B Supplement to "Limit Points of Endogenous Misspecified Learning"

B.1 Omitted Lemmas and Examples

Lemma 1. *For every $a \in A$ and $\varepsilon > 0$, $\hat{\Theta}(a)$ defined in equation (1), $\hat{\Theta}_a(a)$, $\hat{\Theta}^\varepsilon(a)$ defined in equation (2), and $\Delta(\hat{\Theta}(a))$ are compact.*

Proof. Compactness of $\hat{\Theta}(a)$ follows from Assumption 1 and Theorem 2.43 of Aliprantis and Border (2013). Since the projection map is continuous, and $\hat{\Theta}_a(a)$ is the projection of $\hat{\Theta}(a)$, $\hat{\Theta}_a(a)$ is compact as well. Since $\hat{\Theta}_a(a)$ is closed, it immediately follows that $\hat{\Theta}^\varepsilon(a)$ is closed as well, henceforth compact. Given the compactness and separability of $\hat{\Theta}(a)$, $\Delta(\hat{\Theta}(a))$ is compact by, e.g., Theorem 6.4 in Parthasarathy (2005). ■

Lemma 7. *Fix $q \in \Delta(Y)$ with $\operatorname{supp} q \subseteq \operatorname{supp} p_a^*$ and a compact set $C \subseteq \Delta(Y)$ such that all the elements of C are absolutely continuous with respect to p_a^*. Then there exists a $K > 0$ such that for every $f' \in U_\varepsilon(q, p_a^*, \eta)$ with $\operatorname{supp} f \subseteq \operatorname{supp} p_a^*$*

$$|\min_{q' \in C} H\left((1-\eta)p_a^* + \eta q, q'\right) - H\left((1-\eta)p_a^* + \eta q, q\right) - \min_{q' \in C} H\left(f, q'\right) + H\left(f, q\right)| \leqslant K\varepsilon\,.$$

Proof. First, notice that by the Maximum Theorem,

$$\hat{C}(\eta, \varepsilon) := \bigcup_{f \in U_\varepsilon(q, p_a^*, \eta) : \operatorname{supp} f \subseteq \operatorname{supp} p_a^*} \operatorname*{argmin}_{q' \in C} H\left(f, q'\right)$$

is a compact-valued and upper-hemicontinuous correspondence. So, if we let

$$\hat{C} := \bigcup_{\varepsilon \in [0,1]} \bigcup_{\eta \in [0,1]} \hat{C}(\eta, \varepsilon),$$

there is a $K_1 > 0$ such that $\max_{y \in \operatorname{supp} p_a^*} \max_{q' \in \hat{C}} |\log q'(y)| < K_1$.

Then we have that for every $\eta \in [0,1]$, $\varepsilon > 0$, and $f \in U_\varepsilon(q, p_a^*, \eta) : \operatorname{supp} f \subseteq \operatorname{supp} p_a^*$

$$\begin{aligned} & |\min_{q' \in C} H\left((1-\eta)p_a^* + \eta q, q'\right) - H\left((1-\eta)p_a^* + \eta q, q\right) - \min_{q' \in C} H\left(f, q'\right) + H\left(f, q\right)| \\ \leqslant\ & |\min_{q' \in C} H\left((1-\eta)p_a^* + \eta q, q'\right) - \min_{q' \in C} H\left(f, q'\right)| + 2\varepsilon \max_{y \in \operatorname{supp} p_a^*} |\log q\left(y\right)| \\ \leqslant\ & |2K_1\varepsilon| + 2\varepsilon \max_{y \in \operatorname{supp} p_a^*} |\log q\left(y\right)|, \end{aligned}$$

where the inequalities follows from $||f - (1-\eta)p_a^* + \eta q|| \leqslant \varepsilon$, and the definition of K_1. Thus $K: = 2(K_1 + \max_{y \in \operatorname{supp} p_a^*} |\log q(y)|) > 0$ satisfies the statement of the lemma. ■

Computations for Example 1

The monopolist's payoff function when valuation are uniformly distributed on $[0,8]$ is $\mathbb{E}[u(a,y)] = \frac{8-a}{8}a$, so the unique optimal price from the set $\{3,4,5,6,7\}$ equals $a = 4$. If valuations are uniformly distributed on $[2,10]$, the payoff function is $\frac{10-a}{8}a$, so the unique optimal price is $a = 5$.

Let $p^L = (\frac{8-a}{8})_{a\in\{3,4,5,6,7\}}$ be the vector of conditional probabilities when the demand is low and $p^H = (\frac{10-a}{8})_{a\in\{3,4,5,6,7\}}$ be the vector of conditional probabilities when the demand is high. It is easy to check that the KL minimizers are given by

$$\hat{\Theta}(3) = \{p^H\}; \quad \hat{\Theta}(4) = \{p^H\}; \quad \hat{\Theta}(5) = \{p^L, p^H\}; \quad \hat{\Theta}(6) = \{p^L\}; \quad \hat{\Theta}(7) = \{p^L\}.$$

Thus $a = 5$ is the only pure BN-E. Note that $a = 5$ is not a uniform BN-E, because at the low belief the myopically optimal action is 4.

Example 6. *This example shows that Theorem 1 does not hold without Assumption 1(ii). Let the action space be $\{a,b\}$, the outcome space be $\{0,1\}$, and suppose the agent correctly believes that the action has no impact on the outcome distribution, so that each action dependent outcome distribution is indexed by a number in $(0,1)$ corresponding to the probability of outcome 1. Finally, let $p^* = \frac{1}{2}$.*

Assume that the agent assigns positive probabilities to the following countable set:

$$\left\{\frac{3}{4}\right\} \cup \left\{\frac{1}{4} - \frac{1}{n^2} : n \geqslant 3\right\},$$

where distributions are indexed by the probability that they assign to outcome 1. Note that $\frac{1}{4}$ is in Θ even though it doesn't exactly correspond to any of the agent's conceivable outcome distributions. Let $p(n) = \frac{1}{4} - \frac{1}{n^2}$.

Finally, suppose that the agent's utility function is given by $u(a,0) = 0 = u(b,1), u(a,1) = 1, u(b,0) = 4/5$. Then b is not preferred to a for any beliefs with $\nu(\{3/4\}) > 1/2$ and it is strictly preferred to a if $\nu(\{3/4\}) < 1/3$. Then a is a BN-E but not a uniform BN-E, yet play can converge to it with positive probability from a prior μ_0 we specify below.

In the claim below we show that for every $n \in \mathbb{N}$ there exists a $l_n > 0$ such that

$$1 \leqslant p^*(1)\left(\frac{\frac{3}{4}}{p(n)(1)}\right)^{l_n} + p^*(0)\left(\frac{\frac{1}{4}}{p(n)(0)}\right)^{l_n}.$$

Then by Dubins' upcrossing inequality[30]*, for all K_1, and K_2 there exists $C_n \leqslant \frac{\frac{1}{n^2}}{2\sum_{n=3}^{\infty}\frac{1}{n^2}}$ such that if $\mu_0(p(n)) \leqslant C_n$ and $\mu_0\left(\frac{3}{4}\right) > \frac{1}{2}$, the probability that $\limsup_t \frac{\mu_t(p(n))}{\mu_t\left(\frac{3}{4}\right)} > \frac{1}{n^2}K_1$ is smaller then $\frac{1}{n^2}K_2$. Let $\mu_0(p(n)) = C_n$ and $\mu_0\left(\frac{3}{4}\right) = 1 - \sum_{n=3}^{\infty} C_n > \frac{1}{2}$, $K_2 < \frac{1}{\sum_{n=3}^{\infty}\frac{1}{n^2}}$ and $K_1 < \frac{1}{2\sum_{n=3}^{\infty}\frac{1}{n^2}}$. By the union bound with probability*

$$1 - K_2 \sum_{n=3}^{\infty} \frac{1}{n^2} > 0$$

we have that

$$\limsup_t \frac{\sum_{n=3}^{\infty} \mu_t(p(n))}{\mu_t\left(\frac{3}{4}\right)} \leqslant \sum_{n=3}^{\infty} \limsup_t \frac{\mu_t(p(n))}{\mu_t\left(\frac{3}{4}\right)} \leqslant K_1 \sum_{n=3}^{\infty} \frac{1}{n^2} < \frac{1}{2}.$$

Claim 3. *Notice that the outcome distribution most favorable to action b and least favorable to action a is $p(3) = 1/4 - 1/9 = 5/36$. Therefore, if $\nu_t(\{3/4\}) > 1/2$,*

$$\int_{\Delta(Y)} \mathbb{E}_p[u(a,y)]\, d\nu(p) \geqslant \sum_{n=3}^{\infty} p(n)u(a,1)\nu(\{p(n)\}) + \frac{3}{4}u(a,1)\nu(\{3/4\})$$
$$\geqslant \frac{5}{36}u(a,1)(1-\nu(\{3/4\})) + \frac{3}{4}u(a,1)\nu(\{3/4\}) > 4/9$$

and

$$\begin{aligned}\int_{\Delta(Y)} \mathbb{E}_p[u(b,y)]\, d\nu(p) &\leqslant \sum_{n=3}^{\infty}(1-p(n))u(b,0)\nu(\{p(n)\}) + \frac{1}{4}u(b,0)\nu(\{3/4\}) \\ &\leqslant \frac{31}{36}u(b,0)(1-\nu(\{3/4\})) + \frac{1}{4}u(b,0)\nu(\{3/4\}) < 4/9.\end{aligned}$$

[30]See, e.g., page 27 of Neveu (1975)

If $\nu_t(\{3/4\}) < 1/3$,

$$\begin{aligned}\int_{\Delta(Y)} \mathbb{E}_p[u(a,y)]\,d\nu(p) &\leqslant \sum_{n=3}^{\infty} p(n)u(a,1)\nu(\{p(n)\}) + \frac{3}{4}u(a,1)\nu(\{3/4\}) \\ &\leqslant \frac{1}{4}u(a,1)(1-\nu(\{3/4\})) + \frac{3}{4}u(a,1)\nu(\{3/4\}) < \frac{5}{12}\end{aligned}$$

and

$$\begin{aligned}\int_{\Delta(Y)} \mathbb{E}_p[u(b,y)]\,d\nu(p) &\geqslant \sum_{n=3}^{\infty}(1-p(n))u(b,0)\nu(\{p(n)\}) + \frac{1}{4}u(b,0)\nu(\{3/4\}) \\ &\geqslant \frac{3}{4}u(b,0)(1-\nu(\{3/4\})) + \frac{1}{4}u(b,0)\nu(\{3/4\}) = \frac{7}{15}.\end{aligned}$$

Finally, notice that

$$\begin{aligned}1 &\leqslant p^*(1)\left(\frac{\frac{3}{4}}{p(n)(1)}\right)^{l_n} + p^*(0)\left(\frac{\frac{1}{4}}{p(n)(0)}\right)^{l_n} \\ &= \frac{1}{2}\left(\frac{\frac{3}{4}}{\frac{1}{4}-\frac{1}{n^2}}\right)^{l_n} + \frac{1}{2}\left(\frac{\frac{1}{4}}{\frac{3}{4}+\frac{1}{n^2}}\right)^{l_n}\end{aligned}$$

where

$$l_n = \frac{\log\left(1-\frac{1}{\frac{4}{n^2}+3}\right)}{\log\left(\frac{1}{1-\frac{4}{n^2}}\right)+\log 3} > 0.$$

B.2 The Role of Assumption 1(i)

All results in the paper except the non-myopic part of Theorem 1 continue to hold under a weaker version of Assumption 1(i):

Assumption 1(i′) For all $p \in \Theta$ and $\varepsilon > 0$, there exists $p' \in \Theta$ with $||p' - p|| < \varepsilon$ such that for all $a \in A$, if $p^*_a(y) > 0$ then $p'_a(y) > 0$.

Assumption 1(i′) implies that the support of the belief does not change after a finite number of observations. This is the only consequence of Assumption 1(i) that is used in any of the proofs, except for establishing Claim 1 in the proof of Theorem 1 when the agent is not myopic.[31]

The next example shows that without Assumption 1(i′), limit points need not be BN-E.

[31] When the agent is myopic Claim 1 continues to hold under Assumption 1(i′).

Example 7 (Role of Assumption 1(i′)). *Suppose there are two actions a and b, and two outcomes $Y = \{0, 1\}$, and let $u(a, 0) = u(b, 1) = 1 - u(a, 1) = 1 - u(b, 0)$. Identify the elements of $\Delta(Y)$ with the probability they assign to outcome 1, and let $p_a^* = \frac{2}{3}$ and $p_b^* = 1$. Suppose that the agent believes that the outcome distribution does not depend on the action, and that $\Theta = \{\frac{1}{3}, 1\}$. Here b is the unique BN-E, and it is uniformly strict. However, if the prior assigns sufficiently high probability to $1/3$, the agent will start playing a, and with positive probability they will observe outcome 0 in the first period. But after this observation, the posterior assigns probability 1 to $p = 1/3$ and the action converges to a.*

When we weaken Assumption 1(i) to (i′) and allow the supports the various outcome distributions to differ, we need to generalize the definition of observational equivalence as follows:

Definition 14. Two outcome distributions p and p' are *observationally equivalent under action a* if $p_a(y) = p'_a(y)$ for all $y \in \operatorname{supp} p_a^*$.

Thus we now say that two beliefs are observationally equivalent under a if they assign the same probability to each outcome that realizes with positive probability. This definition is equivalent to the one in the main text under Assumption 1(i).

The reason Theorem 1 only holds for myopic agents when we weaken (i) to (i′) is that Claim 1 can fail. The intuition is that even if the agent plays a many times, they may still think that playing a again will give them a non-trivial amount of information, as in the next example.

Example 8. *Let $A = \{a, b, c\}$, $Y = \{0, \bar{y}, y'\}$, and $\Theta = \{\bar{p}, p'\}$. Suppose that $\bar{p}_c(\bar{y}) = 1 - \bar{p}_c(0) = 0.9 = 1 - p'_c(0) = p'_c(y')$ and that $u(c, y) = -0.1$ for all $y \in Y$. Thus, the agent thinks that by playing c they pay a small cost, and with a very high probability they discover the correct model for sure, and otherwise receive an uninformative signal.*

For action b suppose that $\bar{p}_b(0) = 1 = p'_b(0)$ and $u(b, y) = 0$ for all $y \in Y$. That is, the agent thinks that action b is uninformative but safe.

Finally the agent thinks that action a produces the same information of action c but its payoffs are riskier: $\bar{p}_a(\bar{y}) = 1 - \bar{p}_a(0) = 0.9 = 1 - p'_a(0) = p'_a(y')$ $u(a, \bar{y}) = -100$ and $u(a, y') = 1$.

Here, c is not a a BN-E, because it is weakly dominated by action b, and it is never a myopic best reply. However, suppose that $p_c^*(0) = 1$, that the agent starts with a uniform prior over Θ, and the discount factor $\beta = \frac{1}{2}$. Then every optimal policy prescribes starting with action c to get information, and then switching to a forever after observing y', to b

forever after observing $\bar{y}$ and trying c again after observing 0. Since $p^*_c(0) = 1$, the agent will continue to use c forever, because the believe that with high probability the true outcome distribution will be revealed next period.

Assumption 1(i) guarantees that when beliefs concentrate around a set of of outcome distributions that are observationally equivalent under a, i.e. $\nu \in \Delta(\mathcal{E}(a)(p))$ for some $p \in \Theta$, the experimentation value of a is weakly lower than that of some other action. This fact is used in Claim 1 to show that $G(\nu) > 0$ for every $\nu \in \Delta(\mathcal{E}(a)(p))$. Claim 1 holds under Assumption 1(i′) for myopic agents because for these agents all actions have 0 experimentation value.

Assumption 1(i′) is still sufficient for all the problems considered in Section 4.2. More generally, (i′) is sufficient when paired with with this additional assumption:

Assumption 2. $p, p' \in \mathcal{E}(a)(p) \Rightarrow p_a(y) = p'_a(y)$ for all $y \in Y$.

This assumption is trivially satisfied if all beliefs in the support of the agent's subjective prior assign positive probability only to signals which objectively occur with positive probability, i.e. $p_a(y) > 0 \Rightarrow p^*_a(y) > 0$ for all $p \in \Theta, a \in A$.

B.3 Extensions to Signals

Here we expand the probability space of our basic model in the obvious way: The sample space $\Omega = S^\infty \times (Y^\infty)^A$ consists of infinite sequences of signal and action dependent outcome realizations $(s_k, x_{a,s',k})_{k\in\mathbb{N}, a\in A, s'\in S}$ and $x_{a,s',k}$ determines the outcome when the agent takes the action a for the k-th time after s. Formally, we consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is the discrete sigma algebra and the probability measure $\mathbb{P}$ is the product measure induced by independent draws (across signal, actions, and time) according to p^*.

We denote the outcome observed by the agent in period t after action a_t by $y_t = x_{a_t,s_t,k}$, where k is the number of times the agent has taken action a_t after signal s_t up and including period t. A (pure) policy $\pi : \bigcup_{t=0}^\infty S^{t+1} \times A^t \times Y^t \to A$ specifies an action for every history $(s_1, a_1, y_1, s_2, a_2, y_2, \ldots, s_t, a_t, y_t, s_{t+1})$, and an initial action a_1. Throughout, we denote by $a_{t+1} = \pi(s^{t+1}, a^t, y^t)$ the action taken in period t where (s^{t+1}, a^t, y^t) is a sequence of realized signals, actions, and outcomes. For every $p, p' \in \Theta \cup \{p^*\}$, denote the supnorm distance between p and p':

$$||p - p''|| = \max_{s\in S, a\in A, y\in Y} |p_{a,s}(y) - p'_{a,s}(y)|.$$

Given our finite dimensionality assumption, the maximand depends on s only through the finite partition Ξ, so the supremum is attained. In this setting, a policy π converges to a strategy σ if there exists a T such that for all $t \geqslant T$, $\xi \in \Xi$, $p \in \Theta \cup \{p^*\}$ and $y \in Y$

$$\sum_{a \in A} \zeta\left(\left\{s \in \xi : \pi\left(a^T, y^T, s\right) = a\right\}\right) p_{a,s}(y) = \sum_{a \in A} \zeta\left(\left\{s \in \xi : \sigma(s) = a\right\}\right) p_{a,s}(y)$$

that is, there is finite time convergence over the behavior in the finite dimensional partition of signals considered by the agent. This restriction is without loss of generality if S is finite.

Lemma 10. *For every $\sigma \in A^S$ and $\varepsilon > 0$, $\hat{\Theta}(\sigma)$ and $\hat{\Theta}^{\varepsilon}(\sigma)$ are compact.*

Proof of Lemma 10. Compactness of $\hat{\Theta}(\sigma)$ follows from the generalization of Weierstrass Theorem to lower-semicontinuous functions (see e.g. Theorem 2.43 in Aliprantis and Border (2013)). Since the projection map is continuous it follows that $\hat{\Theta}^{\varepsilon}(\sigma)$ is closed, so it is compact. ■

Now we extend Lemma 2 to the case where the agent observes signals and has finite-dimensional beliefs. Since we restricted the policy function of the agent to be measurable in their beliefs, the set of policy functions is

$$\Pi = \left(A^S\right)^{\bigcup_{t=0}^{\infty}\left(A^t \times Y^t \times \Xi^t\right)}.$$

We endow the set A^S of measurable maps from S to A with the metric

$$d_\zeta\left(\sigma, \sigma'\right) = \zeta\left(\left\{s \in S : \sigma(s) \neq \sigma'(s)\right\}\right).$$

Then Π is the (countable) product space of measurable maps with index set $\bigcup_{t=0}^{\infty}\left(A^t \times Y^t \times \Xi^t\right)$.

Lemma 11. *Π is compact in the product topology, and for every $\nu \in \Delta(\Theta)$, $V(\cdot, \nu)$ is continuous with respect to the product topology.*

Proof. By Tychonoff's theorem A^S is compact in the product topology. Suppose that $(\sigma_n)_{n \in \mathbb{N}}$ converges pointwise to σ, and let $C_n = \{s \in S : \forall m \geqslant n, \sigma_m(s) = \sigma(s)\}$. We have that $C_n \uparrow S$,

$$d_\zeta(\sigma_n, \sigma) = \zeta\left(\left\{s \in S : \sigma_n(s) \neq \sigma(s)\right\}\right) \leqslant 1 - \zeta(C_n)$$

and so $d_\zeta(\sigma_n, \sigma) \to 0$. Thus the product topology is finer than the topology induced by d_ζ, and so A^S is also compact in (A^S, d_ζ). Applying Tychonoff's theorem again, Π is compact in the product topology. Continuity follows from the fact that for every period $t \in \mathbb{N}$ the set $(A^t \times Y^t \times \Xi^t)$ is finite, and discounting. ■

We next generalize a couple of definitions given in the text to allow for signals. For every strategy σ and action contingent outcome distribution p, we let

$$p_\sigma = \int_S p^*_{\sigma(s),s}(\cdot) d\zeta(s)$$

denote the distribution over outcomes induced by the use of strategy σ. Let $\hat{\Theta}^\varepsilon(\sigma)$ denote the conceivable outcome distributions that are ε close to one of the elements of $\Theta(a)$:

$$\hat{\Theta}^\varepsilon(\sigma) = \{p \in \Theta : \exists p' \in \hat{\Theta}(\sigma), ||p'_\sigma - p_\sigma|| \leqslant \varepsilon\}.$$

Similarly, we denote the set of beliefs over conceivable distributions that assign at least probability $1 - \varepsilon$ to $\hat{\Theta}^\varepsilon(\sigma)$ by

$$M_{\varepsilon,a} = \{\nu \in \Delta(\Theta) \colon \nu(\hat{\Theta}^\varepsilon(\sigma)) \geqslant 1 - \varepsilon\}.$$

Next we extend Lemma 3 to this setting.

Lemma 12. *If σ is a uniformly strict BN-E, then for every optimal policy π and every $\lambda \in \mathbb{R}_{++}$ there exists an $\hat{\varepsilon} > 0$ such that for all $\varepsilon < \hat{\varepsilon}$*

$$\nu \in M_{\varepsilon,\sigma} \implies |\zeta\left(\{s \in S : \pi\left(\nu, s\right) = a\}\right) - \zeta\left(\{s \in S : \sigma\left(s\right) = a\}\right)| < \lambda. \tag{5}$$

Proof. Fix a belief $\nu \in M_{\varepsilon,\sigma}$. Let π^σ denote the policy that always plays σ, and let Π_λ denote the set of policy functions $\tilde{\pi}$ such that:

$$|\zeta\left(\{s \in S : \tilde{\pi}\left(\nu, s\right) = a\}\right) - \zeta\left(\{s \in S : \sigma\left(s\right) = a\}\right)| \geqslant \lambda$$

Define $G(\varepsilon)$ as the gain from playing σ forever instead of using (one of) the best policies $\tilde{\pi} \in \Pi_\lambda$

$$G(\varepsilon) = \min_{\tilde{\pi} \in \Pi_\lambda} \min_{\nu \in M_{\varepsilon,a}} \left(V\left(\pi^a, \nu\right) - V\left(\tilde{\pi}, \nu\right)\right).$$

Notice that by Lemma 11 the space of the policy functions endowed with the product topology is compact. Since the subset of policy functions that satisfy 5 is closed, this subset is compact as well. Moreover, given that $\beta \in (0, 1)$, the value function is continuous at infinity, and therefore $V(\pi^a, \nu) - V(\cdot, \nu)$ is a continuous function of the policy. Notice also that since $\mathbb{E}_{p,\pi}\left[\sum_{t=1}^{\infty}\left[\beta^{t-1}u(a_t, y_t)\right]\right]$ is continuous in p, $V(\pi^a, \cdot) - V(\tilde{\pi}, \cdot)$ is continuous in ν, so since $\varepsilon \rightarrow M_{\varepsilon,\sigma}$ is an upper hemicontinuous and compact valued correspondence, from the Maximum Theorem G is continuous in ε. Since σ is a uniformly strict BN-E, $G(0) > 0$, and there is an $\hat{\varepsilon}$ such that if $\varepsilon \leqslant \hat{\varepsilon}$, $G(\varepsilon) > 0$. This implies that for any optimal policy π it must be such that $\nu \in M_{\varepsilon,\sigma}$ implies that π satisfies equation (5), which proves the lemma. ■

Lemma 13. *Fix a strategy σ and $\varepsilon > 0$. There exists an $\bar{l} > 0$ such that for all $l \leqslant \bar{l}$ for every KL minimizer $q \in \hat{\Theta}(\sigma)$, every $p' \notin \hat{\Theta}^{\varepsilon}(\sigma)$, and every $\sigma' \in B_l(\sigma)$ we have*

$$f_l(\sigma', q, p') := \sum_{y \in Y} p_{\sigma'}(y) \left(\frac{q_{\sigma'}(y)}{p'_{\sigma'}(y)}\right)^l > 1\,.$$

Proof. As noted by FII in their Lemma 3, for each KL minimizer $q \in \hat{\Theta}(\sigma)$ and every outcome distribution $p' \notin \hat{\Theta}(\sigma)$ there exists an $l(\sigma, q, p')$ such that $f_l(\sigma, q, p') > 1$ for all $l \leqslant l(\sigma, q, p')$. They also pointed out that for all $q, q' \in \Theta$, and $\sigma' \in A^S$, if $\hat{l} > l$ and $f_l(\sigma', q, q') \leqslant 1$, then $f_{\hat{l}}(\sigma', q, q') \leqslant 1$. We will now prove that there exists a uniform l that works for every $q \in \hat{\Theta}(\sigma)$ and $p' \in \hat{\Theta}^{\varepsilon}(\sigma)$, and every strategy σ' sufficiently close to σ.

Suppose by way of contradiction that there was no $\bar{l} > 0$ such that for all $l \leqslant \bar{l}$, $f_l(\sigma', q, p') > 1$ for all $q \in \hat{\Theta}(\sigma)$ and $p' \notin \hat{\Theta}^{\varepsilon}(\sigma)$, $\sigma' \in B_l(\sigma)$. Then we can define a sequence (σ_n, q_n, p'_n) such that $f_{\frac{1}{n}}(\sigma_n, q_n, p'_n) \leqslant 1$, and $\sigma_n \in B_{1/n}(\sigma)$. The sequential compactness of $A^S \times \hat{\Theta}(\sigma) \times cl\{p \in \Delta(\Theta) \colon p_a \notin \hat{\Theta}^{\varepsilon}(\sigma))\}$ derived in Lemma 10 guarantees that this sequence has an accumulation point (σ, q, p'). However, for, $n > \frac{1}{l(\bar{p}, p')}$, $f_{\frac{1}{n}}(\sigma_n, q_n, p'_n) \leqslant 1$ implies $f_{l(q,p')}(\sigma_n, q_n, p'_n) \leqslant 1$, but then the lower semicontinuity of $f_{l(q,p')}$ at (σ, q, p') leads to a contradiction with $f_{l(q,p')}(\sigma, q, p') > 1$. ■

Lemma 14. *Let $p, p', p^* \in \Delta(Y)$, and $l \in (0, 1)$ be such that*

$$\sum_{y \in Y} p^*(y) \left(\frac{p(y)}{p'(y)}\right)^l > 1. \tag{6}$$

Then there is $\varepsilon' > 0$ such that for all $\nu \in \Delta(\Delta(Y))$, if we let

$$\nu(C \mid y) = \frac{\int_{q \in C} q(y) d\nu(q)}{\int_{q \in \Delta(Y)} q(y) d\nu(q)},$$

then

$$\sum_{y\in Y} r(y)\left[\left(\frac{\nu(B_{\varepsilon'}(p)\mid y)}{\nu(B_{\varepsilon'}(p')\mid y)}\right)^{l}\right] \geqslant \left(\frac{\nu(B_{\varepsilon'}(p))}{\nu(B_{\varepsilon'}(p'))}\right)^{l}$$

for all $r \in B_{\epsilon'}(p^*)$.

Proof. The lemma is trivially true if $\nu(B_\varepsilon(p')) = 0$ for some ε. Therefore, without loss of generality, we can assume that $\nu(B_\varepsilon(p')) > 0$ for all ε. Let $C_\varepsilon = B_\varepsilon(p^*) \times \Delta(B_\varepsilon(p)) \times \Delta(B_\varepsilon(p'))$ and define $G : \mathbb{R}_+ \to \mathbb{R}$ by

$$G(\varepsilon) = \min_{(r,\bar{\nu},\nu')\in C_\varepsilon} \sum_{y\in Y} r(y)\left(\frac{\int_{B_\varepsilon(p)} \bar{q}(y)d\bar{\nu}(\bar{q})}{\int_{B_\varepsilon(p')} q(y)d\nu'(q)}\right)^{l}.$$

By the Maximum Theorem, the compactness of $\Delta(B_\varepsilon(p'))$ and $\Delta(B_\varepsilon(p))$ (see, e.g, Theorem 6.4 in Parthasarathy (2005)) and the fact that $G(0) > 1$ by equation (6), there is $\varepsilon' > 0$ such that for all $r, \nu', \bar{\nu} \in C_{\varepsilon'}$

$$\sum_{y\in Y} r(y)\left(\frac{\int_{B_{\varepsilon'}(p)} \bar{q}(y)d\bar{\nu}(\bar{q})}{\int_{B_{\varepsilon'}(p')} q(y)d\nu'(q)}\right)^{l} \geqslant 1. \tag{7}$$

Then,

$$\begin{aligned}
\sum_{y\in Y} r(y)\left(\frac{\nu(B_{\varepsilon'}(p)\mid y)}{\nu(B_{\varepsilon'}(p')\mid y)}\right)^{l} &= \sum_{y\in Y} r(y)\left(\frac{\int_{B_{\varepsilon'}(p)} \nu(B_{\varepsilon'}(p))\bar{q}(y)d\frac{\nu(\bar{q})}{\nu(B_{\varepsilon'}(p))}}{\int_{B_{\varepsilon'}(p')} \nu(B_{\varepsilon'}(p'))\, q(y)d\frac{\nu(q)}{\nu(B_{\varepsilon'}(p'))}}\right)^{l} \\
&= \sum_{y\in Y} r(y)\left(\frac{\int_{B_{\varepsilon'}(p)} \bar{q}(y)d\frac{\nu(\bar{q})}{\nu(B_{\varepsilon'}(p))}}{\int_{B_{\varepsilon'}(p')} q(y)d\frac{\nu(q)}{\nu(B_{\varepsilon'}(p'))}}\right)^{l}\left(\frac{\nu(B_{\varepsilon'}(p))}{\nu(B_{\varepsilon'}(p'))}\right)^{l} \\
&\geqslant \left(\frac{\nu(B_{\varepsilon'}(p))}{\nu(B_{\varepsilon'}(p'))}\right)^{l}
\end{aligned}$$

where the inequality follows from equation (7). ■

Theorem 1′. *Suppose the agent's beliefs are finite dimensional. If* σ *is a limit strategy, then* σ *is a uniform BN-E.*

Proof. If σ is not a uniform BN-E, there is $\bar{p} \in \hat{\Theta}(\sigma)$ such that if $\operatorname{supp}\nu \subseteq \mathcal{E}_\sigma(\bar{p})$, then σ is not a myopic best reply to ν. We fix such a $\bar{p}$ throughout this proof.

Claim 4. *There exists $\varepsilon > 0$ such that if*

$$\frac{\nu\left(\{p \in \Theta\colon \forall s \in S, \forall y \in \operatorname{supp} p^*_{\sigma(s),s}, |p_{\sigma(s),s}(y) - \bar{p}_{\sigma(s),s}(y)| < \varepsilon\}\right)}{1 - \nu\left(\{p \in \Theta\colon \forall s \in S, \forall y \in \operatorname{supp} p^*_{\sigma(s),s}, |p_{\sigma(s),s}(y) - \bar{p}_{\sigma(s),s}(y)| < \varepsilon\}\right)} > \frac{1-\varepsilon}{\varepsilon},$$

then σ is not a myopic best reply to ν.

Proof. Define

$$G(\nu) = \max_{\pi} V(\pi, \nu) - \max_{\tilde{\pi}:\tilde{\pi}(\nu)=\sigma(\cdot)} V(\tilde{\pi}, \nu).$$

From the definition of $\bar{p}$, if

$$\operatorname{supp} \nu \subseteq \{p \in \Theta\colon \forall s \in S, \forall y \in \operatorname{supp} p^*_{\sigma(s),s}, p_{\sigma(s),s}(y) = \bar{p}_{\sigma(s),s}(y)\},$$

then $G(\nu) > 0$. By Lemma 11 the space of policy functions is compact and the value function is continuous in the policy, so $V(\cdot, \nu) - V(\cdot, \nu)$ is a continuous function of the policy, and since $\mathbb{E}_{p,\pi}\left[\sum_{t=1}^{\infty}\left[\beta^{t-1}u(a_t, y_t)\right]\right]$ is continuous in p, $V(\pi, \cdot)$ is continuous in ν. Therefore, we can conclude by the Maximum Theorem that G is continuous.

Now suppose that in contradiction to the claim, for every n there exists a ν_n such that

$$\frac{\nu_n\left(\{p \in \Theta\colon \forall s \in S, \forall y \in \operatorname{supp} p^*_{\sigma(s),s}, |p_{\sigma(s),s}(y) - \bar{p}_{\sigma(s),s}(y)| < 1/n\}\right)}{1 - \nu_n\left(\{p \in \Theta\colon \forall s \in S, \forall y \in \operatorname{supp} p^*_{\sigma(s),s}, |p_{\sigma(s),s}(y) - \bar{p}_{\sigma(s),s}(y)| < 1/n\}\right)} \geqslant \frac{1-1/n}{1/n}$$

and $\sigma \in \pi(\nu_n)$. Because $\Delta(\Theta)$ is sequentially compact, $(\nu_n)_{n\in\mathbb{N}}$ has a converging subsequence $(\nu_{n_i})_{i\in\mathbb{N}} \to \nu^*$. Thus, $\nu^*\left(\{p \in \Theta\colon \forall s \in S, \forall y \in \operatorname{supp} p^*_{\sigma(s),s}, p_{\sigma(s),s}(y) = \bar{p}_{\sigma(s),s}(y)\}\right) = 1$ and $G(\nu^*) = 0$, which would imply that $\sigma \in \pi(\nu^*)$, a contradiction. ■

Now fix such an ε. Because the agent's beliefs are finite dimensional, the agent believes that the outcome distribution depends on the signals only via the partition Ξ. We now define a finer partition of signals Ξ^σ such that for every two signals in the same cell i) the agent thinks they induce the same outcome distribution, i.e., they belong to the same cell of Ξ, and ii) σ prescribes the same action. Formally, Ξ^σ is the collection of subsets of signals of the form

$$\{s \in \xi_i \cap \sigma^{-1}(a) \text{ for some } \xi_i \in \Xi \text{ and } a \in A\}.$$

With a small abuse of notation, for every $\xi \in \Xi^\sigma$ let $\sigma(\xi)$ denote the action that strategy σ prescribes after every signal in ξ, and let $p_{a,\xi}$ be the probability distribution over outcomes induced under p after action a and any signal in ξ. Set $W = \Xi^\sigma \times Y$, and for each $p \in \Theta$, let p^σ be the unique probability measure over W that satisfies

$$p^\sigma(\xi, y) = \zeta(\xi)\, p_{(\sigma(\xi),\xi)}(y) \quad \forall \xi \in \Xi^\sigma, y \in Y.$$

For every $\eta \in (0,1)$, let

$$f_\eta = (1-\eta)p^{*\sigma} + \eta \bar{p}^\sigma.$$

Linearity of H in its first argument implies that for every $\eta \in (0,1)$,

$$p \in \operatorname*{argmin}_{p \in \Theta} H(f_\eta, p^\sigma) \implies p^\sigma = \bar{p}^\sigma.$$

Let g be defined as in the main text with W replacing Y. By the same argument, we still have

$$2g\left((1-\eta)p^{*\sigma} + \eta\bar{p}^\sigma, \varepsilon\right) \geqslant 2\eta\,(\varepsilon)^2.$$

For every $t \in \mathbb{N}$, let $\eta_t = 2t^{-\frac{1}{2}}$. If the empirical frequency is f_{η_t} after t periods, and only strategy σ has been used, then from Lemma 8 and part (ii) of Assumption , there exists $\bar{g} > 0$

$$\begin{aligned}
&\frac{\mu_t\left(\{p \in \Theta \colon \forall s \in S, \forall y \in \operatorname{supp} p^*_{\sigma(s),s}, |p_{\sigma(s),s}(y) - \bar{p}_{\sigma(s),s}(y)| < \varepsilon\}\right)}{1-\mu_t\left(\{p \in \Theta \colon \forall s \in S, \forall y \in \operatorname{supp} p^*_{\sigma(s),s}, |p_{\sigma(s),s}(y) - \bar{p}_{\sigma(s),s}(y)| < \varepsilon\}\right)} \\
= \;& \frac{\bar{\mu}_t\left(\{p \in \Theta \colon \forall w \in \operatorname{supp} p^{*\sigma}, |p^{*\sigma}(w) - \bar{p}^\sigma(w)| < \varepsilon\}\right)}{1-\bar{\mu}_t\left(\{p \in \Theta \colon \forall w \in \operatorname{supp} p^{*\sigma}, |p^{*\sigma}(w) - \bar{p}^\sigma(w)| < \varepsilon\}\right)} \\
\geqslant \;& \mu_0\left(\{p \in \Theta \colon \forall w \in \operatorname{supp} p^{*\sigma}, |p^{*\sigma}(w) - \bar{p}^\sigma(w)| < \varepsilon^2 \frac{2}{\bar{g}t^{\frac{1}{2}}}\}\right) \exp\left(t\eta_t\varepsilon^2\right) \geqslant \Phi\left(\varepsilon^2 \frac{2}{\bar{g}t^{\frac{1}{2}}}\right) \exp\left(t^{\frac{1}{2}}\varepsilon^2\right).
\end{aligned}$$

By Lemma 7 there exists a $\hat{K}, K' > 0$ such that if the empirical frequency is f_t after t periods and $||f_{\eta_t} - f_t|| < ||\bar{p}^\sigma - p^{*\sigma}||t^{-\frac{1}{2}}/K'$ then

$$\frac{\mu_t\left(\{p \in \Theta \colon \forall s \in S, \forall y \in \operatorname{supp} p^*_{\sigma(s),s}, |p_{\sigma(s),s}(y) - \bar{p}_{\sigma(s),s}(y)| < \varepsilon\}\right)}{1-\mu_t\left(\{p \in \Theta \colon \forall s \in S, \forall y \in \operatorname{supp} p^*_{\sigma(s),s}, |p_{\sigma(s),s}(y) - \bar{p}_{\sigma(s),s}(y)| < \varepsilon\}\right)} \geqslant \Psi\left(\hat{K}\varepsilon^2 \frac{2}{\bar{g}t^{\frac{1}{2}}}\right) \exp\left(\hat{K}t^{\frac{1}{2}}\varepsilon^2\right).$$

Fix an outcome $w^0 \in \operatorname{supp} p^{*\sigma}$, and let f_t be the empirical frequency of the other $|\operatorname{supp} p^{*\sigma}| - 1$ outcomes in the support of $p^{*\sigma}$. Denote by $p^{*\sigma}t$ the true probabilities of the same $|\operatorname{supp} p^{*\sigma}| -$

1 outcomes.

An argument that mimics the proof of Claim 2 shows that $f_t \cdot t - p^{*\sigma} t$ is a $|\operatorname{supp} p^{*\sigma}| - 1$ dimensional random walk with nonsingular covariance matrix $\Sigma_{w,w'}$ for the increments.

By the Central Limit Theorem $(f_t - p^{*\sigma})\sqrt{t}$ converges to a Normal random variable with mean 0 and covariance matrix $\Sigma_{w,w'}$. Let $F_t = B_{\frac{||\bar{p}^\sigma - p^{*\sigma}||/K'}{\sqrt{t}}}\left(p^{*\sigma} + \frac{1}{\sqrt{t}}\left(\bar{p}^\sigma - p^{*\sigma}\right)\right)$. We have that

$$\mathbb{P}\left[f_t \in F_t\right] = \mathbb{P}\left[\sqrt{t}(f_t - \bar{p}^*) \in B_{||\bar{p}^\sigma - p^{*\sigma}||/K'}\left(\bar{p}^\sigma - p^{*\sigma}\right)\right].$$

Taking the limit $t \to \infty$ yields that

$$\lim_{t\to\infty} \mathbb{P}\left[f_t \in F_t\right] = \mathbb{P}\left[\tilde{Z} \in B_{||\bar{p}^\sigma - p^{*\sigma}||/K'}\left(\bar{p}^\sigma - p^{*\sigma}\right)\right]$$

where $\tilde{Z}$ is a random variable that is Normally distributed with mean $\vec{0}$ and covariance matrix $\Sigma_{w,w'}$. Consequently, if we denote as E_t the event that $f_t \in F_t$, it follows that $\sum_{t=1}^{\infty} \mathbb{P}\left[E_t\right] = \infty$. Moreover,

$$\liminf_{t\to\infty} \frac{\sum_{s=1}^t \sum_{r=1}^t \mathbb{P}\left[E_s \text{ and } E_t\right]}{\left(\sum_{s=1}^t \mathbb{P}\left[E_s\right]\right)^2} = \liminf_{t\to\infty} \frac{\frac{1}{t^2}\sum_{s=1}^t \sum_{r=1}^t \mathbb{P}\left[E_s \text{ and } E_r\right]}{\left(\frac{1}{t}\sum_{t=1}^{\infty} \mathbb{P}\left[E_t\right]\right)^2} \leqslant \liminf_{t\to\infty} \frac{\frac{1}{t^2}\sum_{s=1}^t \sum_{r=1}^t \mathbb{P}\left[E_r\right]}{\left(\frac{1}{t}\sum_{s=1}^t \mathbb{P}\left[E_s\right]\right)^2}$$
$$= \liminf_{t\to\infty} \frac{\frac{1}{t}\sum_{r=1}^t \mathbb{P}\left[E_r\right]}{\left(\frac{1}{t}\sum_{s=1}^t \mathbb{P}\left[E_s\right]\right)^2} = \frac{1}{\lim_{t\to\infty} \mathbb{P}\left[E_t\right]} = \frac{1}{\mathbb{P}\left[\tilde{Z} \in B_{||\bar{p}^\sigma - p^{*\sigma}||/K'}\left(\bar{p}^\sigma - p^{*\sigma}\right)\right]}.$$

It thus follows from the Kochen-Stone lemma (see Kochen and Stone (1964) or Exercise 2.3.20 in Durrett (2008)) that

$$\mathbb{P}\left[\bigcap_{t=1}^{\infty}\bigcup_{s=t}^{\infty} E_s\right] \geqslant \mathbb{P}\left[\tilde{Z} \in B_{||\bar{p}^\sigma - p^{*\sigma}||/K'}\left(\bar{p}^\sigma - p^{*\sigma}\right)\right] > 0.$$

The event $\bigcap_{t=1}^{\infty}\bigcup_{s=t}^{\infty} E_s$ is invariant under finite permutations of the increments $\left(\mathbf{1}_{w_t = w^1}, \ldots, \mathbf{1}_{w_t = w^{|\operatorname{supp} p^{*\sigma}|-1}} - p^{*\sigma}\right)$ with different time indices, so the Hewitt-Savage zero-one law (see, e.g., Theorem 8.4.6 in Dudley (2018)) implies that the probability of the event $\bigcap_{t=1}^{\infty}\bigcup_{s=t}^{\infty} E_s$ must equal zero or one. As the probability is strictly positive it must equal one.

This implies that $f_t \in F_t$ infinitely often with probability 1. It follows that the agent will eventually want to take an action different from σ:

$$\mathbb{P}\left[a_t \neq \sigma\left(s_t\right) \text{ for some } t\right] = 1.$$

Thus the strategy can not converge to σ with positive probability. ■

Theorem 2′. *Suppose σ is a uniformly strict BN-E. Then there is a belief $\nu \in \Delta(\Theta)$ such that for every $\kappa \in (0,1)$ there exists an $\varepsilon' > 0$ such that starting from any prior belief in $B_{\varepsilon'}(\nu)$:*

$$\mathbb{P}_\pi \left[\lim_{t\to\infty} \frac{1}{t+1} \sum_{r=0}^{t} \mathbf{1}_{\pi(a^r, y^r, s^{r+1}) = \sigma(s_{r+1})} \geqslant 1-\kappa \right] > 1-\kappa\,.$$

Proof. Consider a uniformly strict BN-E σ, an optimal policy π and $\kappa \in (0,1)$. By Lemma 12, for every $\lambda \in (0,1)$ there exists an ε such that if $\nu(\hat{\Theta}^\varepsilon(\sigma)) \geqslant 1-\varepsilon$, then

$$\left|\zeta\left(\{s \in S : \pi(\nu, s) = a\}\right) - \zeta\left(\{s \in S : \sigma(s) = a\}\right)\right| < \lambda.$$

For every $l \in (0,1)$, define the function $f_{l,\sigma} : P \times P \to \bar{\mathbb{R}}$ is defined by

$$f_l(\sigma', \bar{p}, p') = \sum_{y\in Y} p^*_{\sigma'}(y) \left(\frac{\bar{p}_{\sigma'}(y)}{p'_{\sigma'}(y)}\right)^l.$$

By Lemma 13, since $\hat{\Theta}^\varepsilon(\sigma)$ is compact by Lemma 10, and since f_l is lower semicontinuous, there exists $\varepsilon' \in (0,\varepsilon)$ such that $\bar{p} \in \hat{\Theta}^{\varepsilon'}(\sigma)$ implies that $f_l(\sigma, \bar{p}, p') > 1$ for all p' with $p' \notin \hat{\Theta}^\varepsilon(\sigma)$. Let $K = \left(\frac{\varepsilon}{1-\varepsilon}\right)^l$. Then

$$\left(\frac{1-\nu\left(\hat{\Theta}^\varepsilon(\sigma)\right)}{\nu\left(\hat{\Theta}^{\varepsilon'}(a)\right)}\right)^l < K \implies \frac{1-\nu\left(\hat{\Theta}^\varepsilon(\sigma)\right)}{\nu\left(\hat{\Theta}^\varepsilon(\sigma)\right)} < \frac{\varepsilon}{1-\varepsilon}$$
$$\implies \nu\left(\hat{\Theta}^\varepsilon(\sigma)\right) > 1-\varepsilon \implies \pi(\nu) = a.$$

By Lemma 10, $\hat{\Theta}^\varepsilon(\sigma)$ is compact, so it has a finite cover $\{p \in \Theta : ||q^i_a - p_a|| \leqslant \varepsilon\}_{i=1}^n$, where $q^i \in \hat{\Theta}^\varepsilon(\sigma)$.

Let $\bar{\varepsilon}$ be such that $\nu\left(\hat{\Theta}^{\bar{\varepsilon}}(\sigma)\right) > 1-\bar{\varepsilon}$ implies that

$$\left(\frac{1-\nu\left(\hat{\Theta}^\varepsilon(\sigma)\right)}{\nu\left(\hat{\Theta}^\varepsilon(\sigma)\right)}\right)^l < \frac{K(1-\kappa)}{n}.$$

Then if the agent starts with a belief ν_0 with $\nu_0(\hat{\Theta}(\sigma)) > \bar{\varepsilon}$, σ is the unique best reply ν_0'. Moreover, by Lemma 14, Dubins' upcrossing inequality, and the union bound, there is a probability $(1-\kappa)$ that the positive supermartingale

$$\left(\frac{1-\nu_t'\left(\hat{\Theta}^{\varepsilon}(\sigma)\right)}{\nu_t'\left(\hat{\Theta}^{\varepsilon}(\sigma)\right)}\right)^l$$

never rises above K, and with probabilty $(1-\kappa)$

$$|\zeta\left(\{s \in S : \pi\left(\mu_t', s\right) = a\}\right) - \zeta\left(\{s \in S : \sigma\left(s\right) = a\}\right)| \leqslant \lambda,$$

for all $t \in \mathbb{N}$. Then the statement follows from the Hewitt-Savage $0-1$ Law (see, e.g., Theorem 8.4.6 in Dudley (2018)). ■

Theorem 4′. *If signals are finite and subjectively uninformative and outcomes are subjectively exogenous, then any uniformly strict BN-E σ is positively attractive.*

Proof. Under the assumptions of the theorem, $\Theta \subseteq \Delta\left(\Delta\left(Y\right)\right)$. Consider a uniformly strict BN-E σ. By an obvious extension of Lemma 1 to the case with signals, $\Delta\left(\hat{\Theta}(\sigma)\right)$ is compact. Similarly, since S is compact and σ is the unique optimal best reply strategy at the beliefs in $\Delta\left(\hat{\Theta}(\sigma)\right)$, Lemma 3 can be extended to guarantee that there exists $\varepsilon \geqslant 0$ such that if

$$\nu\left(\mathrm{cl}\left(Q_{\varepsilon}\left(\bar{p}_{\sigma}\right)\right)\right) \geqslant (1-\varepsilon)$$

then the myopic best reply to ν is σ. By the same argument of the proof of Theorem 2, there exists an $l \in (0,1)$ and $\varepsilon' \in (0,\hat{\varepsilon})$, such that if $p \in Q_{\varepsilon'}\left(\bar{p}_{\sigma}\right)$ and $p' \notin Q_{\hat{\varepsilon}}\left(\bar{p}_{\sigma}\right)$ then $f_l(p,p') \geqslant 1$.

Using the Maximum Theorem again we can find a sequence of outcome realizations y^t such that if $\hat{p}_t$ is the corresponding empirical frequency, it is sufficiently close to $\bar{p}_{\sigma}$ to have

$$Q_{\hat{\varepsilon}/2}\left(\hat{p}_t\right) \subseteq Q_{\hat{\varepsilon}}\left(\bar{p}_{\sigma}\right).$$

Therefore by Proposition 1, there exists a time period T such that for all $t' > T$, if the empirical frequency $\hat{p}_{t'} = \hat{p}_t$, the agent assigns a relative probability higher than K to an $\hat{\varepsilon}$ Ball around $\bar{p}$. That is,

$$\frac{\mu_{t'}(Q_{\hat{\varepsilon}}\left(\bar{p}_{\sigma}\right))}{1-\mu_{t'}(Q_{\varepsilon'}\left(\bar{p}_{\sigma}\right))} \geqslant \frac{\mu_{t'}(Q_{\hat{\varepsilon}/2}\left(\bar{p}_{\sigma}\right))}{1-\mu_{t'}(Q_{\varepsilon'}\left(\bar{p}_{\sigma}\right))} > 2\frac{(1-\hat{\varepsilon})}{\hat{\varepsilon}}.$$

Notice that replicating the outcome realizations y^t sufficiently many time yields a sequence

$y^{t'}$ such that the empirical frequency $\hat{p}_{t'} = \hat{p}_t$ and $t' > T$. Since $\operatorname{supp} p^*_{a,s} = Y$ for all $(a, s) \in A \times S$, this sequence of outcomes has positive probability, and after it occurs the agent plays σ. By Lemma 4 and the law of iterated expectations, conditional on a being played $\left(\frac{1-\mu_{t'}(Q_{\varepsilon'}(\bar{p}_\sigma))}{\mu_{t'}(Q_{\hat{\varepsilon}}(\bar{p}_\sigma))}\right)^l$ is a positive supermartingale.

Then, by Dubins' upcrossing inequality, there is positive probability that this positive supermartingale never rises above $\frac{\hat{\varepsilon}}{(1-\hat{\varepsilon})}$, that in turns imply that $\mu_{t'}(Q_{\varepsilon'/2}(\hat{p}_t))$ never goes below $(1-\hat{\varepsilon})$ and therefore σ is always played after the sequence y^t. ■

Corollary 4. *Let α be a strongly uniform mixed BN-E in a problem (A, Y, p^*, u, Θ). There is a sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$ such that each $\sigma_{1/n}$ is a uniformly stable BN-E of a $(1/n)$ perturbation of (A, Y, p^*, u, Θ) and*

$$\lim_{n\to\infty} \zeta(\{s : \sigma_n(s) = a\}) = \alpha(a) \quad \forall a \in A.$$

If (A, Y, p^*, u, Θ) is subjectively exogenous and p^* has full support, the $\sigma_{1/n}$ can be chosen to be also positively attractive.

Proof. Let α be a mixed BN-E in a problem (A, Y, p^*, u, Θ). For every $n \in \mathbb{N}$, let $S = \operatorname{supp}\alpha$, $\zeta(a) = \alpha(a)$, and $\tilde{u}(a, y, s) = u(a, y) + \frac{1}{n}\mathbf{1}_{a=s}$, and let $\tilde{p}^*, \tilde{\Theta}$ be as given in part (ii) and (iii) of the definition of a perturbed environment.

Consider the strategy $\sigma(a) = a$. We have that for every $p \in \Theta$

$$\sum_{s\in S} \zeta(s) H\left(\hat{p}^*_{\sigma(s),s}, \Phi(p)_{\sigma(s),s}\right) = \sum_{a\in A} \alpha\left(a\right) p^*_a\left(y\right) \log p_a\left(y\right)$$

by (ii) and (iii) of the definition of a perturbed problem. Therefore, $\hat{\Theta}(\sigma) = \Phi(\hat{\Theta}\left(\alpha\right))$. Fix a signal $s \in S$, and consider any action $a' \neq \sigma(s)$. Since α is a strongly uniform BN-E

$$\mathbb{E}_{p_{\sigma(s)}}\left[u(\sigma(s), y)\right] \geqslant \mathbb{E}_{p_{a'}}\left[u(a', y)\right] \qquad \forall p \in \Theta(\alpha)$$

and by definition of $\tilde{u}$

$$\mathbb{E}_{p_{\sigma(s)}}\left[\tilde{u}(\sigma(s), y, s)\right] \geqslant \mathbb{E}_{p_{a'}}\left[\tilde{u}(a', y, s)\right] + 1/n \qquad \forall p \in \Theta(\alpha)$$

proving that σ is a strictly uniform BN-E. By construction

$$\zeta(\{s : \sigma_n(s) = a\}) = \alpha(a) \quad \forall a \in A.$$

Then the result follows by Theorems $2'$ and $4'$. ∎

B.4 Additional Examples

Example 9 (A uniformly strict BN-E that isn't positively attractive)**.** *In this example the prior has support* $\{p^1, p^2, p^3\}$*. Here* $a = 3$ *is the only BN-E and is uniformly strict. However, if the agent takes an action* $a \in \{1, 2\}$ *then the subjective likelihood assigned to* p^3 *goes down and thus play never converges to* $a = 3$ *if the prior assigns sufficiently low probability to* p^3*. The details are in the following table:*

a	$a=1$			$a=2$			$a=3$			$H(p_a^*,\cdot)$			$A^m(\delta_{(\cdot)})$
y	1	2	3	1	2	3	1	2	3				
u	1	0	0	0	1	0	0	0	1	$a=1$	$a=2$	$a=3$	
p^*	0.1	0.9	0	0.9	0.1	0	0.1	0.1	0.8				
p^1	0.5	0.3	0.2	0.5	0.3	0.2	0.5	0.3	0.2	1.15	$\underline{0.74}$	2.03	$a=1$
p^2	0.3	0.5	0.2	0.3	0.5	0.2	0.3	0.5	0.2	$\underline{0.74}$	1.15	2.03	$a=2$
p^3	0.1	0.1	0.8	0.1	0.1	0.8	0.1	0.1	0.8	2.3	2.3	$\underline{0.64}$	$a=3$

Example 10 (Signal Neglect)**.** *A seller in a physical marketplace can hire one shop assistant to work for the day* a_H *or not hire anyone* a_N*. The outcome* $y \in Y$ *is the percentage of consumers in the marketplace that buy the good, with two possibilities,* $y_h > y_l$*.*

Before choosing whether to hire, the agent observes the number of people at the market that day $s \in \{s_h, s_l\}$*, with* $s_h > s_l$*. The payoff function is* $u(a, y, s) = sy - 1_{a=a_H}$*. The seller realizes that the signal is payoff relevant, but falsely believes that it does not provide any information about the outcome. The agent is uncertain about how useful it is to hire a shop assistant, and in particular they do not know whether hiring is ineffective, i.e., for all* $a \in A, y \in Y$*,* $p_a(y) = 1/2$*, or if it is not, i.e.,* $p'_{a_H}(y_H) = 3/4$ *and* $p'_{a_N}(y_H) = 1/4$*.*

The fraction of consumers who buy varies with the signal: On days with fewer consumers, the ones that actually come to the market are more likely to purchase the good. Formally:

$$p^*_{s_H,a_H}(y_H) = 1/2, \quad p^*_{s_H,a_N}(y_H) = 1/4, \quad p^*_{s_L,a_H}(y_H) = 3/4, \quad p^*_{s_L,a_N}(y_H) = 1/2.$$

Let $\frac{s_l(y_h - y_l)}{4} < 1 < \frac{s_h(y_h - y_l)}{4}$*, so that it is not objectively optimal to hire a shop assistant after* s_L*, and it is objectively optimal to hire an assistant after* s_H*. The following argument* *shows that the only BN-E is that the shop assistant is never hired: If the agent followed the objectively optimal strategy, they would observe the same frequency of sales in days with $s = s_H$ and with the shop assistant hired as in days with $s = s_L$ and without the shop assistant: $p^*_{s_H,a_H}(y_H) = 1/2 = p^*_{s_L,a_N}(y_H)$. This holds because the shop assistant offsets the lower per-customer demand on days with high attendance. However, this observation supports the belief that the shop assistant is useless. Since the myopic best reply to δ_p is to never hire the shop assistant, by Theorem 1' this suboptimal action is the only possible limit action.*

Example 11 (A Uniform BN-E that is not Stable)**.** *There are two actions, a and b, two outcomes, 0 and 1, and two action-dependent outcome distributions, $\Theta = \{p, p'\}$. The utility of the agent is equal to the outcome, i.e., $u(a, y) = y$, and $p_a(1) = p'_a(1) = p^*_a(1) = \frac{1}{2}$, $p_b(1) = \frac{1}{2} < p'_b(1) = p^*_b(1) = \frac{3}{4}$. Here, a is a myopic best reply to the belief δ_p, so it is a BN-E. Moreover, there is a unique class of observationally equivalent outcome distributions under a: $\mathcal{E}_a(p) = \Theta$, so a is a uniform BN-E. However, it is not stable: under every optimal policy of the agent and starting from every belief that assigns positive probability to p', the agent will play action b forever with probability 1.*