Econometrica Supplementary Material

SUPPLEMENT TO "TRUTHFUL EQUILIBRIA IN DYNAMIC BAYESIAN GAMES"

(*Econometrica*, Vol. 83, No. 5, September 2015, 1795–1848)

BY JOHANNES HÖRNER, SATORU TAKAHASHI, AND NICOLAS VIEILLE

THIS SUPPLEMENT CONTAINS ADDITIONAL MATERIAL on Markov decision problems and details on the proof of Theorem 5.

APPENDIX E: MARKOV DECISION PROBLEMS

E.1. *The ACOE*

For the reader's convenience, we provide a statement and a self-contained proof of the average cost optimality equation for MDPs. The material in this section is standard.

We let an irreducible MDP $\mathcal{M}$ with finite primitives be given. The state space is S, the action set is A, the reward function is $r : S \times A \to \mathbf{R}$, and the transition function is $p(\cdot \mid s, a)$.[56] We let Σ denote the set of strategies in $\mathcal{M}$.

For $\delta < 1$ and $N \in \mathbf{N}$, we let

$$v_\delta(s) := \max_{\sigma \in \Sigma} \mathbf{E}_{s,\sigma}\left[(1-\delta)\sum_{n=1}^{\infty} \delta^{n-1} r(s_n, a_n)\right]$$

and

$$v_N(s) := \max_{\sigma \in \Sigma} \mathbf{E}_{s,\sigma}\left[\frac{1}{N}\sum_{n=1}^{N} r(s_n, a_n)\right]$$

denote the values of the discounted and finite-horizon versions of $\mathcal{M}$, as a function of the initial state s.

PROPOSITION 6—ACOE: *There is a unique* $v \in \mathbf{R}$ *and a unique* (*up to an additive constant*) *map* $\theta : S \to \mathbf{R}$ *such that*

$$v + \theta(s) = \max_{a \in A}\left\{r(s, a) + \mathbf{E}_{p(\cdot|s,a)}\theta(\cdot)\right\} \quad \text{for all } s \in S. \tag{21}$$

In addition, $v = \lim_{\delta \to 1} v_\delta(s) = \lim_{N \to +\infty} v_N(s)$ *for all* $s \in S$.

[56]We are thus assuming that the sets S and A are finite, and that for each policy $\rho : S \to \Delta(A)$, the induced Markov chain (s_n) is irreducible.

 DOI: 10.3982/ECTA12182

PROOF: We first prove the existence of a solution to (21). For $\delta < 1$, the dynamic programming principle is written

$$v_\delta(s) = \max_{a \in A}\{(1-\delta)r(s,a) + \delta \mathbf{E}_{p(\cdot|s,a)} v_\delta(\cdot)\} \quad \text{for all } s \in S. \tag{22}$$

Let $a^*(s)$ achieve the maximum in (22), so that $v_\delta(s) = (1-\delta)r(s, a^*(s)) + \delta \mathbf{E}_{p(\cdot|s,a^*_s)} v_\delta(\cdot)$ for each s. This implies that $\delta \mapsto v_\delta(s)$ is a bounded and rational function on $[0, 1)$. In particular, both $v(s) := \lim_{\delta\to 1} v_\delta(s)$ and $\theta(s) := \lim_{\delta\to 1} \frac{v_\delta(s)-v(s)}{1-\delta}$ exist. Irreducibility readily implies that $v(s)$ is independent of s.

Equation (22) then can be rewritten as

$$v + (v_\delta(s) - v) = \max_{a\in A}\{(1-\delta)r(s,a) + \delta \mathbf{E}_{p(\cdot|s,a)}[v_\delta(t) - v] + \delta v\}.$$

Equation (21) follows when dividing by $1-\delta$ and letting $\delta \to 1$.

We next prove uniqueness, and start with v. Let (v, θ) be a solution to (21), so that

$$\theta(s) = \max_{a\in A}\{r(s,a) + \mathbf{E}_{p(\cdot|s,a)}\theta(\cdot)\} - v. \tag{23}$$

Substituting (23) into the right-hand side of (21) yields first

$$2v + \theta(s) = \max_\sigma \mathbf{E}_{s,\sigma}[r(s_1,a_1) + r(s_2,a_2) + \theta(s_3)],$$

and, by induction,

$$v + \frac{\theta(s)}{N} = \max_\sigma \mathbf{E}_{s,\sigma}\left[\frac{1}{N}\sum_{n=1}^{N} r(s_n,a_n) + \frac{\theta(s_{N+1})}{N}\right]$$

for each N. This implies that $\lim_{N\to\infty} v_N(s)$ exists and is equal to v.

We conclude with the uniqueness of θ. Let (v,θ) and (v,ψ) be two solutions to (21). This implies

$$\theta(s) - \psi(s) \le \max_{a\in A} \mathbf{E}_{p(\cdot|s,a)}(\theta(\cdot) - \psi(\cdot))$$

for each s. By irreducibility, it follows that $\theta(\cdot) - \psi(\cdot)$ is constant. *Q.E.D.*

E.2. *Perturbed Markov Chains and Relative Values*

We discuss here two statements on the asymptotic properties of relative values of perturbed Markov chains, as the perturbation parameter converges to zero. These statements readily imply those used in the main body of the paper.

E.2.1. *Result 1*

The setup is as follows. Let (disjoint) sets S_l with $1 \leq l \leq L$ be given. Also, for each l, let an irreducible transition function p_l on S_l with invariant measure ν_l and a "payoff" $r_l : S_l \to \mathbf{R}$ with $\mathbf{E}_{\nu_l}[r_l(s)] = 0$ be given. Let $\theta_l : S_l \to \mathbf{R}$ denote the associated relative value.

In addition, let p be an irreducible transition function on $\mathcal{S} := S_1 \cup \cdots \cup S_L$, and let $r : \mathcal{S} \to \mathbf{R}$ be the function that coincides with r_l on S_l. For $\varepsilon > 0$, we define a transition function p_ε on $\mathcal{S}$ as $p_\varepsilon(t \mid s) := (1-\varepsilon)p_l(t \mid s) + \varepsilon p(t \mid s)$ for $s \in S_l$ and $t \in \mathcal{S}$. Let $\mu_\varepsilon \in \Delta(\mathcal{S})$ be the invariant measure of p_ε, let $\gamma_\varepsilon := \mathbf{E}_{\mu_\varepsilon}[r(s)]$ be the long-run payoff, and let $\theta_\varepsilon : \mathcal{S} \to \mathbf{R}$ be the relative value. To fix ideas, we normalize θ_ε by imposing the condition $\mathbf{E}_{\mu_\varepsilon}[\theta_\varepsilon(\cdot)] = 0$.

PROPOSITION 7: *The map $\varepsilon \mapsto \theta_\varepsilon$ is bounded. In addition,*

$$\lim_{\varepsilon \to 0}\big(\theta_\varepsilon(s') - \theta_\varepsilon(s)\big) = \theta_l(s') - \theta_l(s) \quad \textit{for every } s, s' \in S_l.$$

PROOF: We view each transition $p_\varepsilon(\cdot \mid s)$ as the succession of two random choices. First, it is randomly decided, with probability ε, whether to use p or p_l to draw the next state, which is next drawn accordingly. We denote by τ the random time of *first* "switch" (first round where p is used).

Given any two states $s, s' \in \mathcal{S}$, we denote by (s_n) and (s'_n) two Markov chains with transition functions p_ε starting from s and s', respectively, which are coupled in that (i) the successive switches occur in the same rounds for the two chains and (ii) $s_n = s'_n$ after the first coincidence time $\omega := \inf\{n : s_n = s'_n\}$; yet all other random choices are independent.

CLAIM 6: *The following statements hold*:

- *There exists $c_1 > 0$ such that $\mathbf{E}[\sum_{n=1}^{\tau-1}(r(s_n) - r(s'_n))] \leq c_1$ for all $s, s' \in \mathcal{S}$ and $\varepsilon > 0$.*
- *There exists $c_2 > 0$ such that $\mathbf{P}(\omega \leq \tau) \geq c_2$ for every l, $s, s' \in S_l$ and $0 < \varepsilon \leq \frac{1}{2}$.*

PROOF: Let $s \in \mathcal{S}$ be given, say $s \in S_l$. One has, with obvious notation,

$$\mathbf{E}_\varepsilon\left(\sum_{n=1}^{\tau-1} r(s_n)\right) = \mathbf{E}_l\left(\sum_{n=1}^{\tau-1} r(s_n)\right).$$

By the ACOE, the latter is equal to $\theta_l(s) - \mathbf{E}_l[\theta_l(s_\tau)]$ and is therefore bounded as a function of ε.

The second statement follows from the irreducibility of p_l.[57] *Q.E.D.*

[57]The term $\mathbf{P}(\omega \leq \tau)$ is continuous as a function of ε, converges to 1 as $\varepsilon \to 0$, and is less than 1, except for $\varepsilon = 1$.

Next, we denote by (τ_k) the successive switches, so that $\tau_1 = \tau$. Given $s, s' \in \mathcal{S}$, denote by ϕ the smallest index k such that s_{τ_k+1} and s'_{τ_k+1} belong to the same component S_l. Because p is irreducible, there exists $c_3 > 0$ such that $\mathbf{P}(\phi \leq L) \geq c_3$. Note that

$$\theta_\varepsilon(s) = \mathbf{E}\left[\sum_{n=1}^{\tau_{L+1}}\big(r(s_n) - \gamma_\varepsilon\big) + \theta_\varepsilon(s_{\tau_{L+1}+1})\right]$$

and a similar equality holds for $\theta_\varepsilon(s')$; hence

$$\begin{aligned}\theta_\varepsilon\big(s'\big) - \theta_\varepsilon(s) &= \mathbf{E}\left[\sum_{n=1}^{\tau_{L+1}}\big(r(s_n) - r\big(s'_n\big)\big)\right] + \mathbf{E}\big[\theta_\varepsilon\big(s'_{\tau_{L+1}+1}\big) - \theta_\varepsilon(s_{\tau_{L+1}+1})\big] \\ &\leq Lc_1 + \max_{t,t' \in \mathcal{S}}\big(\theta_\varepsilon\big(t'\big) - \theta_\varepsilon(t)\big) \times \mathbf{P}(\omega > \tau_{L+1}) \\ &\leq Lc_1 + (1 - c_2c_3)\max_{t,t' \in \mathcal{S}}\big(\theta_\varepsilon\big(t'\big) - \theta_\varepsilon(t)\big),\end{aligned}$$

using the previous claim. It follows that $\max_{s,s' \in \mathcal{S}} |\theta_\varepsilon(s') - \theta_\varepsilon(s)| \leq \frac{Lc_1}{c_2c_3}$. Together with the equality $\mathbf{E}_{\mu_\varepsilon}\theta_\varepsilon(\cdot) = 0$, this implies the first statement.

For $\varepsilon > 0$, θ_ε is the unique solution to the linear system $(s \in S_l, l \leq L)$

$$\gamma_\varepsilon + \theta_\varepsilon(s) = r(s) + (1 - \varepsilon)\mathbf{E}_{p_l(\cdot|s)}\theta_\varepsilon(t) + \varepsilon\mathbf{E}_{p(\cdot|s)}\theta_\varepsilon(t),$$

together with the normalization equation.[58] Therefore, $\theta_\varepsilon(s)$ is a bounded and rational function of s. Thus, $\theta(s) := \lim_{\varepsilon\to 0}\theta_\varepsilon(s)$ exists and satisfies the limit system obtained when setting $\varepsilon = 0$; that is, for fixed l and for each $s \in S_l$, one has

$$\theta(s) = r(s) + \mathbf{E}_{p_l(\cdot|s)}\theta(t).$$

All solutions of the latter system are equal to θ_l up to an additive constant, hence the result. *Q.E.D.*

E.2.2. *Result 2*

The setup here is a variant of the previous one. Let two (disjoint) sets S_1 and S_2, an irreducible transition function p_l on S_l with invariant measure ν_l, a function $r_l : S_l \to \mathbf{R}$ $(l = 1, 2)$, and θ_l the relative value be given. In addition, let $f : \mathcal{S} := S_1 \cup S_2 \to \mathcal{S}$ be such that $f(S_1) \subseteq S_2$ and $f(S_2) \subseteq S_1$, and let $r : \mathcal{S} \to \mathbf{R}$ be the map whose restriction to S_l is r_l.

[58]Since μ_ε is the unique solution of a linear system with coefficients linear in ε, $\varepsilon \mapsto \mu_\varepsilon$ is a rational function, hence $\varepsilon \mapsto \gamma_\varepsilon$ is a rational function as well.

For $\varepsilon = (\varepsilon_1, \varepsilon_2) \in (0,1)^2$, we define a transition function p_ε over $\mathcal{S}$ by $p_\varepsilon(t \mid s) = (1-\varepsilon_l)p_l(t \mid s) + \varepsilon_l f(s)$ for $s \in S_l$. Thus, transitions from S_1 to S_2 (resp., from S_2 to S_1) occur with probability ε_1 (resp., ε_2) in each round. Let $\theta_\varepsilon : \mathcal{S} \to \mathbf{R}$ denote the relative value.

PROPOSITION 8: *One has* $\lim_{\varepsilon\to 0}(\theta_\varepsilon(s') - \theta_\varepsilon(s)) = \theta_l(s') - \theta_l(s)$ *whenever* $s, s' \in S_l$.

Note, however, that θ_ε is unbounded as a function of ε as soon as $\mathbf{E}_{\nu_1} r_1(\cdot) \neq \mathbf{E}_{\nu_2} r_2(\cdot)$.

PROOF: We first prove that $\varepsilon \mapsto \theta_\varepsilon(s') - \theta_\varepsilon(s)$ is bounded whenever $s, s' \in S_l$. We use the same notations as in the proof of Proposition 7, and let (s_n) and (s'_n) be two Markov chains starting from s and s', with transition function p_ε and coupled as before. The constants c_1 and c_2 are as before. Whenever $s, s' \in S_l$ (and for ε_l bounded away from 1), one has $\mathbf{P}(\omega \leq \tau) \geq c_2$, hence

$$\left|\theta_\varepsilon\left(s'\right) - \theta_\varepsilon(s)\right| \leq c_1 + (1-c_2) \max_{t,t' \in S_{3-l}} \left|\theta_\varepsilon\left(t'\right) - \theta_\varepsilon(t)\right|.$$

It follows that $\max_{l=1,2} \max_{s,s' \in S_l} |\theta_\varepsilon(s') - \theta_\varepsilon(s)| \leq \frac{c_1}{c_2}$.

The limit claim follows as in the proof of Proposition 7. *Q.E.D.*

E.3. *Proof of Proposition 5*

We let an irreducible MDP $\mathcal{M}_0$ be given, with primitives (Ω, B, q, r). We denote by $v \in \mathbf{R}$ and $\theta : \Omega \to \mathbf{R}$ the limit value and relative values of $\mathcal{M}_0$. For $\omega \in \Omega$, we let

$$B_0(\omega) := \arg\max_{b \in B} \left\{ r(\omega, b) + \mathbf{E}_{\omega' \sim q(\cdot \mid \omega, b)} \theta\left(\omega'\right) \right\}$$

be the set of actions that are optimal at $\omega \in \Omega$.

Thus, for $\omega \in \Omega$ and $b \notin B_0(\omega)$, one has $r(\omega, b) + \mathbf{E}_{q(\cdot\mid\omega,b)}\theta(\omega') < v + \theta(\omega)$, and we let $c_0 > 0$ be strictly smaller than the difference between the two sides, for each ω and $b \notin B_0(\omega)$.

In the absence of transfers, assume that the second agent uses a distribution $\rho(\omega) \in \Delta(B_0(\omega))$ with full support, as a function of the report ω of the first agent. At state ω, it is strictly better to report truthfully ω rather than $\tilde{\omega}$ unless $B(\tilde{\omega}) \subseteq B(\omega)$. The main issue below will be to get rid of such indifference cases and to prevent the first agent from reporting a state $\tilde{\omega}$ with $B(\tilde{\omega}) \subset B(\omega)$. The basic insight of the proof is to reward the first agent for reporting a state with many optimal actions.

We will construct a finite sequence $\mathcal{M}_1, \ldots, \mathcal{M}_n$ of perturbed MDPs. For all MDPs in the sequence, the state space is Ω and the action set is B.

We explain the construction of $\mathcal{M}_1$ before proceeding to the general case. Throughout, we fix an increasing function $\phi : \{1, \dots, |B|\} \to \mathbf{R}$ such that $\phi(|B|) < \frac{1}{|B|}$ (so that $\phi(m) < \frac{1}{m}$ for $m \le |B|$). We then pick $\alpha > 0$ such that (i) $\alpha < \frac{1}{|B|} - \phi(|B|)$ and (ii) $\alpha < \phi(m+1) - \phi(m)$ for all $1 \le m < |B|$.

Given $\varepsilon_1 > 0$, the reward r_1 and transition function q_1 of $\mathcal{M}_1$ are defined as

$$r_1(\omega, b) := (1 - \varepsilon_1) r(\omega, b) + \varepsilon_1 \big(r\big(\omega, \beta_0(\omega)\big) + c_0 \phi\big(\big|B_0(\omega)\big|\big)\big)$$

and

$$q_1(\cdot \mid \omega, b) := (1 - \varepsilon_1) q(\cdot \mid \omega, b) + \varepsilon_1 q\big(\cdot \mid \omega, \beta_0(\omega)\big),$$

where $\beta_0(\omega)$ is the uniform distribution over $B_0(\omega)$. We denote by v_{ε_1} and θ_{ε_1} the limit value and relative values of $\mathcal{M}_1$, and we let

$$B_1(\omega) := \arg\max_{b \in B} \big\{ r_1(\omega, b) + \mathbf{E}_{\omega' \sim q_1(\cdot \mid \omega, b)} \theta\big(\omega'\big) \big\}$$

be the set of optimal actions at ω in $\mathcal{M}_1$. Both v_{ε_1} and θ_{ε_1} are continuous w.r.t. ε_1, with $\lim_{\varepsilon_1 \to 0} v_{\varepsilon_1} = v$ and $\lim_{\varepsilon_1 \to 0} \theta_{\varepsilon_1} = \theta$. Hence $B_1(\omega)$ is upper hemicontinuous as a function of ε_1, so that $B_1(\omega) \subseteq B_0(\omega)$ for all $\varepsilon_1 > 0$ small enough, and $\omega \in \Omega$. We stop with the MDP $\mathcal{M}_1$ if there is a sequence $\varepsilon_1 \to 0$ such that $B_1(\cdot) = B_0(\cdot)$ along the sequence. We otherwise repeat the perturbation process with $\mathcal{M}_1$.

More generally, let $(\varepsilon_k)_{k \in \mathbf{N}}$ be a sequence of positive real numbers with $\sum_k \varepsilon_k < 1$. For $k \in \mathbf{N}$, we set $\vec{\varepsilon}_k := (\varepsilon_1, \dots, \varepsilon_k)$. For any such sequence (ε_k), we define inductively a sequence $\mathcal{M}_k(\vec{\varepsilon}_k)$ of MDPs with state space Ω and action set B, and with limit value denoted $v_{\vec{\varepsilon}_k}$ and $\theta_{\vec{\varepsilon}_k}$. The reward r_k and transition function q_k of $\mathcal{M}_k(\vec{\varepsilon}_k)$ are defined as

$$\begin{aligned} r_k(\omega, b) := {} & \left(1 - \sum_{i=1}^{k} \varepsilon_i\right) r(\omega, b) + \sum_{i=1}^{k} \varepsilon_i \big(r\big(\omega, \beta_{i-1}(\omega)\big) \\ & + c_{i-1} \phi\big(\big|B_{i-1}(\omega)\big|\big)\big) \end{aligned}$$

and

$$q_k(\cdot \mid \omega, b) := \left(1 - \sum_{i=1}^{k} \varepsilon_i\right) q(\cdot \mid \omega, b) + \sum_{i=1}^{k} \varepsilon_i q\big(\cdot \mid \omega, \beta_{i-1}(\omega)\big),$$

where $\beta_i(\omega)$, $B_i(\omega)$, and c_i are defined inductively as follows.

For each i,

$$B_i(\omega) := \arg\max_{b} \big\{ r_i(\omega, b) + \mathbf{E}_{\omega' \sim q_i(\cdot \mid \omega, b)} \theta_{\vec{\varepsilon}_i}\big(\omega'\big) \big\}$$

is the set of actions optimal at ω in $\mathcal{M}_i(\vec{\varepsilon}_i)$, $\beta_i(\omega) \in \Delta(B)$ is the uniform distribution over $B_i(\omega)$, and $c_i > 0$ is any number such that

$$c_i + r_i(\omega, b) + \mathbf{E}_{\omega' \sim q_i(\cdot|\omega,b)} \theta_{\vec{\varepsilon}_i}(\omega') < v_{\vec{\varepsilon}_i} + \theta_{\vec{\varepsilon}_i}(\omega)$$

for each $\omega \in \Omega$ and $b \notin B_i(\omega)$. This definition entails no circularity. Indeed, B_0, β_0, and c_0 are associated with $\mathcal{M}_0$, and for $k \geq 1$, the definition of r_k and q_k, and therefore of $v_{\vec{\varepsilon}_k}$, $\theta_{\vec{\varepsilon}_k}$, B_k, β_k, and c_k, only involves $v_{\vec{\varepsilon}_i}$ and $\theta_{\vec{\varepsilon}_i}$ for $i < k$.

Note also that for given $\vec{\varepsilon}_{k-1}$, $v_{\vec{\varepsilon}_k}$ and $\theta_{\vec{\varepsilon}_k}$ are continuous as functions of ε_k, and $B_k(\omega)$ is therefore upper hemicontinuous. It follows that for every $\vec{\varepsilon}_{k-1}$, one has $B_k(\omega) \subseteq B_{k-1}(\omega)$ provided $\varepsilon_k > 0$ is small enough. In addition, $\lim_{\varepsilon_k \to 0} v_{\vec{\varepsilon}_k} = v_{\vec{\varepsilon}_{k-1}}$ and $\lim_{\varepsilon_k \to 0} \theta_{\vec{\varepsilon}_k} = \theta_{\vec{\varepsilon}_{k-1}}$.

In the sequel, we let a sequence (ε_k) be given such that for each k, ε_k is "very close to zero" given $\vec{\varepsilon}_{k-1}$. By this, we mean that (i) $|v_{\vec{\varepsilon}_k} - v_{\vec{\varepsilon}_{k-1}}|$ and $\|\theta_{\vec{\varepsilon}_k} - \theta_{\vec{\varepsilon}_{k-1}}\|$ are smaller than some positive numbers that only involve $\vec{\varepsilon}_{k-1}$ (and that will appear in the computations below), and (ii) $B_k(\omega) \subset B_{k-1}(\omega)$ for every $\omega \in \Omega$.

We let $n \in \mathbf{N}$ be such that $B_n(\cdot) = B_{n-1}(\cdot)$, and we define $\rho : \Omega \to \Delta(B)$ as

$$\rho(\omega) := \left(1 - \sum_{k=1}^{n} \varepsilon_k\right)\beta_n(\omega) + \sum_{k=1}^{n} \varepsilon_k \beta_{k-1}(\omega).$$

Observe that $\operatorname{supp} \rho(\omega) = B_0(\omega)$ for each ω. We next define $x_{\text{eq}} : \Omega \times B \to \mathbf{R}$ as follows:

- For $b \in B_0(\omega)$, $x_{\text{eq}}(\omega, b)$ is defined by the equation

$$\begin{aligned} &x_{\text{eq}}(\omega, b) + r(\omega, b) + \mathbf{E}_{\omega' \sim q(\cdot|\omega,b)} \theta_{\vec{\varepsilon}_n}(\omega') \\ &\quad = r(\omega, \rho(\omega)) + \mathbf{E}_{\omega' \sim q(\cdot|\omega,\rho(\omega))} \theta_{\vec{\varepsilon}_n}(\omega'). \end{aligned}$$

Observe that $x_{\text{eq}}(\omega, \rho(\omega)) = \mathbf{E}_{b \sim \rho(\omega)} x_{\text{eq}}(\omega, b) = 0$ for each ω.

- For $b \notin B_0(\omega)$, we set $x_{\text{eq}}(\omega, b) = x_{\text{eq}}(\omega, \bar{b})$, where $\bar{b} \in B_n(\omega)$. Note that $x_{\text{eq}}(\omega, b)$ is independent of the choice of $\bar{b}$. Indeed, the actions of $B_n(\omega)$ are those that maximize $r_n(\omega, \cdot) + \mathbf{E}_{q_n(\cdot|\omega,\cdot)} \theta_{\vec{\varepsilon}_n}(\omega')$ or, equivalently, that maximize $r(\omega, \cdot) + \mathbf{E}_{q(\cdot|\omega,\cdot)} \theta_{\vec{\varepsilon}_n}(\omega')$.

Finally, we define $x : \Omega \times B \to \mathbf{R}$ as

$$x(\omega, b) := x_{\text{eq}}(\omega, b) + \sum_{k=1}^{n} \varepsilon_k c_{k-1} \phi(|B_{k-1}(\omega)|).$$

We now prove that the pair (ρ, x) satisfies the desired properties.

CLAIM 7: *The term ρ is an optimal policy in the MDP with stage payoff $r(\omega, b) + x(\omega, b)$.*

PROOF: Recall that $v_{\vec{\varepsilon}_n}$ and $\theta_{\vec{\varepsilon}_n}$ are the limit value and relative values of the MDP $\mathcal{M}_n(\vec{\varepsilon}_n)$, and that $B_n(\omega)$ is the set of actions optimal at ω. Therefore, for each ω and by the ACOE, one has

$$v_{\vec{\varepsilon}_n} + \theta_{\vec{\varepsilon}_n}(\omega) = r_n\big(\omega, \beta_n(\omega)\big) + \mathbf{E}_{q_n(\cdot|\omega,\beta_n(\omega))}\theta_{\vec{\varepsilon}_n}\big(\omega'\big).$$

Given the definition of r_n, q_n, and $\rho(\omega)$, the right-hand side is also equal to

$$r\big(\omega, \rho(\omega)\big) + x\big(\omega, \rho(\omega)\big) + \mathbf{E}_{q(\cdot|\omega,\rho(\omega))}\theta_{\vec{\varepsilon}_n}\big(\omega'\big).$$

Next, it follows from the definition of x_{eq} that

$$r(\omega, b) + x(\omega, b) + \mathbf{E}_{q(\cdot|\omega,b)}\theta_{\vec{\varepsilon}_n}\big(\omega'\big)$$

is independent of $b \in \operatorname{supp} \rho(\omega) = B_0(\omega)$.

On the other hand, for $b \notin B_0(\omega)$ and $\bar{b} \in B_n(\omega)$, one has $r_n(\omega, b) + \mathbf{E}_{q_n(\cdot|\omega,b)}\theta_{\vec{\varepsilon}_n}(\omega') < r_n(\omega, \bar{b}) + \mathbf{E}_{q_n(\cdot|\omega,\bar{b})}\theta_{\vec{\varepsilon}_n}(\omega')$, which implies $r(\omega, b) + \mathbf{E}_{q(\cdot|\omega,b)}\theta_{\vec{\varepsilon}_n}(\omega') < r(\omega, \bar{b}) + \mathbf{E}_{q(\cdot|\omega,\bar{b})}\theta_{\vec{\varepsilon}_n}(\omega')$, which yields, in turn,

$$\begin{aligned} &r(\omega, b) + x(\omega, b) + \mathbf{E}_{q(\cdot|\omega,b)}\theta_{\vec{\varepsilon}_n}\big(\omega'\big) \\ &\quad < r(\omega, \bar{b}) + x(\omega, \bar{b}) + \mathbf{E}_{q(\cdot|\omega,\bar{b})}\theta_{\vec{\varepsilon}_n}\big(\omega'\big). \end{aligned}$$

Together, these observations yield

$$v_{\vec{\varepsilon}_n} + \theta_{\vec{\varepsilon}_n} = \max_{b\in B}\big\{r(\omega, b) + x(\omega, b) + \mathbf{E}_{q(\cdot|\omega,b)}\theta_{\vec{\varepsilon}_n}\big(\omega'\big)\big\},$$

with the maximum being achieved by $\rho(\omega)$. This proves the claim. *Q.E.D.*

CLAIM 8: *For every $\omega, \tilde{\omega} \in \Omega$, one has*

$$\begin{aligned} &r\big(\omega, \rho(\omega)\big) + x\big(\omega, \rho(\omega)\big) + \mathbf{E}_{q(\cdot|\omega,\rho(\omega))}\theta_{\vec{\varepsilon}_n}\big(\omega'\big) \\ &\quad \geq r\big(\omega, \rho(\tilde{\omega})\big) + x\big(\tilde{\omega}, \rho(\tilde{\omega})\big) + \mathbf{E}_{q(\cdot|\omega,\rho(\tilde{\omega}))}\theta_{\vec{\varepsilon}_n}\big(\omega'\big), \end{aligned} \tag{24}$$

with a strict inequality if $\rho(\omega) \neq \rho(\omega')$.

PROOF: Fix $\omega, \tilde{\omega} \in \Omega$. Note that $\rho(\omega) = \rho(\tilde{\omega})$ if and only if $B_k(\omega) = B_k(\tilde{\omega})$ for $k = 0, \ldots, n$. Assume first that $\rho(\omega) = \rho(\tilde{\omega})$. Then, using $x_{\mathrm{eq}}(\omega, \rho(\omega)) = 0$, one has

$$\begin{aligned} x\big(\omega, \rho(\omega)\big) &= \sum_{k=1}^{n} \varepsilon_k c_{k-1}\phi\big(\big|B_{k-1}(\omega)\big|\big) \\ &= \sum_{k=1}^{n} \varepsilon_k c_{k-1}\phi\big(\big|B_{k-1}(\tilde{\omega})\big|\big) = x\big(\tilde{\omega}, \rho(\tilde{\omega})\big). \end{aligned}$$

Thus, (24) holds with equality.

Assume next that $\rho(\omega) \neq \rho(\tilde{\omega})$, and denote by $\bar{k}$ the smallest k such that $B_k(\omega) \neq B_k(\tilde{\omega})$. Since $B_n = B_{n-1}$, one has $\bar{k} < n$. We prove that (24) holds with a strict inequality by looking at the decomposition of ρ as a weighted sum of the uniform distributions β_k.

- For $k < \bar{k}$, one has $\beta_k(\omega) = \beta_k(\tilde{\omega})$, hence

$$\begin{aligned} &r\big(\omega, \beta_k(\omega)\big) + c_k \phi\big(\big|B_k(\omega)\big|\big) + \mathbf{E}_{q(\cdot|\omega,\beta_k(\omega))} \theta_{\vec{\varepsilon}_n}\big(\omega'\big) \\ &\quad = r\big(\omega, \beta_k(\tilde{\omega})\big) + c_k \phi\big(\big|B_k(\tilde{\omega})\big|\big) + \mathbf{E}_{q(\cdot|\omega,\beta_k(\tilde{\omega}))} \theta_{\vec{\varepsilon}_n}\big(\omega'\big). \end{aligned}$$

- For $\bar{k} < k < n$, we will rely on the assumption that ε_k is quite small compared to $\varepsilon_{\bar{k}}$. Plainly, one has, for some constant C that only depends on the primitives of the MDP,

$$\begin{aligned} &r\big(\omega, \beta_k(\omega)\big) + c_k \phi\big(\big|B_k(\omega)\big|\big) + \mathbf{E}_{q(\cdot|\omega,\beta_k(\omega))} \theta_{\vec{\varepsilon}_n}\big(\omega'\big) \\ &\quad \geq r\big(\omega, \beta_k(\tilde{\omega})\big) + c_k \phi\big(\big|B_k(\tilde{\omega})\big|\big) + \mathbf{E}_{q(\cdot|\omega,\beta_k(\tilde{\omega}))} \theta_{\vec{\varepsilon}_n}\big(\omega'\big) - C. \end{aligned}$$

Hence, when multiplied by $\varepsilon_{\bar{k}+1}$, the difference between the two sides of the latter inequality is small compared to $\varepsilon_{\bar{k}+1}$ and, in particular, less than $\alpha\varepsilon_{\bar{k}+1}c_{\bar{k}}$.

- For $k = n$, and since $B_n(\omega)$ are the actions optimal at ω in $\mathcal{M}_n(\vec{\varepsilon}_n)$, one has, as noted previously,

$$r\big(\omega, \beta_n(\omega)\big) + \mathbf{E}_{q(\cdot|\omega,\beta_n(\omega))} \theta_{\vec{\varepsilon}_n}\big(\omega'\big) \geq r\big(\omega, \beta_n(\tilde{\omega})\big) + \mathbf{E}_{q(\cdot|\omega,\beta_n(\tilde{\omega}))} \theta_{\vec{\varepsilon}_n}\big(\omega'\big).$$

We are left with $\bar{k} = k$ and we distinguish two cases. Assume first that $b \notin B_{\bar{k}}(\omega)$ for some $b \in B_{\bar{k}}(\tilde{\omega})$. In that case,

$$\begin{aligned} &r\big(\omega, \beta_{\bar{k}}(\omega)\big) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\omega))} \theta_{\vec{\varepsilon}_{\bar{k}}}\big(\omega'\big) \\ &\quad > r\big(\omega, \beta_{\bar{k}}(\tilde{\omega})\big) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\tilde{\omega}))} \theta_{\vec{\varepsilon}_{\bar{k}}}\big(\omega'\big) + c_{\bar{k}} \times \frac{|B_{\bar{k}}(\tilde{\omega}) \setminus B_{\bar{k}}(\omega)|}{|B_{\bar{k}}(\tilde{\omega})|} \end{aligned}$$

(because all actions in $B_{\bar{k}}(\tilde{\omega}) \setminus B_{\bar{k}}(\omega)$ are played with probability $\frac{1}{|B_{\bar{k}}(\tilde{\omega})|}$ and each leads to a loss of at least $c_{\bar{k}}$). Since $\varepsilon_{\bar{k}+1}, \ldots, \varepsilon_n$ are small (given $\varepsilon_{\bar{k}}$), the latter inequality still holds when $\theta_{\vec{\varepsilon}_n}$ is substituted for $\vec{\varepsilon}_{\bar{k}}$. This implies

$$\begin{aligned} &r\big(\omega, \beta_{\bar{k}}(\omega)\big) + c_{\bar{k}} \phi\big(\big|B_{\bar{k}}(\omega)\big|\big) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\omega))} \theta_{\vec{\varepsilon}_n}\big(\omega'\big) \\ &\quad > r\big(\omega, \beta_{\bar{k}}(\tilde{\omega})\big) + c_{\bar{k}} \phi\big(\big|B_{\bar{k}}(\tilde{\omega})\big|\big) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\tilde{\omega}))} \theta_{\vec{\varepsilon}_n}\big(\omega'\big) \\ &\qquad + c_{\bar{k}} \left(\frac{1}{|B_{\bar{k}}(\tilde{\omega})|} + \phi\big(\big|B_{\bar{k}}(\omega)\big|\big) - \phi\big(\big|B_{\bar{k}}(\tilde{\omega})\big|\big) \right) \\ &\quad > r\big(\omega, \beta_{\bar{k}}(\tilde{\omega})\big) + c_{\bar{k}} \phi\big(\big|B_{\bar{k}}(\tilde{\omega})\big|\big) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\tilde{\omega}))} \theta_{\vec{\varepsilon}_n}\big(\omega'\big) + c_{\bar{k}}\alpha, \end{aligned}$$

using property (i) of α.

Assume now that $B_{\bar{k}}(\tilde{\omega})$ is a proper subset of $B_{\bar{k}}(\omega)$, so that

$$r\big(\omega, \beta_{\bar{k}}(\omega)\big) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\omega))}\theta_{\vec{\varepsilon}_k}\big(\omega'\big) \\ = r\big(\omega, \beta_{\bar{k}}(\tilde{\omega})\big) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\tilde{\omega}))}\theta_{\vec{\varepsilon}_k}\big(\omega'\big),$$

because $\beta_k(\omega)$ is optimal in $\mathcal{M}_k(\vec{\varepsilon}_k)$. This implies

$$r\big(\omega, \beta_{\bar{k}}(\omega)\big) + c_{\bar{k}}\phi\big(\big|B_{\bar{k}}(\omega)\big|\big) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\omega))}\theta_{\vec{\varepsilon}_k}\big(\omega'\big) \\ > r\big(\omega, \beta_{\bar{k}}(\tilde{\omega})\big) + c_{\bar{k}}\phi\big(\big|B_{\bar{k}}(\tilde{\omega})\big|\big) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\tilde{\omega}))}\theta_{\vec{\varepsilon}_k}\big(\omega'\big) + \alpha c_{\bar{k}},$$

using property (ii) of α.

Since $\theta_{\vec{\varepsilon}_n}$ is very close to $\theta_{\vec{\varepsilon}_k}$, the latter inequality still holds when $\theta_{\vec{\varepsilon}_n}$ is substituted for $\theta_{\vec{\varepsilon}_{\bar{k}}}$. The desired inequality follows by summing over all $k = 1, \dots, n$. *Q.E.D.*

APPENDIX F: PROOF OF THEOREM 4

Most computations in Section F.2 will be omitted. Transition phases will rely on the strictly truthful pair $(\rho_{\text{ext},0}, x_0)$ constructed in Section B.2.2, with $K_0 = A$ and $\rho_0 : S \times K_0 \to A$. We supplement the transfers x_0 of Section B.2.2 with transfers $\bar{x}_0 : K_0 \times Y \to \mathbf{R}^I$, which induce obedience to ρ_0, and still denote by $x_0 : S \times K_0 \times Y \to \mathbf{R}^I$ the total transfers. We abbreviate the relative values $\theta_{\rho_0, r+x_0}$ to θ_0, and we let $\bar{r} \geq 1$ be a uniform bound on r and θ_0.

F.1. *Auxiliary Zero-Sum Games*

Throughout this section, we fix a player $i \in I$, and will introduce games between i and $-i$. Without loss of generality, all strategies of player i are here "babbling."

F.1.1. *Preliminaries*

For $k \in \mathbf{N}$ and $j \neq i$, we let $\mathcal{A}^j_k \subset \Delta(A^j)$ be a finite, $\frac{1}{k}$-dense subset of $\Delta(A^j)$; that is, for each $\alpha^j \in \Delta(A^j)$, there exists $\alpha^j_k \in \mathcal{A}^j_k$ such that $\|\alpha^j - \alpha^j_k\|_{L^1} < \frac{1}{k}$. We let Σ^j_k be the set of (repeated game) strategies of player j with the property that the mixed action of j in each round n belongs to $\mathcal{A}^j_k$ and only depends on the past public signals $y^i_1, \dots, y^i_{n-1}$ related to player i. We set $\Sigma^{-i}_k := \bigtimes_{j \neq i} \Sigma^i_k$ and let

$$w^i_k := \lim_{\delta \to 1} \min_{\sigma^{-i} \in \Sigma^{-i}_k} \max_{\sigma^i} \gamma^i_\delta\big(s^i, \sigma^i, \sigma^{-i}\big)$$

be the long-run minmax payoff when players $-i$ are constrained to strategies in Σ^{-i}_k.[59] Thanks to the irreducibility assumption, there exists $c > 0$ such that

[59]It is independent of s^i.

the following statement holds: for each $k \in \mathbf{N}$, $j \neq i$ and each strategy σ^j, there exists $\sigma_k^j \in \Sigma_k^j$ such that

$$\gamma_\delta^j\left(s, \sigma^{-j}, \sigma_k^j\right) < \gamma_\delta^j(s, \sigma) + \frac{c}{k}$$

for every $\delta < 1$ and σ^{-j}.[60] Hence, $\lim_{k \to +\infty} w_k^i = w^i$.

Since strategies of player $-i$ ignore (y_n^j) $(j \neq i)$, we may restrict ourselves to strategies of player i that are independent of the public signals (y_n^j), $j \neq i$, as well.

Let an arbitrary state $\bar{s}^i \in S^i$ be given and let $k \in \mathbf{N}$ be fixed. Given a horizon $T \in \mathbf{N}$, we let $G_k^i(\bar{s}^i, T)$ be the zero-sum game with T rounds between i and $-i$ with no communication, initial state $\bar{s}^i$, and payoff $\frac{1}{T}\sum_{n=1}^T r^i(s_n^i, a_n)$, and in which players $-i$ are restricted to Σ_k^{-i}. Denote by

$$w_k^i(T) := \min_{\sigma^{-i} \in \Sigma_k^{-i}} \max_{\sigma^i} \mathbf{E}_{\bar{s}^i, \sigma}\left[\frac{1}{T}\sum_{n=1}^T r^i\left(s_n^i, a_n\right)\right] \tag{25}$$

the minmax of $G^i(\bar{s}^i, T)$. Using irreducibility, one has $\lim_{T \to +\infty} w_k^i(T) = w_k^i$ for each $\bar{s}^i$.

Given k and T, we fix a strategy profile $\sigma_k^{-i} \in \Sigma_k^{-i}$ that achieves the minimum in (25). For $\alpha_k^{-i} \in \mathcal{A}_k^{-i}$, let $T(\alpha_k^{-i})$ be the (random) set of rounds in which σ_k^{-i} prescribes α_k^{-i}, and let $f_{\alpha_k^{-i}} \in \Delta(Y)$ denote the empirical distribution of the public signals received in $T(\alpha_k^{-i})$. Intuitively, if some player $j \neq i$ is playing according to σ_k^j, the signals (y_n^j) received in $T(\alpha_k^{-i})$ are i.i.d. and drawn from $p^j(\cdot \mid \alpha_k^j)$. Hence, whenever $|T(\alpha_k^{-i})|$ is large, then with high probability, $f_{\alpha_k^{-i}}$ should be close to the distribution $g_{\alpha_k^{-i}}^j \in \Delta(Y)$ defined as

$$g_{\alpha_k^{-i}}^j(y) = f_{\alpha_k^{-i}}\left(y^{-j}\right) p^j\left(y^j \mid \alpha_k^j\right).$$

This motivates the definition of

$$D^j := \sum_{\alpha_k^{-i} \in \mathcal{A}_k^{-i}} \frac{|T(\alpha_k^{-i})|}{T} \left\| f_{\alpha_k^{-i}} - g_{\alpha_k^{-i}}^j \right\|_{L^1}.$$

Claim 9 below formalizes this intuition. In words, and provided that T is large enough, player j can ensure that $D^j < \varepsilon$ with high probability by playing $\sigma_{k,T}^j$.

[60]This assertion also relies on the product monitoring assumption. Under this assumption, public communication and public signals y_n^j for $j \neq i$ cannot be used by players $-i$ as a means to privately correlate their actions against i.

CLAIM 9: *Given* $\varepsilon > 0$, *there exists* $T_0 \geq 0$ *such that*

$$\mathbf{P}_{\sigma^j_{k,T},\sigma^{-j}}\left(D^j > \varepsilon\right) < \varepsilon$$

for all $T \geq T_0$, $j \neq i$, *and* σ^{-j}.

Claim 9 follows from Gossner (1995), who uses Blackwell's theory of approachability. It will be combined with Claim 10 below, which asserts that player i is effectively punished when all players $j \neq i$ pass the test $D^j < \varepsilon$ with high probability.

CLAIM 10: *Let* $\varepsilon > 0$ *and* T *be given*, *and let* σ *be a strategy profile such that* $\mathbf{P}_\sigma(D^j > \varepsilon) < \varepsilon$ *for each* $j \neq i$. *Then*

$$\mathbf{E}_{\bar{s}^i,\sigma}\left[\frac{1}{T}\sum_{n=1}^{N} r^i\left(s^i_n, a_n\right)\right] < w^i_{k,T} + \bar{r}(I+2)\varepsilon.$$

PROOF: On the event $\mathcal{D}^{-i} := \bigcap_{j\neq i}\{D^j \leq \varepsilon\}$, one has $\sum_{\mathcal{A}^{-i}_k} \frac{|T(\alpha^{-i}_k)|}{T}\|f_{\alpha^{-i}_k} - g^j_{\alpha^{-i}_k}\|_{L^1} \leq \varepsilon$ for each $j \neq i$, which implies, by repeated substitution,

$$\sum_{\mathcal{A}^{-i}_k}\frac{|T(\alpha^{-i}_k)|}{T}\left(\sum_y\left|f_{\alpha^{-i}_k}(y) - f_{\alpha^{-i}_k}\left(y^i\right)\bigtimes_{j\neq i} p^j\left(y^j \mid \alpha^j_k\right)\right|\right) < I\varepsilon.$$

We fix now an arbitrary private history (s^i_n, a^i_n, y_n) of player i, and compare the realized payoff $\frac{1}{T}\sum_{n=1}^{n} g^i(s^i_n, a^i_n, y_n)$ to its "expectation," assuming (y^j_n) are drawn using $\sigma^j_{k,T}$. Formally,

$$\begin{aligned}
&\frac{1}{T}\sum_{n=1}^{n} g^i\left(s^i_n, a^i_n, y_n\right)\\
&\quad= \frac{1}{T}\sum_{\mathcal{A}^{-i}_k}\sum_{T(\alpha^{-i}_k)}\Biggl(g^i\left(s^i_n, a^i_n, y_n\right)\\
&\qquad - \sum_{\tilde{y}^{-i}\in Y^{-i}} g^i\left(s^i_n, a^i_n, \tilde{y}^{-i}, y^i_n\right) \times p^{-i}\left(\tilde{y}^{-i} \mid \alpha^{-i}_k\right)\Biggr)\\
&\qquad + \frac{1}{T}\sum_{\mathcal{A}^{-i}_k}\sum_{T(\alpha^{-i}_k)}\sum_{\tilde{y}^{-i}} g^i\left(s^i_n, a^i_n, \tilde{y}^{-i}, y^i_n\right) \times p^{-i}\left(\tilde{y}^{-i} \mid \alpha^{-i}_k\right).
\end{aligned}$$

The expectation of the second term is independent of σ^{-i} and is equal to

$$\mathbf{E}_{\bar{s}^i,\sigma^i,\sigma_{k,T}^{-i}}\left[\frac{1}{T}\sum_{n=1}^{T} r^i\big(s_n^i, a_n\big)\right] \le w_k^i(T).$$

Since the first term is bounded by $2\bar{r}$ and by $\bar{r}I\varepsilon$ on the event $\mathcal{D}^{-i}$, the result follows. *Q.E.D.*

F.1.2. *Auxiliary Games*

From now on and given $\delta < 1$, we set $T := \frac{1}{\sqrt{1-\delta}}$. Given $\delta < 1$, transfers $x : S \times Y^T \to \mathbf{R}^I$, and a state profile $s \in S$, we let $G(s, \delta, x)$ denote the game of T rounds (ending after the draw of s_{T+1}), with initial state profile s, with no communication, and with payoff

$$\frac{1-\delta}{1-\delta^T}\left\{\sum_{n=1}^{T}\delta^{n-1}r(s_n, a_n) + \delta^T x(s, \vec{y}) + \delta^T\theta_0(s_{T+1}, \bar{a}_0)\right\},$$

where $\vec{y} := (y_1, \dots, y_T)$ is the sequence of public signals received along the play and $\bar{a}_0 \in A$ is fixed.

The following result will serve as the building block of the equilibrium construction of punishment phases.

LEMMA 13: *Given $\varepsilon > 0$, there exist $\kappa_* \in \mathbf{R}$ and $\delta_* < 1$ such that for all $\delta > \delta_*$, there exist $x : S \times Y^T \to \mathbf{R}^I$ and $\gamma \in \mathbf{R}^I$ with the following properties*:

(a) *For all $s \in S$, γ is a sequential equilibrium payoff of $G(s, \delta, x)$.*
(b) *We have $\gamma^i < w^i + \varepsilon$.*
(c) *We have $x^i \ge 0$ and $\|x\| \le \kappa_* T$.*

PROOF: Let $\varepsilon > 0$ be given and pick $\varepsilon' < \frac{\varepsilon}{2\bar{r}(I+5)}$. Choose $k \in \mathbf{N}$ such that $|w_k^i - w^i| < \varepsilon'$, choose $C > \frac{4\bar{r}}{\varepsilon'}$, and apply Claim 9 with ε' to get T_0. We will show that the result holds with $\kappa_* := 2C$ and $\delta_* < 1$ large enough so that (i) $T \ge T_0$, (ii) $|w_k^i - w_k^i(T)| < \varepsilon'$ (for each $\bar{s}^i$), and (iii) $\frac{1-\delta}{1-\delta^T}C < 1$, and both of the inequalities displayed below hold for each $\delta > \delta_*$:

$$-\delta^T\bar{r} - \sum_{n=1}^{T}\delta^{n-1}\bar{r} + \delta^T CT\big(1-\varepsilon'\big) > \sum_{n=1}^{T}\delta^{n-1}\bar{r} + \delta^T C\big(1-2\varepsilon'\big) + \delta^T\bar{r}, \tag{26}$$

and for each sequence $(u_1, \dots, u_T)$,

$$\left|\frac{1-\delta}{1-\delta^T}\sum_{n=1}^{T}\delta^{n-1}u_n - \frac{1}{T}\sum_{n=1}^{T}u_n\right| < \varepsilon' \max(u_1, \dots, u_T).$$

Let $\delta > \delta_*$ be arbitrary, define $x_* : Y^T \to \mathbf{R}^I$ by $x^i_*(\cdot) = 0$, and define $x^j_*(\vec{y}) = -CT$ if $D^j > \varepsilon'$ and $x^j_*(\vec{y}) = 0$ otherwise, so that $\|x_*(\cdot)\| \le \frac{1}{2}\kappa_* T$ and $x^i_*(\cdot) \ge 0$.

Given $s \in S$, let σ_s be any sequential equilibrium of $G(s, \delta, x_*)$, with payoff $\gamma_s(\sigma_s) \in \mathbf{R}^I$. By the choice of C and δ_*, one has $\mathbf{P}_{s,\sigma}(D^j > \varepsilon') < 2\varepsilon'$ for every $j \neq i$.[61] Therefore, by Claim 10, one has

$$\mathbf{E}_{s,\sigma}\left[\frac{1}{T}\sum_{n=1}^{T} r^i\left(s^i_n, a_n\right)\right] < w^i_k + \varepsilon' + 2\bar{r}(I+2)\varepsilon',$$

which implies

$$\gamma^i_s(\sigma_s) < w^i + 2\bar{r}(I+4)\varepsilon'.$$

Since $\mathbf{P}_s(D^j > \varepsilon') < 2\varepsilon'$ for each $j \neq i$, it follows from the specification of x_* and δ_* that $\|\gamma_s(\sigma_s)\| \le 14\bar{r}$. Set then $\bar{x}^j(s) := \max_{s' \in S} \gamma^j_{s'}(\sigma_{s'}) - \gamma^j_s(\sigma_s)$ for each $s \in S$ and $j \in I$, and

$$x(s, \vec{y}) := x_*(\vec{y}) + \bar{x}(s).$$

Plainly, σ_s is still a sequential equilibrium of $G(s, \delta, x)$ for each s, and the payoff vector induced by σ_s is now independent of s. Moreover, since $0 \le \bar{x}(\cdot) \le 14\bar{r}$ and by the choice of ε', both (b) and (c) hold as well. *Q.E.D.*

We denote by $\mathcal{G}^i_\varepsilon$ the compact set of all accumulation points of such equilibrium payoffs $\gamma \in \mathbf{R}^I$ as $\delta \to 1$. Before we move on to the equilibrium construction, two remarks are in order. Note first that property (c) can be strengthened to $x^i(\cdot) \ge \varepsilon''$, where $\varepsilon'' < \varepsilon$ is arbitrary. (Indeed, for given $0 < \varepsilon'' < \varepsilon$, it suffices to first apply the current version of Lemma 13 with $\varepsilon - \varepsilon''$ and then add ε'' to x^i.)

Because of irreducibility, there is a constant c (which only depends on the primitives of the game) such that for $j \in I$, $s \in S$, and $t^j \in S^j$, the highest payoffs achievable by j against σ^{-j}_s in the two games $G(t^j, s^{-j}, \delta, x(s, \cdot))$ and $G(s, \delta, x(s, \cdot))$ differ by at most $(1-\delta)c$. Since the latter payoff is equal to γ^j, the former does not exceed $\gamma^j + (1-\delta)c$. Since γ^j is also the payoff induced by $\sigma_{t^j, s^{-j}}$ in the game $G(t^j, s^{-j}, \delta, x(t^j, s^{-j}, \cdot))$, this implies that the benefit to player j of pretending that his initial state is s^j when it is t^j is bounded above by $(1-\delta)c$.

[61] Indeed, by (26), any strategy $\tilde{\sigma}^j$ such that $\mathbf{P}_{\tilde{\sigma}^j, \sigma^{-j}}(D^j > \varepsilon') < \varepsilon'$ is strictly preferred to any strategy $\tilde{\sigma}^j$ such that $\mathbf{P}_{\tilde{\sigma}^j, \sigma^{-j}}(D^j > \varepsilon') > 2\varepsilon'$, and $\sigma^j_{k,T}$ satisfies the former condition by Claim 9.

F.2. *Equilibrium Construction*

We only provide a sketch. We start as in Section B.2. To unify notations, we set $\hat{k}_1(\lambda) = \bar{k}_1(\lambda)$ for $\lambda \neq -e^i$ and $\hat{k}_1(-e^i) = -w^i$ for $i \in I$. Since $\hat{k}_1(\cdot)$ is lower semicontinuous on Λ, there exists $\varepsilon_0 > 0$ such that

$$\forall \lambda \in \Lambda, \quad \max_{Z_\eta} \lambda \cdot z + 2\varepsilon_0 < \hat{k}_1(\lambda).$$

For each player i, we apply Lemma 13 with $0 < \varepsilon_0' < \varepsilon_0$ (so that $x^i(\cdot) \geq \varepsilon_0'$) and get κ_* and δ_*. We next pick $\varepsilon_0'' < \frac{\varepsilon_0'}{\kappa_*}$. With these choices, for fixed i and $\delta > \delta_*$, the payoff vector γ and the transfers x satisfy $\gamma^i < w^i + \varepsilon_0$, $\|x\| < \kappa_* T$ and $\lambda \cdot x_*(\cdot) \leq 0$ whenever $\|\lambda - (-e^i)\| < \varepsilon_0''$.

Parameters are chosen as follows. We first pick the parameter $0 < \beta < \frac{1}{2}$ of the length of transition phases; then we choose κ to be large enough. Next, as before, pick $\varepsilon > 0$ small enough. Finally, we choose $\bar{\delta} < 1$ high enough. Computations are highly similar to those in Sections A.1.2 and A.1.3. They are, therefore, omitted, and we do not list conditions to be satisfied by κ, ε, and $\bar{\delta}$.

We let $z \in Z$ be given and let $\pi_1 \in \bigtimes_i \Delta(S^i)$ be the distribution of the initial state. The play is divided into a sequence of phases, with odd phases being transition phases. Slight adjustments in the strategies are needed (as compared with Section A.1.2), and we detail the updating from one transition phase to the following transition phase. The transition phase k starts with a target payoff $z_{(k)}$, which is deduced from past public play. We set $(\rho_{(k)}, x_{(k)}) = (\rho_{\text{ext},0}, x_0)$, $v_{(k)} := \mathbf{E}_{\mu_{\rho_{\text{ext},0}}}[r(s, a) + x_0(s, k_0, y)]$ and $\theta_{(k)} := \theta_0$. In each round, the p.r.d. chooses with probability $\xi_* := (1 - \delta)^\beta$ whether to start a new phase. In the first round $n = \tau_{(k+1)}$ of the following phase $k + 1$, we first define the auxiliary target $w_{(k+1)}$ according to

$$\xi_* w_{(k+1)} + (1 - \xi_*) z_{(k)} = \frac{1}{\delta} z_{(k)} - \frac{1-\delta}{\delta} v_{(k)} + \frac{1-\delta}{\delta} x_{(k)}(\omega_{\text{pub},n-1})$$

and then we apply Lemma 1 with $z := w_{(k)}$ to get $\lambda_{(k+1)}$.

If $\|\lambda_{(k+1)} - (-e^i)\| \geq \varepsilon_0''$ for all i, we apply Lemma 2 to get $(v_{(k+1)}, \rho_{(k+1)}, x_{(k+1)}) \in \mathcal{S}$, and finally update $z_{(k+1)}$ as

$$\begin{aligned} z_{(k+1)} = w_{(k+1)} + (1-\delta)\biggl(\biggl(1 + \frac{1-\delta}{\delta\xi}\biggr)\theta_{(k)}(m_{n-1}, m_n) \\ - \theta_{(k+1)}(\omega_{\text{pub},n-1}, m_n)\biggr). \end{aligned}$$

Then in each round, the p.r.d. chooses with probability ξ whether to start a new phase. In round $\tau_{(k+2)}$, the auxiliary target will be updated to $w_{(k+2)}$ according to (4) and $z_{(k+2)}$ in the following transition phase is defined by (5).

If instead $\|\lambda_{(k+1)} - (-e^i)\| < \varepsilon_0''$ for some i, we apply Lemma 13 with player i, and get $x : S \times Y^T \to \mathbf{R}^I$ and γ. We set $v_{(k+1)} = \gamma$ and $x_{(k+1)} = x$. In that case, the duration of phase $k+1$ is T. In round $\tau_{(k+2)} := \tau_{(k+1)} + T$, we set

$$z_{(k+2)} = \frac{1}{\delta^T} z_{(k+1)} - \frac{1-\delta^T}{\delta^T} v_{(k+1)} + (1-\delta) x_{(k+1)}(m_{\tau_{(k+1)}}, y_{\tau_{(k+1)}}, \dots, y_{\tau_{(k+1)}+T-1}).$$

That this recursive construction is well defined follows as in Lemma 5.

Under σ, players report truthfully and play $\rho_{(k)}$ in any phase k that is not a punishment phase. If $\|\lambda_{(k)} - (-e^i)\| < \varepsilon_0''$, we let $\sigma_{(k)}$ be a sequential equilibrium in $G(m_{\tau_{(k)}}, \delta, x_{(k)})$ with payoff $v_{(k)}$. Under σ, player j plays $\sigma^j_{(k)}$ if his report in round $\tau_{(k)}$ is truthful and otherwise plays a (sequentially) best reply to $\sigma^{-j}_{(k)}$ in the game $G(s^j_{\tau_{(k)}}, m^{-j}_{\tau_{(k)}}, \delta, x_{(k)})$.

As in Section A.1.3, one can establish that the continuation payoff under σ is equal to $z_{(k)}$ at the beginning of a punishment phase and equal to $z_{(k)} + (1-\delta)\theta_{(k)}$ in any round that does not belong to a punishment phase.

That a player cannot profitably deviate at the action step follows from the definition of σ in a punishment phase and follows as in Theorem 2 otherwise. That a player cannot profitably deviate at the reporting step of a nontransition phase is clear during punishment phases since reports are ignored, and otherwise follows as before.

Consider finally the reporting step in a round n belonging to a transition phase. In the specific case where n is the first round following a punishment phase, reports are ignored, and the action being played is $\bar{a}_0$; hence truthful reporting is trivially optimal. Otherwise, the belief of player j over S^{-j} has full support, and the optimality of truth-telling follows along earlier lines, using that (i) $(\rho_{\text{ext},0}, x_0)$ is strictly truthful, and that (ii) the (ex post) marginal benefit of having misreported, conditional on the p.r.d. choosing to start a new phase, is at most on the order of $(1-\delta)$; see the remark at the end of Section F.1.2.

REFERENCE

GOSSNER, O. (1995): "The Folk Theorem for Finitely Repeated Games With Mixed Strategies," *International Journal of Game Theory*, 24, 95–107. [12]

Cowles Foundation, Yale University, 30 Hillhouse Ave., New Haven, CT 06520, U.S.A.; johannes.horner@yale.edu,

Dept. of Economics, Faculty of Arts and Social Sciences, National University of Singapore, AS2-06-11, 1 Arts link, Singapore 117570, Singapore; ecsst@nus.edu.sg,

and

HEC Paris, 1 Rue de la Libération, 78351 Jouy-en-Josas, France; vieille@hec.fr.

Manuscript received January, 2014; final revision received April, 2015.