse on U. Next, note that $\boldsymbol{\rho}^g$ solves

$$\Psi_g(\boldsymbol{\rho}^g) = 0.$$

Since $\|g\|_1 \leq \gamma$, for γ small enough, $\Psi(\boldsymbol{\rho}^g) \in U$ must hold. Then there is some $L_{\Psi^{-1}} > 0$ such that

$$\begin{aligned} \|\boldsymbol{\rho}^g - \boldsymbol{\rho}^*\| &= \|\Psi^{-1}(\Psi(\boldsymbol{\rho}^g)) - \Psi^{-1}(0)\| \\ &\leq L_{\Psi^{-1}} \|\Psi(\boldsymbol{\rho}^g)\| \leq L_{\Psi^{-1}} \gamma, \end{aligned}$$

where the first inequality follows because Ψ^{-1} is differentiable and $\Psi(\boldsymbol{\rho}^*) = 0$, and the second by the definition of $\Psi(\boldsymbol{\rho}^g)$. Since the right hand side is independent of g, the bound is uniform.

For point (2), we first need to verify that all $\boldsymbol{\rho}^g$ close enough to $\boldsymbol{\rho}^*$ must also be asymptotically stable. The next Lemma gives a more general result:

Lemma 1. *Suppose $\boldsymbol{\rho}^*$ is hyperbolic. Let $D\Psi(\boldsymbol{\rho}), D\Psi_g(\boldsymbol{\rho})$ be the Jacobian of Ψ, Ψ_g, respectively. Then the eigenvalues of $D\Psi_g(\boldsymbol{\rho}^g)$ converge to the eigenvalues of $D\Psi(\boldsymbol{\rho}^*)$ uniformly over $g \in \mathcal{B}^1_\gamma$ as $\gamma \to 0$. Thus, for small enough γ, $\boldsymbol{\rho}^g$ has the same stability properties as $\boldsymbol{\rho}^*$.*

Proof. I will show that eigenvalues of a hyperbolic matrix $D\Psi(\boldsymbol{\rho}^*)$ vary continuously in $\mathcal{C}^1$ perturbations g to Ψ.

Palis Jr, Melo, et al. (1982, Proposition 2.18) shows that eigenvalues vary continuously for any matrix A. Thus, if $\|D\Psi(\boldsymbol{\rho}^*) - D\Psi_g(\boldsymbol{\rho}^g)\|$ is small enough, the eigenvalues of the two matrices must be close to each other. Now write

$$\begin{aligned} \|D\Psi(\boldsymbol{\rho}^*) - D\Psi_g(\boldsymbol{\rho}^g)\| &= \|D\Psi(\boldsymbol{\rho}^*) - D\Psi(\boldsymbol{\rho}^g)\| + \|Dg(\boldsymbol{\rho}^g)\| \\ &\leq \|D\Psi(\boldsymbol{\rho}^*) - D\Psi(\boldsymbol{\rho}^g)\| + \gamma, \end{aligned}$$

where the equality follows from the definition of Ψ_g. Since $D\Psi$ is continuous, and $\boldsymbol{\rho}^g \to \boldsymbol{\rho}^*$ uniformly for $g \in \mathcal{B}^1_\gamma$ as $\gamma \to 0$ (see above proof of point 2), we get that

$$\sup_{g \in \mathcal{B}^1_\gamma} \|D\Psi(\boldsymbol{\rho}^*) - D\Psi_g(\boldsymbol{\rho}^g)\| \to 0$$

as $\gamma \to 0$. Then applying Palis Jr, Melo, et al. (1982, Proposition 2.18) finishes the result. □

Since we know that all $\boldsymbol{\rho}^g$ must be asymptotically stable for γ small enough, one can apply Faure and Roth (2010, Theorem 2.15) . To prove convergence to an attractor $\{\boldsymbol{\rho}^g\}$ with positive probability, a stronger result than Proposition 1 is first needed:

Assumption 3 (Condition 11, Faure and Roth (2010))**.** *There exists a map* $\omega : \mathbb{R}^3_+ \to \mathbb{R}_+$ *such that*

(1) For any $\varepsilon > 0$, $T > 0$,

$$\mathbb{P}\left(\sup_{m' \geq n} \sup_{m' \leq k \leq m(\tau_{m'}+T)} \left\| \sum_{i=n}^{k-1} \alpha_{i+1}\Big(\delta_{i+1} + M_{i+2}\Big) \right\| > \varepsilon \;\middle|\; \mathcal{F}_n \right) \leq \omega(n, \varepsilon, T),$$

almost surely in $\mathcal{F}_0$*.*

(2) $\lim_{n\to\infty} \omega(n, \varepsilon, T) = 0$.

Faure and Roth (2010, Proposition 2.16) states that Condition 11 above is satisfied for our M_{n+1} martingale difference sequence (i.e. if $\delta_n = 0$ for all n). I show next that this result extends to our case of (2):

Lemma 2. *Suppose* δ_n, M_n *satisfy Assumption 1 (1),(2),(4). Then condition* 11 *is satisfied.*

Proof. Note first that

$$\begin{aligned} &\left\| \sum_{i=n}^{k-1} \alpha_{i+1}\Big(\delta_{i+1} + M_{i+2}\Big) \right\| \\ &\leq \left\| \sum_{i=n}^{k-1} \alpha_{i+1}\Big(M_{i+2}\Big) \right\| + \left\| \sum_{i=n}^{k-1} \alpha_{i+1}\Big(\delta_{i+1}\Big) \right\| \\ &= R_n + \Psi_n^k, \end{aligned}$$

similarly as stated in the proof above. For R_n, Proposition 2.16 in Faure and Roth (2010) immediately applies, as it only requires 1 (i) on α_n, and (4) is satisfied for M_n. The remaining term Ψ_n^k can be treated analogously to the proof of Proposition 1. □

Finally, Faure and Roth (2010, Theorem 2.15) states that if condition 11 is satisfied, $\mathbb{P}[L_{S,g} = \{\boldsymbol{\rho}^g\}] > 0$ holds as long as $\{\boldsymbol{\rho}^g\}$ is *attainable* by the process $\boldsymbol{\rho}_n$. This can be verified analogously to the approach in the proof of Theorem 1 of the main text. Thus, Faure and Roth (2010, Theorem 2.15) applies, concluding this proof. □

Theorem 2

The following generalizes Theorem 2 of the main text:

Theorem 2. *Let $\boldsymbol{\rho}^* \in E_S^*$ be linearly unstable for Ψ_S. Then for all γ small enough and all $g \in \mathcal{B}_\gamma^1$ there is an open neighborhood U_γ with $\boldsymbol{\rho}^* \in U_\gamma$ such that*

$$\mathbb{P}[L_{S,g} \subseteq U_\gamma] = 0.$$

Proof. The proof will use the Hartman-Grobman Theorem (c.f.Chicone (2006, Theorem 4.8)), which connects the flow of a nonlinear ODE in the neighborhood of a hyperbolic rest point to the flow of a linearized ODE. Since it works fully locally, our analysis only requires that $\Psi(\boldsymbol{\rho})$ be single valued and $\mathcal{C}^1$ in a neighborhood of rest point $\boldsymbol{\rho}^*$, and we can allow $\Psi(\boldsymbol{\rho})$ to be multivalued otherwise. Call this neighborhood $U_{\boldsymbol{\rho}^*}$.

First, define invariant sets for given differential equations:

Definition 6. *Let $z(t, z_0)$ be the solution to some given differential equation $\dot{z} = f(z)$ with initial value z_0. Then a set S*

- *is invariant for f, if $z(t, z_0) \in \mathbf{S}$ holds for all $t \in \mathbb{R}$ and all $z_0 \in \mathbf{S}$.*
- *isolated invariant for f if there is an open set N such that $S \subset N$ and*

$$S = \{z' : z(t, z') \in N \, \forall t \in \mathbb{R}\}.$$

Given a $g \in \mathcal{B}_\gamma^1$, we know from Proposition 1 that only ICT sets (recall Definition 5) subset of a neighborhood of $\boldsymbol{\rho}^g$ are candidates to being limiting points of the algorithm (2). The singleton $\{\boldsymbol{\rho}^g\}$ is an ICT set, and we show first that this is a limiting set of the algorithm with probability zero. Then we go on to show that for small enough γ, no other ICT sets can exist in a neighborhood around $\boldsymbol{\rho}^*$, which finishes the proof.

1) $\{\boldsymbol{\rho}^g\}$ is a limiting set of (2) with probability zero.

Note that by Lemma 1, there are $\gamma > 0$ small enough such that all $\boldsymbol{\rho}^g$ for $g \in \mathcal{B}_\gamma^1$ are linearly unstable, just as $\boldsymbol{\rho}^*$. We can thus apply Benaïm and Faure (2012, Theorem 3.12) to prove $\mathbb{P}[L_{S,g} = \{\boldsymbol{\rho}^g\}] = 0$ in the following. Importantly, note that the conditions and analysis sufficient for the proof of Benaïm and Faure (2012)'s Theorem are local with respect to $\boldsymbol{\rho}^g$. Thus, the fact that Ψ_g is globally potentially multivalued is of no importance, since in a small enough neighborhood around $\boldsymbol{\rho}^g$ it must be single-valued and $\mathcal{C}^1$.

Benaïm and Faure's result is concerned with time-interpolations of iterations such as (2). Their Theorem 3.12 states, translated in terms of this paper, that under an Assumption the authors refer to as Hypothesis 2.2, and Assumption 1 (4), (5), the result to be proved here holds true.

In fact, Benaïm and Faure (2012, Hypothesis 2.2) is equivalent[2] to Assumption 3, which was shown to hold for our algorithm in Lemma 2. Thus, the result applies, concluding the proof.

2) No other ICT sets exist in a neighborhood of $\boldsymbol{\rho}^*$ and $\boldsymbol{\rho}^g$.

We will prove that there are no other invariant sets in such a neighborhood. Since ICT sets are subsets of invariant sets, this will complete the proof.

We can use Hartman-Grobman to show that there are open neighborhoods N_g, N_0 with $\boldsymbol{\rho}^* \in N_0, \boldsymbol{\rho}^g \in N_g$ such that $\boldsymbol{\rho}^*, \boldsymbol{\rho}^g$ are isolated invariant sets in their respective neighborhoods. These neighborhoods are nontrivial for all γ small enough, which follows from both $\boldsymbol{\rho}^*, \boldsymbol{\rho}^g$ being hyperbolic:

By Hartman-Grobman and hyperbolicity there exists a homeomorphism H on a neighborhood $N \subseteq U_{\boldsymbol{\rho}^*}$ of $\boldsymbol{\rho}^*$ with $H(\boldsymbol{\rho}^*) = \boldsymbol{\rho}^*$ such that

$$H(\phi(t, \boldsymbol{\rho})) = \psi(t, H(\boldsymbol{\rho})),$$

where $\phi(t, \cdot)$ is a solution (flow) to the differential inclusion $\dot{\boldsymbol{\rho}} \in \Psi(\boldsymbol{\rho})$, and $\psi(t, \cdot)$ is the solution to the ODE $\dot{y} = D\Psi(\boldsymbol{\rho}^*)(y - \boldsymbol{\rho}^*)$. Given a neighborhood $U \subseteq N$ of $\boldsymbol{\rho}^*$, define

$$inv(U) = \{\boldsymbol{\rho} \in U : \phi(t, \boldsymbol{\rho}) \in U \, \forall t \in \mathbb{R}\}.$$

We will show that $\{\boldsymbol{\rho}^*\} = inv(U)$, and therefore, it is isolated invariant.
Notice that $inv(U)$ can be rewritten as

$$inv(U) = \{y \in H(U) : H^{-1}(\psi(t, y)) \in U \, \forall t \in \mathbb{R}\} = \{y \in H(U) : \psi(t, y) \in H(U) \, \forall t \in \mathbb{R}\},$$

[2] See Faure and Roth (2010, Remark 2.14)

since H is bijective. We know that $\boldsymbol{\rho}^*$ is an isolated invariant set for the linear ODE solution $\psi(t,y) = Ce^{tD\Psi(\boldsymbol{\rho}^*)}y + \boldsymbol{\rho}^*$. Thus, we must also have that

$$inv(U) = \boldsymbol{\rho}^*,$$

and $\{\boldsymbol{\rho}^*\}$ is an isolated invariant set for $\phi(t, \boldsymbol{\rho})$.

Since $\boldsymbol{\rho}^g$ are hyperbolic for γ small enough, an analogous argument gives us that $\boldsymbol{\rho}^g$ are isolated invariant also. Let N_g be the neighborhood on which the homeomorphism is defined that connects flows of Ψ_g to flows of the linearized system $D\Psi_g(\boldsymbol{\rho}^g)$. By definition, $\boldsymbol{\rho}^g \in N_g$, and we know that $\boldsymbol{\rho}^g$ is isolated invariant in N_g. We are left to show that for γ small enough, for all $g \in \mathcal{B}^1_\gamma$, $\boldsymbol{\rho}^* \in N_g$:

To prove this, we will argue that each N_g contains a ball $B^g_z(\boldsymbol{\rho}^g)$, for which the radius $z > 0$ can be lower bounded by a number that depends only on the eigenvalues of $D\Psi(\boldsymbol{\rho}^*)$ and γ. First, we need an auxiliary Lemma to show how eigenvalues of $D\Psi_g(\boldsymbol{\rho}^g)$ vary continuously in γ. First, some more notation:

For small enough γ, all $\boldsymbol{\rho}^g$ are hyperbolic when $g \in \mathcal{B}^1_\gamma$. Fix such a g. Define $\boldsymbol{\rho}_l > 0$ to be the smallest positive eigenvalue of $D\Psi_g(\boldsymbol{\rho}^g)$, and $\boldsymbol{\rho}_u < 0$ be the largest negative eigenvalue of $D\Psi_g(\boldsymbol{\rho}^g)$. Now let $a_g \in (0,1)$ be any number such that

$$\max\{e^{\boldsymbol{\rho}_u}, e^{-\boldsymbol{\rho}_l}\} < a_g < 1.$$

For the original system $D\Psi(\boldsymbol{\rho}^*)$, let $a_0 \in (0,1)$ be any such number.

Lemma 3. *For any $\delta > 0$ with $a_0 < 1 - \delta$ there exists $\overline{\gamma} > 0$ such that for all $\gamma \in (0, \overline{\gamma}]$, there is a set of $\{a_g\}_{g \in \mathcal{B}^1_\gamma}$ as defined above with*

$$\sup_{g \in \mathcal{B}^1_\gamma} |a_g - a_0| < \delta.$$

Proof. Apply Lemma 1. Since there is a one-to-one mapping between eigenvalues and $\{e^{\boldsymbol{\rho}_u}, e^{-\boldsymbol{\rho}_l}\}$, one can find numbers a_g. The result follows. □

Given this continuity in eigenvalues, we can prove the following Lemma to finish our result:

Lemma 4. *Suppose $\boldsymbol{\rho}^*$ is hyperbolic for Ψ. Fix a small $\underline{z} > 0$. Then there is $\bar{\gamma}$ such that for all $\gamma \leq \bar{\gamma}$, and all $g \in \mathcal{B}^1_\gamma$, there is $B^g_z(\boldsymbol{\rho}^g) \subseteq N_g$ with $z \geq \underline{z}$.*

Proof. For small enough γ, all $\boldsymbol{\rho}^g$ are hyperbolic when $g \in \mathcal{B}^1_\gamma$. Fix such a g. Given some $\varepsilon > 0$, let r_ε be defined as

$$\sup\{r > 0 : \|\boldsymbol{\rho} - \boldsymbol{\rho}^g\| < r;\ \|D\Psi_g(\boldsymbol{\rho}) - D\Psi_g(\boldsymbol{\rho}^g)\| < \varepsilon\}.$$

Since $D\Psi_g$ is continuous, $r_\varepsilon > 0$ must hold. Pick $a_g \in (0, 1)$ as defined previously. Then define

$$\overline{\varepsilon}_g = \frac{1 - a_g}{a_g} > 0.$$

By Palis Jr, Melo, et al. (1982, Lemmas 4.3 and 4.4), $B_{r_\varepsilon}(\boldsymbol{\rho}^g) \subseteq N_g$, if $\varepsilon < \overline{\varepsilon}_g$.

We are left to show that r_ε can be made to depend only on the eigenvalues of $D\Psi(\boldsymbol{\rho}^*)$ and γ. Notice that small enough $\underline{z} > 0$ pins down the $\delta > 0$ referred to in Lemma 3: Let

$$\hat{z}(\overline{\gamma}) = \inf_{\gamma \in (0, \overline{\gamma}]} \inf_{g \in \mathcal{B}^1_\gamma} \overline{\varepsilon}_g.$$

For $\delta > 0$ small enough, choose $\overline{\gamma} > 0$ such that Lemma 3 holds. It follows from the Lemma that $\hat{z}(\overline{\gamma}) > 0$. Then any $\underline{z} < \hat{z}(\overline{\gamma})$ satisfies our conditions and the conclusion follows. □

Now recall that by the proof of Theorem 1 point 2, $\boldsymbol{\rho}^g \to \boldsymbol{\rho}^*$ uniformly over $g \in \mathcal{B}^1_\gamma$ as $\gamma \to 0$. Thus, there is γ small enough for which $\sup_{g \in \mathcal{B}^1_\gamma} |\boldsymbol{\rho}^g - \boldsymbol{\rho}^*| < \underline{z}$ and therefore $\boldsymbol{\rho}^* \in N_g$ for all $g \in \mathcal{B}^1_\gamma$. Let $U_\gamma = \cap_{g \in \mathcal{B}^1_\gamma} N_g$. Since $\boldsymbol{\rho}^g$ for $g \in \mathcal{B}^1_\gamma$ are isolated invariant in U_γ by construction, the result follows. □

III. Numerical Example and Simulations

I continue here the numerical example outlined in the main text. Consider stage game with linear demand as discussed in the example in section 3.C of the main text, where $D = 20, b = 1, \gamma = 1/2$, and $c = 2$. One can verify that in this model, $p_N = \frac{D-c}{2b-\gamma} = 12$. Suppose that agents discount time with $\delta = 0.9$. The agents are ACQ learners and commonly observe a binary state variable with $O_S = \{A, B\}$, where transition probability functions take a logit form. Specifically,

$$T_{AB}(P) = \left(1 + \exp\left(k_A\left(P - d_A\right)\right)\right)^{-1},$$
$$T_{BA}(P) = \left(1 + \exp\left(k_B\left(P - d_B\right)\right)\right)^{-1}, \tag{7}$$

where $k_A = 1.02, d_A = 24.91, k_B = 1.38, d_B = 23.94$. I verify numerically that this set of transition functions supports two symmetric collusive equilibria $\boldsymbol{\rho}^{*(1)}, \boldsymbol{\rho}^{*(2)}$ where $\boldsymbol{\rho}^{*(1)}(A) \approx 13.05 > p_N > \boldsymbol{\rho}^{*(1)}(B) \approx 11.21$, and $\boldsymbol{\rho}^{*(2)}(A) \approx 11.54 < p_N < \boldsymbol{\rho}^{*(2)}(B) \approx 12.64$. Hence, one state realization supports a collusive state of high prices, and the other induces a punishment state needed to incentive the collusive price.[3]

As noted in the main text, this scenario satisfies that static Nash will not be learnt, while either of the two collusive euilibria may be learnt under ACQ and ACG learning. I show now the results of simulation studies that are even stronger than the statements proved in the main text. Below, I verify in simulations that not only do ACQ, ACG learners globally converge to $\boldsymbol{\rho}^{*(1)}$ or $\boldsymbol{\rho}^{*(2)}$, but also that settings where ACQ and ACG learners are in competition (the asymmetric critics case) and settings where learners have asymptotically different learning rates (asymmetric rates case) satisfy this strong convergence property.

Based on notation from the main text, I simulate the reduced form algorithms, following Assumption 1. Let Ψ be an admissible critic profile. Fix a binary state variable S.

For $i \in \{1, 2\}$ and all $s \in O_S$,

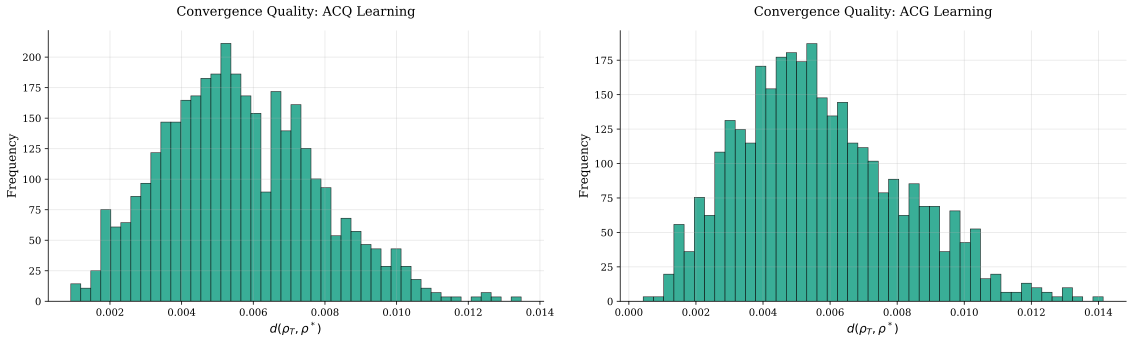
$$\boldsymbol{\rho}^i_{t+1}(s) \in \boldsymbol{\rho}^i_t(s) + \alpha^i_t \Big[\Psi^i(\boldsymbol{\rho}_t) + M^i_{t+1}\Big], \tag{8}$$

where $M^i_{t+1} \sim N(0, .1)$ is an i.i.d mean-zero Normal noise variable with variance 0.1, and $\Psi(\boldsymbol{\rho}_t)$ is computed exactly. The default stepsize is $\alpha^i_t = (t + 1)^{-2/3}$ for all i.

In each simulation exercise, I run 1,000 separate simulations, and each for 10^6 periods. The algorithms are initialized randomly, where $\boldsymbol{\rho}^i(s)_0 \sim U[0, 2p_N]$ for all i and s. As will be seen, depending on the state variables of the algorithms involved, iterations move closer to the equilibrium in the neighborhood of which they started at, or move away from it, confirming the theory developed in this paper.

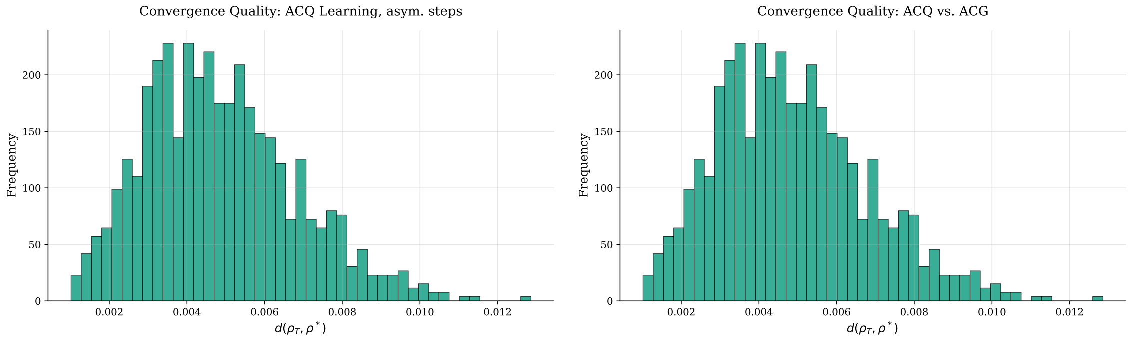
First, I report the last iterates of simulation runs where both agents are ACQ or both are ACG learners. To this end, define $d(\boldsymbol{\rho}, \boldsymbol{\rho}^*) = \min\{\|\boldsymbol{\rho} - \boldsymbol{\rho}^{*(1)}\|, \|\boldsymbol{\rho} - \boldsymbol{\rho}^{*(2)}\|\}$ as the smallest distance between policy profile $\boldsymbol{\rho}$ and either collusive equilibrium.

[3] Note that as both T_{AB}, T_{BA} are decreasing in P, this is intuitive: e.g. for $\boldsymbol{\rho}^{*(1)}$, at high prices in state A, optimal one-shot deviations ask for reduction in price, which leads to an increase in probability of the punishment state. On the other hand, in the punishment state B, optimal one-shot deviations require increases in price, which leads to an increase in the probability of remaining in B.



(A) (B)

FIGURE 1. In (A), 52.3 % of simulations converged to $\boldsymbol{\rho}^{*(1)}$, and in (B), 46.2 % did.



(A) (B)

FIGURE 2. In (A), 49.2 % of simulations converged to $\boldsymbol{\rho}^{*(1)}$, and in (B), 42.6 % did.

Figure 1 shows how indeed, all simulation runs converge to one of the two collusive equilibria, and never converge to the static Nash equilibrium. Next, Figure 2 robustifies the theoretical results stated in Section 4 of the main text. First I consider a setting of asymmetric stepsizes, where $\alpha_t^1 = (t+1)^{-2/3}$ as before, but $\alpha_t^2 = (\frac{1}{2} - (t+1)^{-2})(t+1)^{-2/3}$. Hence, $\lim_{t\to\infty} \frac{\alpha_t^2}{\alpha_t^1} = \frac{1}{2}$. Second, I consider the asymmetric critic setting where stepsizes are equal, but player 1 follows ACQ, and player 2 follows ACG learning. As can be seen, the global convergence property also holds in these simulation runs, further strengthening the robustness properties of the analytical results given in the main text.

Finally, I consider Proposition 1 in the main text and apply the ideas of covariate restrictions to my numerical example. I take $\beta \in [0, 1]$ to mix transitions defined in (7) with uninformative noise. Specifically, let

$$\tilde{T}_{AB}(P;\beta) = T_{AB}(P)(1-\beta) + \beta\frac{1}{2},$$
$$\tilde{T}_{BA}(P;\beta) = T_{BA}(P)(1-\beta) + \beta\frac{1}{2}.$$

Hence, with probability $1-\beta$, transitions follow the functions sensitive to P defined in (7), but with complementary probability, price has no impact on transitions and hence no information is carried by transitions. Clearly, when $\beta = 0$, we are back at the original regime, while $\beta = 1$ corresponds to the case where policies cannot condition on any past information about price choices, and the only equilibrium that exists is $\boldsymbol{\rho}_N$.

I consider how learning outcomes under ACQ or ACG learners change as I vary β. To this end, for a range of $\beta \in [0,1]$, I run an extensive equilibrium search for each β, compute the payoff-maximal equilibrium, and check whether this equilibrium, and / or static Nash, is attracting under ACQ and ACG learning. Figure 3 shows the result. I plot the value ratio $W(\boldsymbol{\rho}^*;\beta)/W(\boldsymbol{\rho}_N;\beta)$, the ratio of best equilibrium payoff given β relative to static Nash payoff. As indicated by Proposition 1 in the main text, the ratio falls as β increases. For a value of $\bar{\beta} \approx 0.42$, the best equilibrium converges to static Nash (in fact, for values of β above $\bar{\beta}$, only static Nash is found as equilibrium). Furthermore, whenever collusive equilibria exist (i.e. $\beta < \bar{\beta}$), they are attracting under both ACQ and ACG learners, while static Nash is unstable. The opposite holds for $\beta > \bar{\beta}$. This further strengthens the insights of Proposition 1: restricting the monitoring ability of the state variable (increasing β) will both decrease the value of the most concerning collusive equilibrium that may be learnt, and also at one point ensure that no collusive equilibria remain to be learnt, while static Nash is learnt with positive probability.

The example and simulations were generated using the python programming language and packages provided in Van Rossum, Drake, et al. 1995, Harris et al. 2020,Virtanen et al. 2020, and Hunter 2007.

IV. Approximating the best equilibrium

I provide here a result connecting the binary strategies considered in section 3.B of the main text to payoff maximal equilibria of an associated finite action game. APS provide a

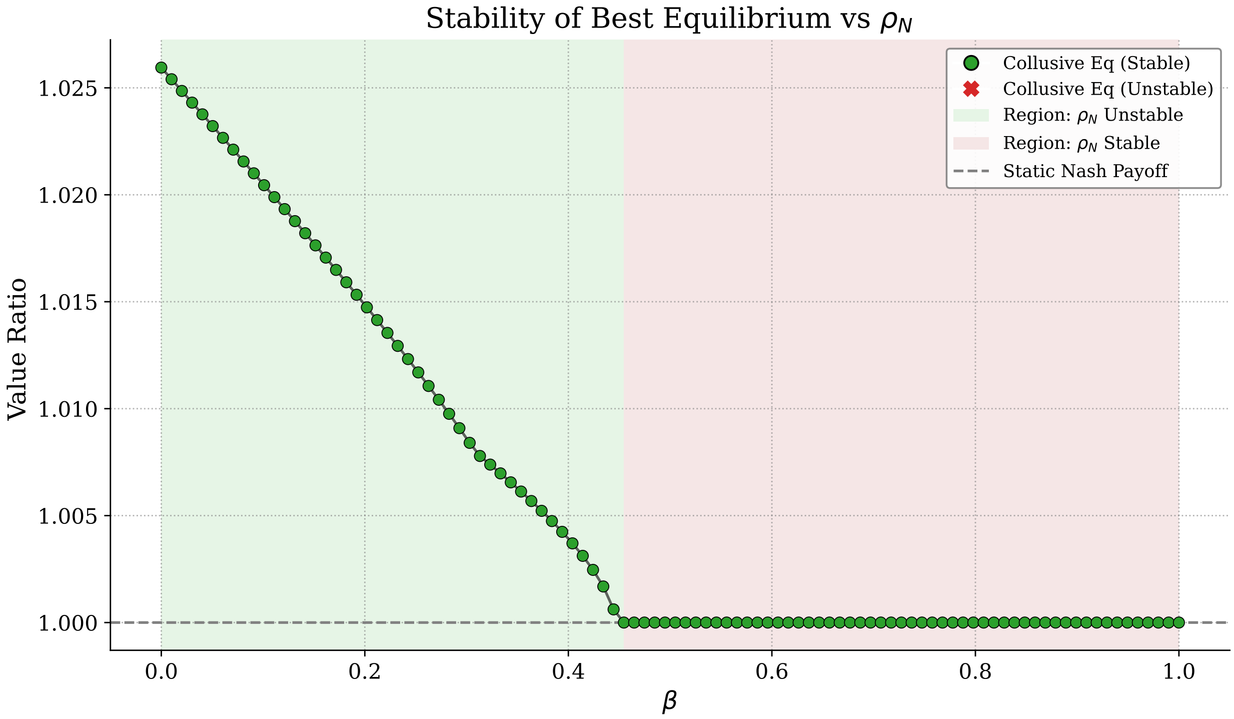


FIGURE 3.

result stating that the best strongly symmetric sequential equilibrium (SSE) of the repeated game can be supported by a bang-bang solution, under their setting.

Such a bang-bang strategy, by definition, is constructed using subsets of Ω, intended as punishment and reward regions. Translated into this paper's setup, there exists a state variable S^* with $O^*_S = \{A, B\}$ and sets Ω_A, Ω_B such that $s = A$ and $\tilde{A} \in \Omega_A$ implies the next period's state is A, and $s = B$ and $\tilde{A} \in \Omega_B$ implies next period's state is B. The reverse holds for $\Omega \setminus \Omega_s$.

Notice that any binary partition of Ω affects payoffs of players only by pinning down their transition probability functions $T_{ss'}(P)$. As $g(a; p)$ is twice differentiable in p by Assumption 3 in the main text, one can restrict attention to twice continuously differenetiable $T_{ss'}$. Call the space of such transition probability functions $\mathcal{T}$. For any $T_{ss'} \in \mathcal{T}$, let $E^*(T_{ss'})$ be the set of symmetric Nash equilibria given expected discounted payoffs $W(\rho)$ when $T_{ss'}$ governs state transitions. This set is nonempty for all $T_{ss'} \in \mathcal{T}$ due to the repetition of the static Nash equilibrium p_N.

As APS' result was shown only for finite strategy sets, I introduce an approximation result to the continuous action case. Let Y_L be a discretization of cardinality $0 < L < \infty$ of price-set Y, where for any discretization I impose that $p_N \in Y_L$.

Define $W(\boldsymbol{\rho}, T)$ for symmetric profiles $\boldsymbol{\rho} \in \overline{Y}$, transition probabilities $T \in \mathcal{T}$ as long run expected payoffs. Define $E_L(T)$ as the set of symmetric equilibria given discretization Y_L. Here, APS's bang-bang result directly applies. By the observation above, I can alternatively characterise the maximal SSE as

$$V_L = \sup_{\substack{\boldsymbol{\rho} \in E_L(T) \\ T \in \mathcal{T}}} W(\boldsymbol{\rho}, T).$$

Analogously, define

$$V = \sup_{\substack{\boldsymbol{\rho} \in E^*(T) \\ T \in \mathcal{T}}} W(\boldsymbol{\rho}, T). \tag{9}$$

Define V^* to be the best SSE payoff among all SSE of Γ^∞.

Proposition 2. *Given additional regularity conditions on W, $\mathcal{T}$[4], there exists an SSE $\boldsymbol{\rho}$ of Γ^∞ supported by a binary-state policy, under some $T^* \in \mathcal{T}$ such that $V = W(\boldsymbol{\rho}, T^*)$. It holds that*

(1) $V \leq V^*$.

(2) For any ε there exists $\bar{L}$ such that for all $L \geq \bar{L}$, $|V - V_L| < \varepsilon$.

Proposition 2 indicates that there exist binary state variables such that if used by algorithms, they may learn to play strategies that achieve the best SSE payoff for any discretization of their game. For any interior $\boldsymbol{\rho} \in E^*(T_{ss'})$, let $\boldsymbol{J}(\boldsymbol{\rho})$ be the 2×2 matrix of best response derivatives, i.e. the Jacobian of the best response function at the equilibrium $\boldsymbol{\rho}$, $B^1_S(\boldsymbol{\rho}_2)$. I require the following additional assumption:

Assumption 4. *For all $T_{ss'} \in \mathcal{T}$, all $\boldsymbol{\rho} \in E^*(T_{ss'})$ are interior, with negative definite Hessian, and all eigenvalues of $\boldsymbol{J}(\boldsymbol{\rho})$ are different from* 1.

For any fixed $T_{ss'}$, the above assumption is a generic property over the space of regular payoff functions. The assumption has additional strength as it imposes that given the regular

[4] These stronger conditions ensure continuity of the equilibrium correspondence, see Assumption 4

stage game payoff function π, there exists no $T_{ss'} \in \mathcal{T}$ that could lead to a singular Hessian at some equilibrium, or a $\boldsymbol{J}(\boldsymbol{\rho})$ with eigenvalue equal to 1. For any discretization Y_L, define $W^L(\boldsymbol{\rho}, T) : Y^2 \times \mathcal{T} \to \mathbb{R}$ as restriction of the payoff function to Y_L, given some T.

$$W^L(\boldsymbol{\rho}, T) = W(f^L(\boldsymbol{\rho}), T),$$

where

$$f^L(\boldsymbol{\rho}) = \arg \min_{\boldsymbol{\rho}' \in Y_L^2} \left\| \boldsymbol{\rho} - \boldsymbol{\rho}' \right\|,$$

for any norm on Y^2, the projection of $\boldsymbol{\rho}$ onto discrete space Y_L. For every sequence Y_L there is an associated sequence $\alpha_L(T)$ with

$$\alpha_L(T) = \max_{\boldsymbol{\rho} \in Y^2} \left\| W^L(\boldsymbol{\rho}, T) - W(\boldsymbol{\rho}, T) \right\|.$$

Continuity of W and the Lipschitz property of density $g(a;p)$ implies that $\alpha_L(T) \to 0$ for all $T \in \mathcal{T}$. Write $\alpha_L(Y_L, T)$ for a sequence of α_L given a fixed sequence of discretizations and transition function T. Say that a discretization sequence Y_L is *covering* if $\alpha_L(Y_L, T) \to 0$ (and $p_N \in Y_L$). From now on fix a covering sequence of discretizations Y_L and transition probability T.

Notice that $E_L(T)$ is closed-valued, trivially by finiteness of Y_L. Furthermore, $E^*(T)$ is closed-valued: W is continuous, Y compact, and thus Berge gives us that the best response correspondence is closed and compact-valued. Then, applying the closed-graph theorem gives us that the equilibrium set $E^*(Y)$, as a set of fixed points of a closed and compact correspondence, must be closed. To get to claim (1), I will show that any converging sequence $\boldsymbol{\rho}_L \in E_L(T)$ has its limit in $E^*(T)$, and any $\boldsymbol{\rho} \in E^*(T)$ has a converging sequence in $E_L(T)$ approaching it. In other words, upper and lower hemicontinuity properties hold for the equilibrium correspondence along sequences of covering discretizations.

Lemma 5. *For all covering sequences Y_L,*

$$\lim_{K \to \infty} H\left(E_L(T), E^*(T)\right) = 0,$$

where $H(\cdot, \cdot)$ is the Hausdorff-distance.

Proof. I first show upper hemicontinuity in K. Suppose u.h.c. is not satisfied. Then there exists a subsequence $\boldsymbol{\rho}_{L_t} \in E_{L_t}(T)$ with $\boldsymbol{\rho}_{L_t} \to_t \bar{\boldsymbol{\rho}} \notin E^*(T)$. The converging subsequence exists since Y^2 is compact. To ease notation, re-define $L = L_t$ for the rest of the proof. Not being an equilibrium, I have that there exists $\boldsymbol{\rho}_T \neq \bar{\boldsymbol{\rho}}$ that maximizes the deviation payoff

$$\Delta_T = W(\boldsymbol{\rho}_{T,i}, \bar{\boldsymbol{\rho}}_{-i}, T) - W(\bar{\boldsymbol{\rho}}, T) > 0.$$

Pick $\varepsilon \in (0, \Delta_T)$. By convergence of $\boldsymbol{\rho}_L$, and by continuity of W, I have that there exists $L_{1,T}$ such that for all $L \geq L_{1,T}$,

$$\left| W(\boldsymbol{\rho}_{T,i}, \boldsymbol{\rho}_{L,-i}, T) - W(\boldsymbol{\rho}_{T,i}, \bar{\boldsymbol{\rho}}_{-i}, T) \right| \leq \frac{\varepsilon}{3}. \tag{10}$$

By the same argument, there is a $L_{2,T}$ s.t. for all $L \geq L_{2,T}$,

$$\left| W(\boldsymbol{\rho}_L, T) - W(\bar{\boldsymbol{\rho}}, T) \right| \leq \frac{\varepsilon}{3}. \tag{11}$$

Furthermore, one can always choose $\bar{L}_T \geq \max\{L_{1,T}, L_{2,T}\}$ large enough so that $\alpha_L \leq \frac{\varepsilon}{3}$, implying

$$\left| W(f^L(\boldsymbol{\rho}_{T,i}), \boldsymbol{\rho}_{L,-i}, T) - W(\boldsymbol{\rho}_{T,i}, \boldsymbol{\rho}_{L,-i}, T) \right| \leq \frac{\varepsilon}{3}. \tag{12}$$

Take $L \geq \bar{L}_T$. Define the best deviation under the discrete game as

$$\hat{\boldsymbol{\rho}_{L,i}} = \arg \max_{\boldsymbol{\rho}_i \in Y_L^2 \setminus \boldsymbol{\rho}_L} W(\boldsymbol{\rho}_i, \boldsymbol{\rho}_{L,-i}, T).$$

Now

$$\begin{aligned}
W(\hat{\boldsymbol{\rho}_{L,i}}, \boldsymbol{\rho}_{L,-i}, T) - W(\boldsymbol{\rho}_L, T) &\geq W(f^L(\boldsymbol{\rho}_{T,i}), \boldsymbol{\rho}_{L,-i}, T) - W(\boldsymbol{\rho}_L, T) \\
&= W(\boldsymbol{\rho}_{T,i}, \boldsymbol{\rho}_{L,-i}, T) - W(\boldsymbol{\rho}_L, T) + \beta_{1,L} \\
&= W(\boldsymbol{\rho}_{T,i}, \bar{\boldsymbol{\rho}_{-i}}, T) - W(\bar{\boldsymbol{\rho}}, T) + \beta_{1,L} + \beta_{2,L} + \beta_{3,L} \\
&\geq \Delta_T + \beta_{1,L} + \beta_{2,L} + \beta_{3,L},
\end{aligned}$$

where $\beta_{1,L}$ corresponds to the projection error (12), and $\beta_{2,L}, \beta_{3,L}$ correspond to (10),(11) respectively. Note that $|\beta_{i,L}| \leq \frac{\varepsilon}{3}$, and thus

$$W(\hat{\boldsymbol{\rho}}_{L,i}, \boldsymbol{\rho}_{L,\text{-}i}, T) - W(\boldsymbol{\rho}_L, T) \geq \Delta_T - \varepsilon > 0,$$

implying that $\boldsymbol{\rho}_L \notin E_L(T)$, a contradiction.

For lower hemicontinuity, Assumption 4 imposes that for all equilibria in $E^*(T)$ for all players, Hessians at the equilibrium are negative definite. Thus, small deviations must lead to strict payoff loss. Fix T, then pick any equilibrium $\boldsymbol{\rho}^* \in E^*(T)$. Analogously to the one above, define $\Delta_T > 0$ as the best deviation payoff:

$$\Delta_T = W(\boldsymbol{\rho}^*, T) - \sup_{\boldsymbol{\rho}_i \in Y \setminus \boldsymbol{\rho}} W(\boldsymbol{\rho}_i, \boldsymbol{\rho}^*_{\text{-}i}, T) > 0.$$

Since $\Delta_T > 0$ and $\boldsymbol{\rho}^*$ is an equilibrium in the unconstrained game, I can find a fine enough discretization s.t. $\boldsymbol{\rho}^*$ can be approximated arbitrarily closely, in which case incentives must also align, by continuity of W. The result follows. □

Continuity of the equilibrium correspondence gives us that for all $\varepsilon > 0$ there is $K > 0$ large enough so that

$$\|V_L(T) - V^*(T)\| < \varepsilon,$$

with $V_L(T), V^*(T)$ being the maximal payoff over the equilibrium sets $E_L(T), E^*(T)$.

To make judgements about $\sup_{T \in \mathcal{T}} V_L(T)$, a uniform continuity property of $V_L(T)$ will be useful. By Assumption 4, all equilibria in E_L and E^* are hyperbolic, for K large enough. Hyperbolicity implies that an implicit function theorem holds: For any $\boldsymbol{\rho} \in E_L$, there exists neighborhoods $\mathcal{N}_{\boldsymbol{\rho}}, \mathcal{N}_T$ of $\boldsymbol{\rho}, T$ and a continuous map $h : \mathcal{N}_T \to \mathcal{N}_{\boldsymbol{\rho}}$ such that $h(T) \in E_L(T)$ for all $T \in \mathcal{N}_T$. Thus, the equilibrium correspondences $E_L(T), E^*(T)$ are continuous in T for all K large enough.

As $W(\boldsymbol{\rho}, T)$ is continuous both in $\boldsymbol{\rho}$ and T[5], and equilibrium correspondences are continuous in T, Berge's Maximum Theorem applies. So, $V_L(T)$ is continuous in T. Moreover, as the payoff functions $W(\boldsymbol{\rho}, T)$ are bounded, twice differentiable in $\boldsymbol{\rho}$ and transition probabilities T (evaluated at $\boldsymbol{\rho}$), these payoff functions are Lipschitz both in $\boldsymbol{\rho}$ and T. Then,

[5] Regarding functions T, consider the sup-norm as metric on $\mathcal{T}$.

$V_L(T)$ are bounded, Lipschitz as well for K large enough. This follows first from continuity of $E^*(T)$, and then by the Lipschitz property of $W(\boldsymbol{\rho}, T)$. Finally, boundedness and the Lipschitz property imply that $V_L(T)$ are equicontinuous in T.

By the Arzelá-Ascoli Theorem, boundedness and equicontinuity of $V_L(T)$ implies the existence of a subsequence K_m so that V_{K_m} converges uniformly to some V. Pointwise convergence of $V_L(T)$ implies that this limit satsifies $V = V^*$. A simple contradiction argument with another application of Arzelá-Ascoli shows that indeed $V_L(T)$ converges uniformly to V^*. It follows that

$$\lim_{K\to\infty} \sup_{T\in\mathcal{T}} V_L(T) = \sup_{T\in\mathcal{T}} V^*(T).$$

■

References

Benaïm, Michel and Mathieu Faure (2012). "Stochastic approximation, cooperative dynamics and supermodular games". In: *The Annals of Applied Probability* 22.5, pp. 2133–2164.

Borkar, Vivek S (2009). *Stochastic approximation: a dynamical systems viewpoint.* Vol. 48. Springer.

Chicone, Carmen (2006). *Ordinary differential equations with applications.* Vol. 34. Springer Science & Business Media.

Faure, Mathieu and Gregory Roth (2010). "Stochastic approximations of set-valued dynamical systems: Convergence with positive probability to an attractor". In: *Mathematics of Operations Research* 35.3, pp. 624–640.

Harris, Charles R, K Jarrod Millman, Stéfan J Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. (2020). "Array programming with NumPy". In: *nature* 585.7825, pp. 357–362.

Hunter, John D (2007). "Matplotlib: A 2D graphics environment". In: *Computing in science & engineering* 9.03, pp. 90–95.

Mertikopoulos, Panayotis, Ya-Ping Hsieh, and Volkan Cevher (2024). "A unified stochastic approximation framework for learning in games". In: *Mathematical Programming* 203.1, pp. 559–609.

Palis Jr, J, W de Melo, et al. (1982). *Geometric Theory of Dynamical Systems.* Springer New York.

Papadimitriou, Christos and Georgios Piliouras (2018). "From nash equilibria to chain recurrent sets: An algorithmic solution concept for game theory". In: *Entropy* 20.10, p. 782.

Robbins, Herbert and Sutton Monro (1951). "A stochastic approximation method". In: *The annals of mathematical statistics*, pp. 400–407.

Van Rossum, Guido, Fred L Drake, et al. (1995). *Python reference manual.* Vol. 111. Centrum voor Wiskunde en Informatica Amsterdam.

Virtanen, Pauli, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. (2020). "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature methods* 17.3, pp. 261–272.