Learning in Society*

Braz Camargo†

July 3, 2006

Abstract

In the canonical learning model, the multi-armed bandit with independent arms, a decision maker learns about the different alternatives by his experience only. It is well known that an optimal experimentation strategy for this problem sometimes leads the best alternative to be dropped altogether, the so-called Rothschild effect. Many situations of interest, however, involve learning from individual experience *and* the experience of others. This paper shows that learning in society can overcome the Rothschild effect. We consider an economy with a continuum of infinitely lived agents where each one of them faces a two-armed bandit and the unknown stochastic payoffs of each arm are the same for all agents. In each period, agents are randomly and anonymously matched in pairs, where they observe their partners' current action choice. We establish that if initial beliefs are sufficiently heterogeneous, then the fraction of agents who choose the superior arm converges to one in any perfect bayesian equilibrium of this game.

JEL Classification: C73, D82, D83.

Key words: Multi-Armed Bandit, Social Learning, Strategic Experimentation.

*I am very grateful to George Mailath for his encouragement and advice. I would also like to thank Jan Eeckhout, Andrew Postlewaite, and Rafael Rob for their comments.

†Address: University of Western Ontario, Department of Economics, Social Science Centre, London, Ontario N6A 5C2, Canada. E-mail: bcamargo@uwo.ca.

1 Introduction

The defining characteristic of an experimentation problem is the tradeoff between learning about the different available alternatives, which provides valuable information for future decisions, and maximizing immediate rewards. In the canonical learning model, the multi-armed bandit with independent arms, this tradeoff implies that an optimal experimentation strategy can lead the superior alternative to be dropped altogether, see Banks and Sundaram (1992). This is sometimes referred to as the Rothschild effect. However, in many situations of interest an agent can learn from the experience of other agents that face the same problem. Examples include consumers learning about product quality and doctors learning about the efficacy of different treatments for the same disease. In this paper we show how the presence of information flows across individuals can overturn the Rothschild effect.

We consider a discrete time economy populated with a continuum of infinitely lived agents where each one of them faces a two-armed bandit with independent arms. The unknown stochastic payoffs to each of the available action choices are the same for all agents, i.e., agents are homogeneous. Information is transmitted across agents in the following way. In every period, agents are randomly and anonymously matched in pairs, where they observe their partner's current action choice. We refer to these matches as meetings in society.

In this environment, an agent's flow payoff depends only on his action choice. Differently from the standard bandit problem though, the action choices of the other individuals in the population reveal information about the available alternatives. At any point in time, the likelihood an agent has of being matched to someone else that chooses a specific action is determined by the behavior of the other agents. If these matching probabilities depend on the true payoffs of the action choices, the meetings in society reveal payoff relevant information even if outcomes are not observable within a match. Consequently, when choosing an action, an agent has to take into consideration: (i) the tradeoff between the flow payoffs from the different alternatives and the information they provide; (ii) the fact that, independently of what he chooses, his meetings in society provide information about the available alternatives. The hypothesis of random and anonymous matchings implies that each agent only cares about the aggregate behavior of the other agents in the population.

It turns out that the individual learning problem, i.e., the experimentation problem each agent has to solve when he takes the behavior of all the other agents as given, is formally equivalent to a multi-armed bandit with correlated arms. Easley and Kiefer (1988) study limiting behavior in a large class of infinite horizon individual experimentation problems, including multi-armed bandits with correlated arms. They only consider stationary problems, however, while here we must consider non-stationary ones. Indeed, aggregate behavior changes over time if learning takes place. Hence, matching probabilities also change over time, and with them the informational content of the meetings in society.

Most of the literature on social learning only considers purely informational interaction among individuals. For example, Smallwood and Conlisk (1979), Ellison and Fudenberg (1993, 1995), and Bala and Goyal (1999) consider models of social learning with boundedly rational agents. Banarjee (1992), Bikhchandani et. al. (1992), and Smith and Sorensen (2000) consider sequential decision models with rational agents. Of the few papers that deal with strategic and informational interaction among individuals, Aoyagi (1998) and Bolton and Harris (1999) are the closest in spirit to this paper.[1] Both of them consider games of strategic experimentation with a finite number of players. In Bolton and Harris there is no asymmetric information, since the outcome of each player's action choice is public. In Aoyagi, an individual's action choice is public, but not its outcome. He shows, under certain restrictions, that in any Nash equilibrium of the corresponding game all players eventually settle on the same action choice, not necessarily the superior one.

This article is structured as follows. The model is introduced in the next section. Section 3 contains some preliminary discussion. Section 4 considers the individual learning problem. There we establish a characterization result that plays an important role in the analysis of long-run aggregate behavior. The main result of the paper, that the fraction of the population choosing the superior action converges to one in any perfect bayesian equilibrium of this game when initial beliefs are sufficiently heterogenous, is established in Section 5. Section 6 concludes and several appendices contain omitted proofs.

Some conventions, definitions, and facts. Unless otherwise stated, measurability is always understood to be Borel measurability. Finite sets are endowed with the discrete topology and products of topological spaces are endowed with the product topology. For any set B, I_B denotes its indicator function and B^{-t} denotes the set $\times_{k=t+1}^{\infty} B$.

[1] See also Keller et. al. (2005).

If S is a metric space, $\mathcal{B}(S)$ denotes its Borel σ-algebra and $\mathcal{P}(S)$ denotes the set of all Borel probability measures on S. Let $(\Omega, \mathcal{F})$ be a measure space, S be a complete separable metric space (a Polish space), and endow $\mathcal{P}(S)$ with the topology of weak convergence of probability measures. A $\mathcal{F}$-measurable map $\lambda : \Omega \to \mathcal{P}(S)$ is transition probability from $(\Omega, \mathcal{F})$ into $(S, \mathcal{B}(S))$.[2] In what follows, we always omit $\mathcal{B}(S)$ and say λ is a transition probability from $(\Omega, \mathcal{F})$ into S. Moreover, if Ω is a metric space and $\lambda : \Omega \to \mathcal{P}(S)$ is measurable, we also omit $\mathcal{F}(= \mathcal{B}(\Omega))$. In particular, if $\lambda : \Omega \to \mathcal{P}(S)$ is continuous, then λ is a (continuous) transition probability from Ω into S.

2 The Model

Time is discrete and indexed by $t \in \mathbb{Z}_+$. The economy is populated with a continuum of mass one of infinitely lived agents that we identify with $([0,1], \Sigma, \mu)$, where Σ is the σ-algebra of Lebesgue measurable subsets of $[0,1]$ and μ is the Lebesgue measure on $[0,1]$. All agents have the same discount factor $\beta \in [0,1)$. We denote a typical element of $[0,1]$ by i.

Every period has two parts. First, each agent privately chooses one of N actions, labelled 1 to N, observes a stochastic outcome $y \in Y$, and collects a reward $r(j, y)$, where $j \in A = \{1, \ldots, N\}$ denotes his action choice. We also use k to denote an action choice. The outcome space Y is a Polish space and the value of y is determined independently for each agent in the population. Then, all agents are randomly and anonymously matched in pairs, where they observe the current action choice of their partners. We assume that each agent randomly chooses an action in period zero.[3]

The outcome of each action j depends on a parameter θ_j that is the same for all agents. The set Θ_j of possible values of θ_j is finite. We refer to the value of θ_j as the true type of j and to the set $\Theta = \Theta_1 \times \cdots \times \Theta_N$, with typical element $\theta = (\theta_j)$, as the set of states of the world. The value of θ is initially unknown to the agents. To each pair (j, θ_j) is associated a Borel probability measure $\mu_j(\theta_j)$ on Y that governs the realization of outcomes when j is chosen and its true type is θ_j. The maps $\theta_j \mapsto \mu_j(\theta_j)$ are assumed to be one-to-one.

[2]This definition coincides with the usual one; i.e., $\lambda : \Omega \to \mathcal{P}(S)$ is $\mathcal{F}$-measurable if, and only if, for each $D \in \mathcal{B}(S)$, the map $\omega \mapsto \lambda(\omega)(D)$ is $\mathcal{F}$-measurable.

[3]This assumption is irrelevant for our results. With it, the individual learning problem fits the description of a dynamic programming problem with unknown transition probabilities, see Rieder (1975). It also allows us to introduce the requirement of sequential rationality in the agents' behavior in a compact way.

Let $\Pi = \mathcal{P}(\Theta)$ denote the set of beliefs about the state of the world, beliefs for short. Then, $\Pi = \Delta_S$, where $S+1 = \Pi_{j=1}^{N}|\Theta_j|$. The case of interest is when $\sum_{j=1}^{N}|\Theta_j| \geq N+1$. Denote a typical element of Π by π and the probability that π assigns to θ, the belief that the state of the world is θ, by $\pi(\theta)$. Now let $\Pi^d = \{\pi \in \Pi : \pi = \pi_1 \times \cdots \times \pi_N,\ \text{with}\ \pi_j \in \mathcal{P}(\Theta_j)\}$. Each agent begins period zero with a non-dogmatic prior in Π^d. Hence, in the absence of the meetings in society, the problem of the agents is a multi-armed bandit with independent arms.

Prior beliefs may be heterogeneous. There is a measurable function $\Phi : I \to \Pi$ such that $\Phi(i)$ is the prior belief of the agent i. Notice that the range of Φ must be in the (relative) interior of Π^d. We also assume that the range of Φ is countable.

Let ω be the element of $\mathcal{P}(Y)$ given by $\sum_{j=1}^{N}\sum_{\theta_j\in\Theta_j}\mu_j(\theta_j)$. By definition, the measures $\mu_j(\theta_j)$ are absolutely continuous with respect to ω. Denote the density of $\mu_j(\theta_j)$ with respect to ω by $g_j(\cdot,\theta_j)$. Now let $r_j(\theta_j) = \int r(j,y)g_j(y,\theta_j)\omega(dy)$ be the expected flow payoff from j when its true type is θ_j. We make the following assumptions. The first three are regularity assumptions. The fourth rules out the case where there is at least one state of the world where the expected rewards from two or more of the available action choices are the same. The fifth implies that the maps $\theta_j \mapsto r_j(\theta_j)$ are one-to-one. The last assumption implies that for each j there is at least one state of the world where this action is the best alternative.

ASSUMPTION 1. *The reward function* $r : A \times Y \to \mathbb{R}$ *is bounded.*

ASSUMPTION 2. *For each* j *and* $y \in Y$ *there exists* θ *such that* $g_j(y,\theta_j) > 0$.

ASSUMPTION 3. $\int y^m g_j(y,\theta_j)\omega(dy) < \infty$ *when* $m = 1, 2$ *for all* j *and* θ.

ASSUMPTION 4. $r_j(\theta_j) \neq r_k(\theta_k)$ *for all* θ *and* $j \neq k$.

ASSUMPTION 5. $\theta_j \neq \theta_j'$ *implies that* $r_j(\theta_j) \neq r_j(\theta_j')$.

ASSUMPTION 6. *For each* j *there exists* θ *such that* $r_j(\theta_j) > r_k(\theta_k)$ *for all* $k \neq j$.

By Assumption 4, there exists $\underline{J} \subset A$ with $N-1$ elements such that if $j \in \underline{J}$, then there is θ_j with the property that $r_j(\theta_j) < \min_{\theta_k} r_k(\theta_k)$ for $k \in A \setminus \underline{J}$. Let $\underline{\Theta}_j$ be the set of all such θ_j's. Assumption 6 implies that $\underline{\Theta}_j$ is a proper subset of Θ_j for all $j \in \underline{J}$. In what follows, we assume that $\underline{J} = \{1, \ldots, N-1\}$.

3 Preliminaries

An agent's experience in a given period is a triple (j, y, k), where j is his action choice, y is the outcome of j, and k is the (current) action choice of his partner. Let $X = Y \times A$ be the observation space, i.e., the set of possible one-period observations an agent can make, and let $Z = A \times X$ be the set of possible action-observation pairs. Then, $H_t = Z^t$ is the set of period t histories and $H_\infty = Z^\infty$ is the set of infinite histories. Denote a typical element of X by $x = (x_1, x_2)$, a typical element of Z^t by z_t, and a typical element of H_t by $h_t = (j_0, x_0, \ldots, j_{t-1}, x_{t-1})$. A behavior strategy is a sequence $f = \{f_t\}_{t \in \mathbb{N}}$, where $f_t : H_t \to \Delta_{N-1}$ is the measurable function describing how period $t \geq 1$ histories are mapped into probability measures over A. By convention, the jth coordinate of an element of Δ_{N-1} denotes the probability that j is chosen.

We now define what a strategy profile is in this environment. For this, some terminology is needed. Let S be a metric space and X be a Banach space with norm $||\cdot||$. The set of all bounded and measurable functions from S into X is denoted by $\mathrm{B}_b(S, X)$ (or $\mathrm{B}_b(S)$ when $X = \mathbb{R}^1$). We always take this set to be endowed with the sup-norm, in which case it is a Banach space. A function $f : [0, 1] \to X$ is simple if there exist $x_1, \ldots, x_m \in X$ and $E_1, \ldots, E_m \in \Sigma$ such that $f = \sum_{k=1}^m x_k I_{E_k}$. A function $f : [0, 1] \to X$ is strongly Lebesgue measurable if there exists a sequence $\{f_n\}$ of simple functions such that $\lim_n ||f_n(i) - f(i)|| = 0$ for almost all $i \in [0, 1]$. If f is simple and $E \in \Sigma$, the integral $\int_E f\mu$ is defined as $\sum_k x_k \mu(E_k \cap E)$. A strongly Lebesgue measurable function $f : [0, 1] \to X$ is Bochner integrable if there is a sequence $\{f_n\}$ of simple functions such that $\int ||f_n(i) - f(i)||\mu(di) = 0$, in which case $\int_E f\mu$ is defined as the limit of $\int_E f_n\mu$. The set of all equivalence classes of Bochner integrable functions is denoted by $\mathrm{L}_1(\mu, X)$ (or $\mathrm{L}_1(\mu)$, in case $X = \mathbb{R}$).

Let $\Gamma_t = \mathrm{B}_b(H_t, \mathbb{R}^N)$ and $\Omega_t = \{f \in \Gamma_t : f(h_t) \in \Delta_{N-1}$ for all $h_t \in H_t\}$. Then, $\Omega = \times_{t \in \mathbb{N}} \Omega_t$ is the set of all possible behavior strategies. Now let $\Psi_t = \mathrm{B}_b(\Pi, \Omega_t)$ and $\Psi = \times_{t \in \mathbb{N}} \Omega_t$. An element of Ψ is what we call a prior-contingent behavior strategy. Denote an arbitrary element of $\mathrm{L}_1(\mu, \mathrm{B}_b(\Pi, \Gamma_t))$ by F_t and define, in an abuse of notation, $\mathrm{L}_1(\mu, \Psi_t)$ to be set of F_t in $\mathrm{L}_1(\mu, \mathrm{B}_b(\Pi, \Gamma_t))$ such that $\mu(\{i : F_t(i) \in \Psi_t\}) = 1$. We take $\Lambda = \times_{t \in \mathbb{N}} \mathrm{L}_1(\mu, \Psi_t)$ as the set of possible strategy profiles. This definition captures the notion that each agent in the population is non-atomic, and so his particular choice of prior-contingent behavior strategy has no aggregate effect.

Suppose $F = \{F_t\} \in \Lambda$ is the strategy profile under play and denote the canonical basis of $\mathbb{R}^N$ by $\{e_1, \ldots, e_N\}$. Given $h_1 \in H_1$, the probability that agent i chooses $j \in A$ is $\langle F_1(i)(\Phi(i))(h_1), e_j \rangle$. By assumption, all agents randomize in period zero. Since there is no aggregate uncertainty, this implies that the fraction of the population that chooses a given action in period zero is the same in every state of the world, $\frac{1}{N}$. Hence, when the state of the world is θ, the probability that i chooses j in period one is

$$p_1(i, j, \theta, F) = \sum_{j,k=1}^{N} \frac{1}{N^2} \int \langle G_1(i)(j, y, k), e_j \rangle \mu_j(dy|\theta_j),$$

where $G_1(i) = F_1(i)(\Phi(i))$. The above integral is well-defined by Corollary 2 in Appendix A together with Theorem 13.4 in Aliprantis and Border (1999). By Lemma 9 in Appendix A, the function $p_1 : [0,1] \times A \times \Theta \times \Lambda \to [0,1]$ just defined is Lebesgue measurable in i. Consequently, once more because there is no aggregate uncertainty, the measure of agents who choose j in period one when the state of the world is θ is

$$m_1(j, \theta, F) = \int p_1(i, j, \theta, F) \mu(di).$$

This number is also the probability, when the state of the world is θ and F is under play, that in period one an agent is matched to a partner whose current action is j.

From the matching probabilities $m_1(j, \theta, F)$ we can construct the map $\tau_2 : [0,1] \times \Theta \times \Lambda \to \mathcal{P}(H_2)$ such that if $D \in \mathcal{B}(H_2)$, then $\tau_2(i, \theta, F)(D)$ is the probability that agent i experiences $h_2 \in D$ when the state of the world is θ and F is under play. It is possible to show (see Appendix B) that for each θ and F, the map $\tau_2(\cdot, \theta, F)$ is a transition probability from $([0,1], \Sigma)$ into H_2. The probability, as a function of θ and F, that i chooses j in period two is then

$$p_2(i, j, \theta, F) = \int \langle G_2(i)(h_2), e_j \rangle \tau_2(dh_2|i, \theta),$$

where $G_2(i) = F_2(i)(\Phi(i))$. Given the functions $p_2 : [0,1] \times A \times \Theta \times \Lambda \to [0,1]$ we can then construct, in the same way as in the previous paragraph, the period two matching probabilities $m_2(j, \theta, F)$.

Continuing with this process, we obtain a sequence $m = \{m_t\}$, where $m_t = \{m_t(j, \theta, F)\}_{j \in A, \theta \in \Theta}$ is the vector of period t matching probabilities; i.e., $m_t(j, \theta, F)$ is the fraction of agents who choose j in period t when the state of the world is θ and the strategy profile is F. The details can be found in Appendix B. Let $\Xi = \times_{t \in \mathbb{Z}_+} \Delta_{N-1}^{S+1}$ and denote by M the map that takes an element F of Λ into its corresponding sequence $m(F) \in \Xi$ of matching vectors.

Because the matching process is random and anonymous, the infinite sequence m of matching vectors subsumes all the informational content of the meetings in society. In other words, besides outcomes and rewards, the individual learning problem is characterized by the sequence m of matching vectors.

Denote by $\sigma_\theta(h_1, f, m)$ the Borel probability measure on Z^∞ induced by the period zero action-observation pair h_1, the behavior strategy f, and the infinite sequence of matching vectors m when the state of the world is θ.[4] Now let $z_t = (j_t, x_t)$ be the period t action-observation pair and define $R : Z^\infty \to \mathbb{R}$ to be such that if $z_\infty = \{z_t\}$, then $R(z_\infty) = \sum_{t=1}^{\infty} \beta^{t-1} r(j_t, x_{1t})$, where x_{1t} is the first coordinate of x_t.[5] The objective of an agent with prior π_0 is to choose $f^* \in \Omega$ such that

$$\sum_{\theta \in \Theta} \int R(z_\infty)\sigma_\theta(dz_\infty|h_1, f^*, m)\pi_0(\theta) = \sup_{f\in\Omega} \sum_{\theta\in\Theta} \int R(z_\infty)\sigma_\theta(dz_\infty|h_1, f, m)\pi_0(\theta) \qquad \forall h_1 \in H_1. \quad (1)$$

Observe that if $\widetilde{r} : A \times X \to \mathbb{R}$ is such that $\widetilde{r}(a_t, x_t) = r(a_t, x_{1t})$, then $R(z_\infty) = \sum_{t=1}^{\infty} \beta^{t-1}\widetilde{r}(a_t, x_t)$. Hence, the individual learning problem is equivalent to a non-stationary multi-armed bandit with correlated arms where the reward function is $\widetilde{r}$ and the outcome space is X.

Definition: *A perfect bayesian equilibrium is a profile $F^* \in \Lambda$ such that if $m = M(F^*)$, then, for almost all $i \in [0, 1]$, $F^*(i)(\pi_0)$ satisfies (1) for all $\pi_0 \in \Pi^d$.*

4 The Individual Learning Problem

We first describe the individual learning problem and establish a few basic results. We then discuss a result concerning belief updating in the presence of meetings in society. We finish with a characterization result, Theorem 1, that plays a central role in the next section.

4.1 Description and Basic Results

Define $\eta_t^\theta : \Xi \to \mathcal{P}(A)$ to be such that if $m = \{m_t\} \in \Xi$, then $\eta_t^\theta(m) = \sum_{j=1}^{N} m_t(j, \theta)\delta_j$, where δ_j is the Dirac measure on A with mass on j. By construction, $\eta_t^\theta(m)$ is the probability measure describing the outcome of the matching process in period t when the state of the world is θ and the sequence of matching vectors is m. Now let $\nu_t^\theta : A \times \Xi \to \mathcal{P}(X)$ be such that $\nu_t^\theta(j, m) =$

[4]Notice that $\sigma_\theta(h_1, f, m)$ also depends on the measures $\mu_j(\theta_j)$.

[5]We ignore the period zero payoffs.

$\mu_j(\theta_j) \times \eta_t^\theta(m)$. By definition, $\nu_t^\theta(j,m)$ is the probability measure describing the distribution of possible period t observations as a function of θ, j, and m.

For each prior π_0 and sequence of matching vectors m, the individual learning problem is a non-stationary markovian decision problem (DP) with state space X and unknown transition probabilities $\{\nu_t^\theta\}_{t\in\mathbb{Z}_+,\theta\in\Theta}$, see Rieder (1975).[6] In this section we make use of the fact that this problem is equivalent to a certain non-stationary non-markovian DP with state space X and known transition probabilities. From now on we use ILP to refer to the individual learning problem and $\overline{\text{ILP}}$ to refer to the equivalent DP with known transitions.

We first derive the transition probabilities for $\overline{\text{ILP}}$. For this, let $B_t : \Pi \times A \times X \times \Xi \to \Pi$ be such that if $x = (x_1, x_2)$ and $m = \{m_t(j,\theta)\}$, then

$$B_t(\pi, j, x, m)(\theta) = \frac{g_j(x_1, \theta_j) m_t(x_2, \theta)\pi(\theta)}{\sum_{\theta'\in\Theta} g_j(x_1, \theta'_j) m_t(x_2, \theta')\pi(\theta')}$$

if the denominator is not zero. Otherwise,

$$B_t(\pi, j, x, m)(\theta) = \frac{g_j(x_1, \theta_j)\pi(\theta)}{\sum_{\theta'\in\Theta} g_j(x_1, \theta'_j)\pi(\theta')}.$$

The term on the right-hand side of the last equation is well-defined by Assumption 2. By definition, if the sequence of matching vectors is m, an agent who in t has belief π, chooses j, and observes x revises his belief to $B_t(\pi, j, x, m)$. Now let $\{\pi_t\}_{t\in\mathbb{N}}$, with $\pi_t : H_t \times \Pi \times \Xi \to \Pi$, be such that $\pi_1(j_0, x_0, \pi_0, m) = B_1(\pi_0, j_0, x_0, m)$ and $\pi_{t+1}(h_t, j_t, x_t, \pi_0, m) = B_t(\pi_t(h_t, \pi_0, m), j_t, x_t, m)$ for $t \geq 1$. By construction, $\pi_t(h_t, \pi_0, m)$ is the period t belief of an agent with history h_t when his prior is π_0 and the sequence of matching vectors is m. Notice that π_1 is independent of m as the meetings in society when $t = 0$ are uninformative. By Lemma 10 in Appendix C, for each $\pi_0 \in \Pi$ and $m \in \Xi$, the maps $\pi_t(\cdot, \pi_0, m)$ are measurable. To finish, let $q_t : H_t \times A \times \Pi \times \Xi \to \mathcal{P}(X)$ be given by

$$q_t(h_t, j, \pi_0, m) = \sum_{\theta\in\Theta} \pi_t(h_t, \pi_0, m)(\theta)\nu_t^\theta(j, m).$$

From above, for each $\pi_0 \in \Pi$ and $m \in \Xi$, $q_t(\cdot, \pi_0, m)$ is a transition probability from $H_t \times A$ into X. The maps $\{q_t(\cdot, \pi_0, m)\}$ are the transitions probabilities for $\overline{\text{ILP}}$ as a function of the prior belief and the sequence of matching vectors.

[6] In Rieder's terminology, the initial distributions for the individual learning problem are the measures $\{\tau_1(\theta)\}_{\theta\in\Theta}$, where $\tau_1 : \Theta \to \mathcal{P}(H_1)$ is the transition probability defined in Appendix B.

The set $\overline{H}_t$ of period t histories for $\overline{\text{ILP}}$ is the same as the set of period t histories for ILP, and so both problems have the same set of behavior strategies. For consistency in notation, denote the set of behavior strategies for $\overline{\text{ILP}}$ by $\overline{\Omega}$. Fix $\pi_0 \in \Pi$, $m \in \Xi$, and let $\overline{r}_j(h_t) = \sum_{\theta\in\Theta} r_j(\theta_j)\pi_t(h_t, \pi_0, m)(\theta)$ be the expected flow payoff from choosing j in period t as a function of h_t. Now let $\overline{R} : \overline{H}_\infty \to \mathbb{R}$ be such that $\overline{R}(h_\infty) = \sum_{t=1}^{\infty} \beta^{t-1}\overline{r}_{j_t}(h_\infty^t)$, where h_∞^t is the restriction of h_∞ to H_t and j_t is the period t action choice. Moreover, let $\sigma(\pi_0, h_1, f, m) = \sum_\theta \pi_0(\theta)\sigma_\theta(h_1, f, m)$. The objective in $\overline{\text{ILP}}$ for an agent with prior π_0 is to choose $f^* \in \overline{\Omega}$ such that

$$\int \overline{R}(h_1 z_\infty)\sigma(dz_\infty|\pi_0, h_1, f^*, m) = \sup_{f\in\overline{\Omega}} \int \overline{R}(h_1 z_\infty)\sigma(dz_\infty|\pi_0, h_1, f, m) \qquad \forall h_1 \in H_1,$$

where if $h_k \in \overline{H}_k$ and $z_t \in Z^t$, then $h_k z_t$ is the concatenation of h_k and z_t. It is straightforward to show that f^* is an optimal strategy for $\overline{\text{ILP}}$ if, and only if, it is an optimal strategy for ILP.

As mentioned above, $\overline{\text{ILP}}$ is a non-markovian decision problem. Nevertheless, we can analyze it using dynamic programming techniques. The Bellman equations for $\overline{\text{ILP}}$ are

$$\overline{V}_t(h_t) = \max_{j\in A}\left\{\overline{r}_j(h_t) + \beta\int \overline{V}_{t+1}(h_t, j, x)q_t(dx|h_t, j, \pi_0, m)\right\} = \overline{T}_t\overline{V}_{t+1}(h_t), \qquad t \in \mathbb{N}. \tag{2}$$

We start with a basic result. For each $f \in \overline{\Omega}$, let $f|_{h_k} = \{g_t\}_{t\in\mathbb{Z}_+}$, where $g_t : Z^t \to \Delta_{N-1}$ is such that $g_t(z_t) = f_{k+t}(h_k z_t)$. Now let $\sigma_t(\pi_0, h_t, f, m) = \sum_\theta \pi_0(\theta)\sigma_\theta(f|_{h_t}, m^t)$, where $\sigma_\theta(f|_{h_t}, m^t)$ is the Borel probability measure on Z^∞ induced by $f|_{h_t}$ and the sequence $m^t = \{m_k\}_{k=t}^{\infty}$ of matching vectors when the state of the world is θ.[7] Finally, let $\overline{V}_t^* : \overline{H}_t \to \mathbb{R}$ be given by

$$\overline{V}_t^*(h_t) = \sup_{f\in\overline{\Omega}} \int \overline{R}(h_t z_\infty)\sigma_t(dz_\infty|\pi_0, h_t, f, m). \tag{3}$$

The proof of the following result is in Appendix C.

Lemma 1. *For each $\pi_0 \in \Pi$ and $m \in \Xi$, the sequence $\{\overline{V}_t^*\}$ defined by (3) is the unique sequence of bounded and measurable functions that solves the Bellman equations (2).*

The second result we prove is central for the characterization result established at the end of this section. It shows that the sequence of maps $\{\pi_t\}$ constitutes a sufficient statistic for $\overline{\text{ILP}}$. Let $\nu_t : \Pi\times A\times\Xi \to \mathcal{P}(X)$ be given by $\nu_t(\pi, j, m) = \sum_{\theta\in\Theta} \pi(\theta)\nu_t^\theta(j, m)$ and define $\rho_t : \Pi\times A\times\Xi \to \mathcal{P}(\Pi)$ to be such that if $D \in \mathcal{B}(\Pi)$, then

$$\rho_t(\pi, j, m)(D) = \int I_D(B_t(\pi, j, x, m))\nu_t(dx|\pi, j, m). \tag{4}$$

[7]Notice that $\sigma_\theta(f|_{h_1}, m) = \sigma_\theta(h_1, f, m)$.

By Lemma 11 in Appendix C, for each $m \in \Xi$, $\rho_t(\cdot, m)$ is a continuous transition probability from $\Pi \times A$ into Π. Now let $s_j : \Pi \to \mathbb{R}$ be such that $s_j(\pi) = \sum_{\theta \in \Theta} r_j(\theta_j)\pi(\theta)$ and, for each $m \in \Xi$, consider the sequence of functional equations given by

$$V_t(\pi) = \max_{j \in A} \left\{ s_j(\pi) + \beta \int V_{t+1}(\pi')\rho_t(d\pi'|\pi, j, m) \right\} = T_t V_{t+1}, \qquad t \in \mathbb{N}. \tag{5}$$

It is convenient to introduce the operators $T_{t,j}$ such that if $v : \Pi \to \mathbb{R}$ is measurable, then

$$T_{t,j}v(\pi) = s_j(\pi) + \beta \int v(\pi')\rho_t(d\pi'|\pi, j, m).$$

Lemma 2. *For each $m \in \Xi$, there exists a unique sequence $\{V_t^*\}$ of bounded and measurable functions that satisfies (5). Moreover, for each $\pi_0 \in \Pi$, $\overline{V}_t^*(h_t) = V_t^*(\pi_t(h_t, \pi_0, m))$.*

Proof: The first part follows from Lemma 7 in Appendix A. For the second part, let $w_t : H_t \to \mathbb{R}$ be such that $w_t(h_t) = V_t^*(\pi_t(h_t))$, where the dependence of π_t on π_0 and m is omitted. By Lemma 10 in Appendix C, w_t is bounded and continuous. Moreover,

$$\begin{aligned} &\overline{r}_j(h_t) + \beta \int w_{t+1}(h_t, j, x)q_t(dx|h_t, j, m) \\ &\qquad = s_j(\pi_t(h_t)) + \beta \int V_{t+1}^*(\pi_{t+1}(h_t, j, x))\nu_t(dx|\pi_t(h_t), j, m) \\ &\qquad = s_j(\pi_t(h_t)) + \beta \int V_{t+1}^*(B_t(\pi_t(h_t), j, x, m)\nu_t(dx|\pi_t(h_t), j, m) \\ &\qquad = s_j(\pi_t(h_t)) + \beta \int V_{t+1}^*(\pi')\rho_t(d\pi'|\pi_t(h_t), j, m), \end{aligned}$$

where the first equality follows from the definitions of s_j and ν_t, the second follows from the definition of π_t, and the third follows from the definition of ρ_t. Hence, $\{w_t\}$ satisfies the Bellman equations (2) for $\overline{\text{ILP}}$, from which we can conclude that $w_t(h_t) = \overline{V}_t^*(h_t)$. □

4.2 Learning

What happens when an agent with a prior that assigns positive probability to the true type of a particular action j chooses this action an infinite number of times? If it were the case that he only observed the outcomes of his action choices, he would learn the true type of j with probability one. This result on the consistency of Bayes estimates follows from the fact that each of the available actions has a finite number of possible types, the densities $g_j(y, \theta_j)$ have finite first and second moments, and the maps $\theta_j \mapsto \mu_j(\theta_j)$ are one-to-one, see Section 10.5 in DeGroot (1970). It turns out that this is also true in the presence of the meetings in society. This is the content of Lemma 1 in Aoyagi (1998), that we state below adapted to our framework.

Fix the prior belief π_0, the sequence of matching vectors m, and the behavior strategy f. Let $\pi^j_{\theta,t} : H_\infty \to [0,1]$ be given by $\pi^j_{\theta,t}(h_\infty) = \pi_t(h^t_\infty, \pi_0, m))(\{\theta_j\} \times \Theta_{-j})$, where $\Theta_{-j} = \times_{k \neq j} \Theta_k$. Define $\sigma_\theta(f,m) \in \mathcal{P}(H_\infty)$ to be such that if $D_1 \in \mathcal{B}(Y)$ and $D_2 \in \mathcal{B}(Z^\infty)$, then

$$\sigma_\theta(f,m)(D_1 \times D_2) = \int_{D_1} \sigma_\theta(h_1, f, m)(D_2)\tau_1(dh_1|\theta),$$

where $\tau_1(\theta)(D)$ is the probability an agent has of experiencing a period one history in $D \in \mathcal{B}(H_1)$. See Appendix B for the construction of τ_1. Now let $\lambda(\pi_0, f, m)$ be the Borel probability measure on $\Theta \times H_\infty$ such that if $\Theta' \subseteq \Theta$ and $G \in \mathcal{B}(H_\infty)$, then $\lambda(\pi_0, f, m)(\Theta' \times G) = \sum_{\theta \in \Theta'} \pi_0(\theta)\sigma_\theta(f,m)(G)$. Finally, let $E_j \in \mathcal{B}(H_\infty)$ be the event where j is chosen infinitely many times.

Lemma 3. *For each $\pi_0 \in \Pi$, $m \in \Xi$, and $f \in \Omega$, $\lambda(\pi_0, f, m)((\{\theta_j\} \times \Theta_{-j}) \times E_j) > 0$ implies that $\pi^j_{\theta,t}$ converges to one on $(\{\theta_j\} \times \Theta_{-j}) \times E_j$ almost surely with respect to $\lambda(\pi_0, f, m)$.*

4.3 Characterization

We now prove the main result of this section. For this, let $r(\theta) = \max_j r_j(\theta_j)$ and define $l_j : \Pi \to \mathbb{R}$ to be such that

$$l_j(\pi) = \frac{s_j(\pi)}{1-\beta} - \left\{ \max_{k \neq j} s_k(\pi) + \frac{\beta}{1-\beta} \sum_{\theta \in \Theta} r(\theta)\pi(\theta) \right\}.$$

Notice that if π is such that $l_j(\pi) > 0$, then $s_j(\pi) > s_k(\pi)$ for all $k \neq j$. This, in turn, implies that $l_k(\pi) < 0$ for all $k \neq j$. In other words, if $\Pi^j = \{\pi \in \Pi : l_j(\pi) > 0\}$, then $\bigcap_{j \in A} \Pi^j = \emptyset$. Also notice that if $\Theta(j) = \{\theta \in \Theta : r_j(\theta_j) > r_k(\theta_k) \text{ for } k \neq j\}$ is the set of states of the world where j is the best alternative, then $\theta \in \Theta(j)$ implies that $l_j(\pi) > 0$ for all π that put sufficiently high probability on the event $\{\theta_j\} \times \Theta_{-j}$.

Theorem 1. *Fix $\pi_0 \in \Pi$ and $m \in \Xi$, and suppose $f^* = \{f^*_t\}$ is an optimal strategy for $\overline{\text{ILP}}$. If $\pi_t(h_t, \pi_0, m) \in \Pi^j$, then $f^*_t(h_t)$ puts probability one on j.*

The following result follows immediately from Theorem 1 together with Lemma 3.

Corollary 1. *No optimal strategy for $\overline{\text{ILP}}$ has an agent choosing all actions infinitely many times.*

What makes Theorem 1 useful is that the functions l_j, and so the sets Π^j, are independent of matching probabilities. Hence, this theorem provides "bounds" on behavior that hold regardless of the informational content of the meetings in society, i.e., they hold no matter the sequence m of matching vectors.

To understand the meaning of the condition $l_j(\pi) > 0$, consider the hypothetical situation where an agent learns the true value of θ if he chooses any action other than j. In this case, if his belief is π, his expected lifetime payoff from choosing $k \neq j$ is $s_k(\pi) + \beta(1-\beta)^{-1} \sum_{\theta \in \Theta} r(\theta)$. Since, in truth, learning about θ does not happen immediately (if it happens at all), the above payoff is an upper bound for this agent's lifetime expected payoff when he chooses k. The condition $l_j(\pi) > 0$ then says that even if an agent settles on j, which is not necessarily optimal, he is still better off then if he chooses any other action and behaves optimally in the periods that follow. In other words, $l_j(\pi) > 0$ means that the agent is so pessimistic about the other actions that even if he could learn θ by choosing one of them, he would still prefer to settle on j.

Motivated by the last paragraph, consider the functional equation

$$W_j(\pi) = \max_{k \neq j} \{S_{jk} W_j(\pi), S_{jj} W_j(\pi)\}, \tag{6}$$

where S_{jk} and S_{jj}, with $k \neq j$, are the operators such that if $v : \Pi \to \mathbb{R}$, then

$$S_{jk} v(\pi) = s_k(\pi) + \beta \sum_{\theta \in \Theta} v(\delta_\theta) \pi(\theta) \quad \text{and} \quad S_{jj} v(\pi) = s_j(\pi) + \beta v(\pi).$$

For each j, equation (6) has a unique continuous solution, that we denote by W_j^*. This function is, moreover, convex and such that $W_j^*(\delta_\theta) = (1-\beta)^{-1} r(\theta)$. Since $W_j^*(\pi) \geq (1-\beta)^{-1} s_j(\pi)$ for all $\pi \in \Pi$, $S_{jj} W_j^*(\pi) > S_{jk} W_j^*(\pi)$ for all $k \neq j$ when $\pi \in \Pi^j$. The proof of the next lemma is in appendix C.

Lemma 4. *Fix $j \in A$ and $m \in \Xi$. Then, for all $k \neq j$, $t \in \mathbb{N}$, and $\pi \in \Pi$,*

$$T_{t,k} V_{t+1}^*(\pi) - T_{t,j} V_{t+1}^*(\pi) \leq S_{jk} W_j^*(\pi) - S_{jj} W_j^*(\pi).$$

Proof of Theorem 1: By Lemma 2, if $\pi_t(h_t, \pi_0, m) = \pi$, then

$$\int \overline{V}_{t+1}^*(h_t, j, x) q_t(dx|h_t, j, m) = \int V_{t+1}^*(\pi') \rho_t(d\pi'|\pi_t(h_t, \pi_0, m), j, m)$$

for all actions j. Hence, the desired result follows from Lemma 4 and the principle of optimality for dynamic programming. $\square$

5 Long Run Behavior

We now establish the main result of this article. We say the full support assumption is satisfied if $\mu\{i : \Phi(i) \in A\} > 0$ for every open subset A of Π^d.[8]

Theorem 2. *Suppose $N = 2$ and the full support assumption holds. Then, in all perfect bayesian equilibria of this game the fraction of agents who choose the best alternative converges to one regardless of the state of the world.*

The conclusion of Theorem 2 is not necessarily true when the full support assumption is not satisfied. As an example, suppose that $\Theta_1 = \{\theta_1\}$, $\Theta_2 = \{\theta_{21}, \theta_{22}\}$, and $r_2(\theta_{21}) < r_1(\theta_1) < r_2(\theta_{22})$. This corresponds to the case of an one-armed bandit where the unknown arm can be of two types. Moreover, suppose that a measure one of the agents has the same prior belief π_0, where π_0 has the property that $l_1(\pi_1(\pi_0, a_0, x_0)) > 0$ for all $(a_0, x_0) \in H_1$.[9] In this particular case, the game has an unique perfect bayesian equilibrium where almost all agents always choose a_1 regardless of the value of θ. This follows immediately from Theorem 1 together with the fact that the meetings in society are uninformative when a measure one of agents choose the same action no matter the state of the world.

The following two lemmas are needed for Theorem 2. Their proofs are in appendix D. Recall that if $\theta_j \in \underline{\Theta}_j$, then $r_j(\theta_j) < r_N(\theta_N)$ for all θ_N, and that $\Theta(j)$ denotes the set of states of the world where j is the best alternative.

Lemma 5. *Let F^* be a perfect bayesian equilibrium and suppose θ is such that $\theta_j \in \underline{\Theta}_j$. Then, $m_t(j, \theta, F^*)$ converges to zero.*

By Lemma 3, an agent who chooses a particular action j infinitely often learns its true type as long as his prior belief assigns positive probability to this type. Therefore, when θ is such that $\theta_j \in \underline{\Theta}_j$, Theorem 1 and the assumption that all agents have non-dogmatic priors imply that an agent following an optimal strategy plays j only a finite number of times. Hence, in any equilibrium, the measure of agents who choose j converges to zero in the long-run for all θ such that $\theta_j \in \underline{\Theta}_j$.

[8] Suppose $\{\pi_n\}_{n \in \mathbb{Z}_+}$ is countable dense subset of the interior of Π^d and let $\{I_n\}_{n \in \mathbb{N}}$ be such that $I_n = (\frac{1}{n+1}, \frac{1}{n}]$. Now define $\Phi' : I \to \Pi$ to be such that $\Phi'(i) = \pi_0$ if $i = 0$ and $\Phi'(i) = \pi_n$ if $i \in I_n$. Notice that Φ' is measurable. If $\Phi = \Phi'$, the full support assumption is satisfied.

[9] This requires the likelihood ratios $g_1(\cdot, \theta_1)/g_2(\cdot, \theta_2)$ to be bounded away from both zero and infinity.

Lemma 6. *Let* F^* *be a perfect bayesian equilibrium and suppose* $\theta \in \Theta(j)$*. Then, under the full support assumption, there exists* $\underline{m} = \underline{m}(F^*) > 0$ *such that* $m_t(j, \theta, F^*) \geq \underline{m}$ *for all* $t \in \mathbb{N}$*.*

Lemma 6 states that if the full support assumption holds, then whenever a particular action j is the best alternative, the measure of agents who choose this action is bounded away from zero in any perfect bayesian equilibrium of this game. The proof of this result is done in two parts. We first establish that if in period zero an agent knows the true type of all action choices but j and assigns high enough probability to j's true type, then there is a positive probability that he always chooses j. This part of the argument borrows from the techniques used in the proof of Theorem 5.1 in Banks and Sundaram (1992). In the second part, we use a continuity argument to show that the same result holds for any agent with a prior that attaches sufficiently high probability to the true state of the world. The desired result then follows from the full support assumption.

To illustrate the idea behind the proof of Theorem 2, consider the case where $\Theta_1 = \{\theta_{11}, \theta_{12}\}$, $\Theta_2 = \{\theta_{21}, \theta_{22}\}$, and $r_1(\theta_{11}) < r_2(\theta_{21}) < r_1(\theta_{12}) < r_2(\theta_{22})$. By Lemma 5, if θ is such that $\theta_1 = \theta_{11}$, then the measure of agents who choose a_1 converges to zero in any equilibrium. Suppose now that $\theta_2 = \theta_{21}$ and $\theta_1 = \theta_{12}$. By Lemma 6, the measures of agents who choose a_1 are bounded away from zero in all equilibria. Suppose then, by contradiction, that the measures of agents who choose a_2 are also bounded away from zero. Any agent who chooses a_2 infinitely many times, and there is a positive measure of them who do so, learns that $\theta_2 = \theta_{12}$ by Lemma 3. At the same time, this agent also observes that the measure of agents who choose a_1 does not converge to zero in the long-run. Since he knows that this measure converges to zero if $\theta_1 = \theta_{11}$, this agent learns that $\theta_1 > \theta_2$. This, however, is a contradiction by Theorem 1. To finish, consider the case where $\theta_1 = \theta_{12}$ and $\theta_2 = \theta_{22}$. By Lemma 6, the measures of agents who choose a_2 are bounded away from zero in any equilibrium. Suppose, once more by contradiction, that the measures of agents who choose a_1 are also bounded away from zero. Any agent who chooses a_1 infinitely often learns that $\theta_1 = \theta_{12}$. At the same time, this agent observes that the fraction of agents who choose a_2 does not converge to zero. Since he knows that this fraction converges to zero when $\theta_1 = \theta_{12}$ and $\theta_2 = \theta_{21}$, he then learns that $\theta_2 > \theta_1$, a contradiction.

Proof of Theorem 2: Order the elements of Θ_1 and Θ_2, from lowest to highest, in terms of their expected rewards. This is possible since the maps $\theta_j \mapsto r_j(\theta_j)$ are one-to-one. Then, $\Theta_1 =$

$\{\theta_1^1, \ldots, \theta_1^L\}$ and $\Theta_2 = \{\theta_2^1, \ldots, \theta_2^K\}$, where $L+K \geq 3$. If G is a subset of Θ_j and H is a subset of Θ_k, with $j, k \in \{1, 2\}$, write $G < H$ when, $r_j(\theta_j) < r_k(\theta_k)$ for all $\theta_j \in A$ and $\theta_k \in B$. By assumption, $r_1(\theta_1^1) < r_2(\theta_2^1)$. There are two cases to consider, then. Either $r_1(\theta_1^L) > r_2(\theta_2^K)$ or $r_1(\theta_1^L) < r_2(\theta_2^K)$. We consider the second case only. The modifications required to adapt the argument that follows to the first case are transparent.

Since $r_2(\theta_2^K) > r_1(\theta_1^L)$, there exist $m \in \mathbb{N}$ and sets Θ_j^n, with $j \in \{1, 2\}$ and $n \in \{1, \ldots, m\}$, such that $\Theta_1 = \bigcup_{n=1}^m \Theta_1^n$, $\Theta_2 = \bigcup_{n=1}^m \Theta_2^n$, $\Theta_j^1 < \cdots < \Theta_j^m$ for all $j \in \{1, 2\}$, and $\Theta_1^n < \Theta_2^n$ for all $n \in \{1, \ldots, m\}$. Let $z_j : \Theta_j \to \{1, \ldots, m\}$ be such that $z_j(\theta_j) = n$ if $\theta_j \in \Theta_j^n$. Then, by construction, $z_1(\theta_1) \leq z_2(\theta_2)$ implies that $r_1(\theta_1) < r_2(\theta_2)$ and $z_1(\theta_1) > z_2(\theta_2)$ implies that $r_1(\theta_1) > r_2(\theta_2)$. Now let $\Upsilon : \Theta \to \{1, \ldots, 2m\}$ be such that

$$\Gamma(\theta) = \begin{cases} 2z_1(\theta_1) - 1 & \text{if} \quad z_1(\theta_1) \leq z_2(\theta_2) \\ 2z_2(\theta_2) & \text{if} \quad z_2(\theta_2) < z_1(\theta_1) \end{cases} .$$

The proof is by induction in the range of Υ.

Suppose first that $\Upsilon = 1$. Then, $\theta \in \Theta$ is such that $\theta_1 \in \underline{\Theta}_1$ and we know, from Lemma 5, that $m_t(1, \theta, F^*)$ converges to zero if F^* is a perfect bayesian equilibrium (PBE). Suppose now, by induction, that there exists $\overline{\Upsilon} \in \mathbb{N}$ such that: (i) if $\Upsilon(\theta) \leq \overline{\Upsilon}$ is even, then $m_t(2, \theta, F^*)$ converges to zero if F^* is a PBE; (ii) if $\Upsilon(\theta) \leq \overline{\Upsilon}$ is odd, then $m_t(1, \theta, F^*)$ converges to zero if F^* is a PBE. We only consider the case where $\overline{\Upsilon}$ is odd, as the argument when $\overline{\Upsilon}$ is even is identical. By assumption, if $\Upsilon(\theta) = \overline{\Upsilon} + 1$, then $r_1(\theta_1) > r_2(\theta_2)$ and $z_2 = (\overline{\Gamma} + 1)/2$. By Lemma 6, if F^* is a PBE, there exists $\underline{m} > 0$ such that $m_t(1, \theta, F^*) \geq \underline{m}$ for all $t \in \mathbb{N}$. Hence, there is a subsequence of $\{m_t(1, \theta, F^*)\}$ that converges to some $\alpha > 0$. Assume, without loss, that $m_t(1, \theta, F^*)$ itself converges to α.

Suppose then, by contradiction, that $m_t(2, \theta, F^*)$ does not converge to zero (so that $\alpha < 1$). The same argument used in the proof of Lemma 5 shows that a positive measure of agents chooses 2 infinitely many times. Now observe, by the induction hypothesis, that $m_t(1, \theta, F^*)$ converges to zero when $z_2(\theta_2) = (\overline{\Upsilon} + 1)/2$ and $z_1(\theta_1) \leq z_2(\theta_2)$, as this corresponds to the case where $\Upsilon(\theta) \leq \overline{\Upsilon}$. Hence, by Lemma 13 in Appendix D, an agent who chooses 2 infinitely often learns that $z_1 > z_2$ with probability one. This, however, implies that almost all of the agents who choose 2 infinitely many times eventually have a belief π with $l_1(\pi) > 0$, a contradiction by Theorem 1. Therefore, $m_t(2, \theta, F^*)$ must converge to zero, and so the induction hypothesis is true for $\overline{\Upsilon} + 1$. This concludes the argument. □

6 Conclusion

This paper shows how learning in society can overcome the incomplete learning results typical of multi-armed bandits. We consider an environment with a continuum of agents where each one of them faces a two-armed bandit and the unknown stochastic payoffs of each arm are the same for all agents. Information flows in a decentralized way: in every period all agents are randomly and anonymously matched in pairs and they observe the current action choice of their partner. We show that if initial beliefs are sufficiently heterogenous, then all perfect bayesian equilibria of this game are ex-post efficient; i.e., the fraction of the population choosing the best alternative converges to one in these equilibria.

References

[1] ALIPRANTIS, C. and K. BORDER (1999). *Infinite Dimensional Analysis*, 2nd Edition. Berlin: Springer-Verlag.

[2] AOYAGI, M. (1998). "Mutual Observability and the Convergence of Actions in a Multi-Armed Bandit Model", *Journal of Economic Theory*, **82**, 405-424.

[3] BALA, V. and S. GOYAL (1998). "Learning from Neighbours", *Review of Economic Studies*, **65**, 595-621.

[4] BANARJEE, A. (1992). "A Simple Model of Herd Behavior", *Quaterly Journal of Economics*, **107**, 797-817.

[5] BANKS, J. and R. SUNDARAM (1992). "Denumerable-Armed Bandits", *Econometrica*, **60**, 1071-1096.

[6] BIKHCHANDANI, S., D. HIRSHLEIFER, and I. WELCH (1992). " A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades", *Journal of Political Economy*, **100**, 992-1026.

[7] BOLTON, P. and C. HARRIS (1999). "Strategic Experimentation", *Econometrica*, **67**, 349-374.

[8] CRIMALDI, I. and L. PRATELLI (2005). "Two Inequalities for Conditional Expectations and Convergence Results for Filters", *Statistics & Probability Letters*, **74**, 151-162.

[9] DEGROOT, M. (1970). *Optimal Statistical Decisions.* New York: McGraw-Hill.

[10] EASLEY, D. and N. KIEFER (1988). "Controlling a Stochastic Process with Unkown Parameters", *Econometrica*, **56**, 1045-1064.

[11] ELLISON, G. and D. FUDENBERG (1993). "Rules of Thumb for Social Learning", *Journal of Political Economy*, **101**, 612-643.

[12] ELLISON, G. and D. FUDENBERG (1995). "Word-of-Mouth Communication and Social Learning", *Quaterly Journal of Economics*, **110**, 93-125.

[13] HINDERER, K. (1970). *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*, Lecture Notes in Operations Research and Mathematical System, Vol. 33. Berlin and New York: Springer-Verlag.

[14] KELLER, G. RADY, S., and M. CRIPPS (2005). "Strategic Experimentation with Exponential Bandits", *Econometrica*, **73**, 39-69.

[15] RIEDER, U. (1975). "Bayesian Dynamic Programming", *Advances in Applied Probability*, **7**, 330-348.

[16] SHIRYAEV, A. (1996). *Probability*, Graduate Texts in Mathematics. New-York: Springer-Verlag.

[17] SMALLWOOD, D. and J. CONLISK (1979). "Product Quality in Markets where Consumers are Imperfectly Informed". *Quaterly Journal of Economics*, **93**, 1-23.

[18] SMITH, L., and P. SORENSEN (2000). "Pathological Outcomes of Observational Learning", *Econometrica*, **68**, 371-398.

Appendix A

In this appendix we establish some auxiliary results. We begin with a generalization of the contraction mapping theorem used in dynamic programming. For this, let $\{S_t\}$ be a sequence of Polish spaces, Ξ be a Polish space, and A be a finite set. Denote a typical element of S_t by s_t, a typical element of Ξ by ξ, and a typical element of A by a. Moreover, for each $t \in \mathbb{N}$, let $q_t : S_t \times A \times \Xi \to \mathcal{P}(S_{t+1})$ be such that $q_t(\,\cdot\,, \xi)$ is a transition probability from $S_t \times A$ into S_{t+1} for all $\xi \in \Xi$. Finally, for each $t \in \mathbb{N}$, let $r_t : S_t \times A \to \mathbb{R}$ be bounded.

Fix $\xi \in \Xi$, let $\beta \in (0, 1)$, and consider the sequence of functional equations given by

$$V_t(s_t) = \max_{a \in A} \left\{ r_t(s_t, a) + \beta \int V_{t+1}(s_{t+1}) q_t(ds_{t+1}|s_t, a, \xi) \right\} = T_t V_{t+1}(s_t), \qquad t \in \mathbb{N}. \tag{A.1}$$

Lemma 7. *For each $\xi \in \Xi$, there is a unique sequence $\{V_t^*\}$ of bounded and measurable functions that satisfies* (A.1)*.*

Proof: Fix $\xi \in \Xi$ and let $\mathrm{B}_\infty = \times_{t \in \mathbb{N}} \mathrm{B}_b(S_t)$. The space B_∞ is complete and metrizable when endowed with the product topology. A metric d on B_∞ that is compatible with the product topology is the following: if $g = \{g_t\}, h = \{h_t\} \in \mathrm{B}_\infty$, then

$$d(g, h) = \max_{t \in \mathbb{N}} \frac{c_t ||g_t - f_t||_{\sup}}{1 + ||f_t - g_t||_{\sup}} \tag{A.2}$$

where $\{c_t\}$ is any sequence of strictly positive numbers that converges to zero.

Consider now the map T such that if $v = \{v_t\} \in \mathrm{B}_\infty$, then $Tv = \{T_t v_{t+1}\}$. By assumption, if $v \in \mathrm{B}_b(S_{t+1})$, the function $\int v(s_{t+1}) q_t(ds_{t+1}|s_t, a, \xi)$ is jointly measurable in s_t and a and is bounded in s_t. Hence, since A is finite, $T_t v$ is a bounded and measurable function of s_t when $v \in \mathrm{B}_b(S_{t+1})$. Consequently, T maps B_∞ into itself. If we show that T is a contraction, the desired result is a consequence of the Banach fixed point theorem.

Suppose, by contradiction, that there exist $g, h \in \mathrm{B}_\infty$ such that $d(Tg, Th) \geq d(g, h)$. This implies that for all $n \in \mathbb{N}$ there exists $t(n) \in \mathbb{N}$ such that

$$\begin{aligned} \frac{c_{t(n)} ||T_{t(n)} g_{t(n)} - T_{t(n)} h_{t(n)}||_{\sup}}{1 + ||T_{t(n)} g_{t(n)} - T_{t(n)} h_{t(n)}||_{\sup}} &> \left(1 - \frac{1}{n}\right) \max_t \frac{c_t ||g_t - h_t||_{\sup}}{1 + ||g_t - h_t||_{\sup}} \\ &\geq \left(1 - \frac{1}{n}\right) \frac{c_{t(n)} ||g_{t(n)} - h_{t(n)}||_{\sup}}{1 + ||g_{t(n)} - h_{t(n)}||_{\sup}}. \end{aligned}$$

Rearranging terms, we have that

$$\frac{||T_{t(n)}g_{t(n)} - T_{t(n)}h_{t(n)}||_{\sup} - \beta||g_{t(n)} - h_{t(n)}||_{\sup}}{d_{t(n)}} >$$
$$\left(1 - \frac{1}{n} - \beta\right) \frac{||g_{t(n)} - h_{t(n)}||_{\sup}}{d_{t(n)}} - \frac{1}{n} \underbrace{\frac{||T_{t(n)}g_{t(n)} - T_{t(n)}h_{t(n)}||_{\sup}||g_{t(n)} - h_{t(n)}||_{\sup}}{d_{t(n)}}}_{e_{t(n)}},$$

where $d_t = (1 + ||T_tg_t - T_th_t||_{\sup})(1 + ||g_t - h_t||_{\sup})$. The sequence $\{e_{t(n)}\}$ is, however, bounded, and so $\frac{1}{n}e_{t(n)}$ converges to zero. Since $1 - \frac{1}{n} - \beta$ converges to $1 - \beta$, we then have that the right-hand side of the above inequality is positive if n is sufficiently large. Therefore,

$$||T_{t(n)}g_{t(n)} - T_{t(n)}h_{t(n)}||_{\sup} > \beta||g_{t(n)} - h_{t(n)}||_{\sup}$$

for n large enough, contradicting the fact that the maps T_t are contractions of modulus β. We can then conclude that T contraction. $\square$

Lemma 8. *Let S be a metric space X be a Banach space. Suppose $F : [0, 1] \to \mathrm{B}_b(S, X)$ is strongly Lebesgue measurable and let $G_s : [0, 1] \to X$ be such that $G_s(i) = F(i)(s)$, where s is a fixed element of S. Then, G_s is strongly Lebesgue measurable for all $s \in S$.*

Proof: Let $|| \cdot ||_X$ denote the norm in X and $|| \cdot ||$ denote the norm in $\mathrm{B}_b(S, X)$. By assumption, there is a sequence $\{F_n\}$, with $F_n : [0, 1] \to \mathrm{B}_b(S, X)$ simple, such that $||F_n(i) - F(i)|| \to 0$ for almost all $i \in [0, 1]$. Fix $s \in S$ and define $G_{n,s} : [0, 1] \to X$ to be such that $G_{n,s}(i) = F_n(i)(s)$. Then, $\{G_{n,s}\}$ is a sequence of simple functions. Moreover,

$$||G_{n,s}(i) - G_s(i)||_X = ||F_n(i)(s) - F(i)(s)||_X \leq ||F_n(i) - F(i)||,$$

and so $||G_{n,s}(i) - G_s(i)|| \to 0$ for almost all $i \in [0, 1]$. Hence, G_s is strongly Lebesgue measurable. $\square$

Corollary 2. *Let S, X, and F be as in Lemma 8. Suppose $\Phi : [0, 1] \to S$ is measurable and has a countable range. Then, $G : [0, 1] \to X$ given by $G(i) = F(i)(\Phi(i))$ is strongly Lebesgue measurable.*

Proof: Let $\{s_n\} \subset S$ be the range of Φ and notice that $G(i) = \sum_{n=1}^{\infty} G_{s_n}(i) I_{\{s_n\}}(\Phi(i))$, where G_s is defined as above. The desired result now follows from Lemma 8 and the fact that the pointwise limit of a sequence of strongly Lebesgue measurable functions is strongly Lebesgue measurable, see Lemma 11.37 in Aliprantis and Border (1999). $\square$

For the next two results, recall that Σ denotes the σ-algebra of the Lebesgue measurable subsets of the unit interval $[0,1]$.

Lemma 9. *Let S be a separable metric space, $\lambda : [0,1] \to \mathcal{P}(S)$ be a transition probability from $([0,1],\Sigma)$ into S, and suppose $F : [0,1] \to \mathrm{B}_b(S)$ is strongly Lebesgue measurable. The function $h : [0,1] \to \mathbb{R}$ such that $h(i) = \int F(i)(s)\lambda(ds|i)$ is Lebesgue measurable.*[10]

Proof: By Lemma 11.36 in Aliprantis and Border (1999), F is Lebesgue measurable and there exist a separable subset X of $\mathrm{B}_b(S)$ and a Lebesgue measurable subset I_0 of $[0,1]$ such that $\mu(I_0) = 0$ and $F(i) \in X$ for all $i \notin I_0$. Define $T : [0,1] \times X \to \mathbb{R}$ to be such that $T(i,f) = \int f(s)\lambda(ds|i)$. Notice that T is a Carathéodory function; i.e., $T(i,\cdot) : X \to \mathbb{R}$ is continuous for each $i \in [0,1]$, and $T(\cdot, f) : [0,1] \to \mathbb{R}$ is Lebesgue measurable for each $f \in X$. Hence, by Lemma 4.50 in the same book, T is $\Sigma \otimes \mathcal{B}(X)$ measurable. Now let $\widehat{f}$ be an element of X and define $F' : [0,1] \to \mathrm{B}_b(S)$ to be such that $F'(i) = F(i)$ if $i \in (I_0)^c$ and $F'(i) = \widehat{f}$ if $i \in I_0$. Since $\Psi : [0,1] \to X \times I$ given by $\Psi(i) = (F'(i), i)$ is Lebesgue measurable, so is $T \circ \Psi : [0,1] \to \mathbb{R}$. Because $T \circ \Psi$ differs from h on a set of Lebesgue measure zero, we have the desired result. $\square$

Corollary 3. *Let S be a metric space, and S_1, S_2 be separable metric spaces. Suppose $\nu \in \mathcal{P}(S_2)$, $\lambda_1 : [0,1] \to \mathcal{P}(S_1)$ is a transition probability from $([0,1],\Sigma)$ into S_1, $F : [0,1] \to \mathrm{B}_b(S, \mathrm{B}_b(S_1))$ is strongly Lebesgue measurable, and $\Phi : I \to S$ is measurable and has a countable range. Define $\lambda_2 : [0,1] \to \mathcal{P}(S_1 \times S_2)$ to be such that if $D_1 \in \mathcal{B}(S_1)$ and $D_2 \in \mathcal{B}(S_2)$, then*

$$\lambda_2(i)(D_1 \times D_2) = \left(\int_{D_1} F(i)(\Phi(i))(s_1)\lambda_1(ds_1|i) \right) \nu(D_2).$$

The map λ_2 is a transition probability from $([0,1],\Sigma)$ into $S_1 \times S_2$.

Proof: Let $\mathcal{D}$ be the subset of $\mathcal{B}(S_1 \times S_2)$ such that if $D \in \mathcal{D}$, then the map $i \mapsto \lambda_2(i)(D)$ is Lebesgue measurable. By Lemmas 8 and 9, $\mathcal{D}$ contains the algebra generated by the rectangles of $S_1 \times S_2$. Moreover, a straightforward argument shows that $\mathcal{D}$ is a monotone class. Hence, by the monotone class lemma, see Lemma 4.12 in Aliprantis and Border (1999), $\mathcal{D} = \mathcal{B}(S_1 \times S_2)$, and the desired result holds. $\square$

[10] Notice that if $g : [0,1] \times S \to \mathbb{R}$ is such that $g(i,s) = F(i)(s)$, then it is not necessary that g is jointly measurable. Hence, this result is not immediate.

Appendix B

Here we construct the map M that takes strategy profiles into (infinite) sequences of matching vectors. The starting point is the map $\tau_2 : [0,1] \times \Theta \times \Lambda \to \mathcal{P}(H_2)$. For this, let $\tau_1 : \Theta \to \mathcal{P}(H_1)$ be the transition probability such that if $D \in \mathcal{B}(Y)$, then $\tau_1(\theta)(\{j\} \times D \times \{k\}) = \frac{1}{N^2}\mu_j(D|\theta_j)$. By definition, $\tau_1(\theta)(D)$ is the probability that any agent in the economy experiences a period zero action-observation pair that lies in $D \in \mathcal{B}(H_1)$. Fix $\theta \in \Theta$ and $F \in \Lambda$, and suppose $D_1 \in \mathcal{B}(Z)$ and $D_2 \in \mathcal{B}(Y)$. Now define the set function $\widehat{\tau}_2(i)$ to be such that

$$\widehat{\tau}_2(i)(D_1 \times \{j\} \times D_2 \times \{k\}) = \left(\int_{D_1} \langle F_1(i)(\Phi(i))(h_1), e_j \rangle \tau_1(dh_1|\theta) \right) \mu_j(D_2|\theta_j) m_1(k, \theta, F).$$

The unique extension of $\widehat{\tau}_2(i)$ to $\mathcal{B}(H_2)$ is $\tau_2(i, \theta, F)$. By Corollary 3 in Appendix A, for each $\theta \in \Theta$ and $F \in \Lambda$, the map $\tau_2(\,\cdot\,, \theta, F)$ is a transition probability from $([0,1], \Sigma)$ into H_2. From τ_2 we can construct, by the process described in Section 3, the map $m_2 : A \times \Theta \times \Lambda \to [0,1]$ such that $m_2(j, \theta, F)$ is the fraction of agents who choose j in period two when the state of the world is θ and the strategy profile under play is F.

Suppose then, by induction, that for some $t \geq 2$ there exist:

(1) A map $\tau_t : [0,1] \times \Theta \times \Lambda \to \mathcal{P}(H_t)$ such for each $\theta \in \Theta$ and $F \in \Lambda$, $\tau_t(\,\cdot\,, \theta, F)$ is the transition probability from $([0,1], \Sigma)$ into H_t with the property that if $D \in \mathcal{B}(H_t)$, then $\tau_t(i, \theta, F)(D)$ is the probability that agent i experiences a period t history in D when the state of the world is θ and the strategy profile under play is F.

(2) A map $m_t : A \times \Theta \times \Lambda : [0,1] \to [0,1]$ such that $m_t(j, \theta, F)$ is the fraction of agents who choose j in period t when the state of the world is θ and the strategy profile under play is F.

Fix $\theta \in \Theta$ and $F \in \Lambda$, and define $\widehat{\tau}_{t+1}(i)$ to be the set function such that

$$\widehat{\tau}_{t+1}(i)(D_1 \times \{j\} \times D_2 \times \{k\}) = \left(\int_{D_1} \langle F_t(i)(\Phi(i))(h_t), e_j \rangle \tau_t(dh_t|\theta) \right) \mu_j(D_2|\theta_j) m_t(k, \theta, F)$$

for all $D_1 \in \mathcal{B}(H_t)$ and $D_2 \in \mathcal{B}(Y)$. This set function admits an unique extension to $\mathcal{B}(H_{t+1})$ that we denote by $\tau_{t+1}(i, \theta, F)$. By definition, $\tau_{t+1}(i, \theta, F)(D)$ is the probability that agent i experiences a period $t+1$ history in $D \in \mathcal{B}(H_{t+1})$ as a function of the state of the world θ and the strategy profile F. Corollary 3 in Appendix A implies that the map $\tau_{t+1}(\,\cdot\,, \theta, F)$ is a transition probability from $([0,1], \Sigma)$ in H_{t+1}.

Now define $p_{t+1} : [0,1] \times A \times \Theta \times \Lambda \to [0,1]$ to be such that

$$p_{t+1}(i,j,\theta,F) = \int \langle F_{t+1}(i)(\Phi(i))(h_{t+1}), e_j \rangle \tau_{t+1}(dh_{t+1}|i,\theta,F).$$

This map is well-defined by Theorem 13.4 in Aliprantis and Border (1999) together with Corollary 2 in Appendix A. By Lemma 9 in Appendix A, $p_{t+1}(\,\cdot\,,j,\theta,F)$ is Lebesgue measurable for all $j \in A$, $\theta \in \Theta$, and $F \in \Lambda$, and so the map $m_{t+1} : A \times \Theta \times \Lambda \to [0,1]$ given by

$$m_{t+1}(j,\theta,F) = \int p_{t+1}(i,j,\theta,F)\mu(di)$$

is also well-defined. By construction, $m_{t+1}(j,\theta,F)$ is the fraction of agents who choose j in period $t+1$ when the state of the world is θ and the strategy profile under play is F.

We can then conclude, by induction, that there exists a map $M = \{M_t\} : \Lambda \to \Xi$ such that $M_t(F) = \{m_t(j,\theta,F)\}_{j \in A, \theta \in \Theta}$ is the vector of period t matching probabilities if F is the strategy profile being played.

Appendix C

Lemma 10. *For each $\pi_0 \in \Pi$ and $m \in \Xi$, the maps $\pi_t(\cdot,\pi_0,m) : H_t \to \Pi$ are measurable.*

Proof: Fix π_0 and $m = \{m_t\}$. First notice that Assumption 2 implies that B_t is continuous in x_1, and so is jointly continuous in (j,x), for all $t \in \mathbb{N}$. In particular, $\pi_1(\,\cdot\,,\pi_0,m) = B_1(\pi_0,\,\cdot\,,m)$ is measurable. Suppose then, by induction, that there exists $k \in \mathbb{N}$ such that $\pi_k(\,\cdot\,,\pi_0,m)$ is measurable. If we show that the maps B_t are jointly measurable in (π,j,x), we are done, since this implies that $\pi_{k+1}(\,\cdot\,,\pi_0,m) = B_k(\pi_k(\,\cdot\,,\pi_0,m),\,\cdot\,,m)$ is measurable. For this, fix j and x, and let $\{m_n\}$ be given by $m_{t,n} = \frac{1}{n}\widehat{m}_t + \frac{n-1}{n}m_t$, where $\widehat{m} = \{\widehat{m}_t\}$ is such that $\widehat{m}_t(j,\theta) \equiv \frac{1}{S+1}$. Moreover, let $B_{t,n} : \Pi \to \Pi$ be such that $B_{t,n}(\pi) = B_t(\pi,j,x,m_n)$. By construction, $B_{t,n}$ is continuous for all $t,n \in \mathbb{N}$, as its denominator is always positive. It is straightforward to show that for each $\pi \in \Pi$, $B_{t,n}(\pi)$ converges to $B_t(\pi,j,x,m)$. Therefore, $B_t(\,\cdot\,,j,x,m)$ is measurable. By Lemma 4.50 in Aliprantis and Border (1999), we can then conclude that B_t is jointly measurable in (π,j,x), the desired result. $\square$

Lemma 11. *For each $m \in \Xi$, the map $\rho_t(\,\cdot\,, m)$ given by (4) is a continuous transition probability from $\Pi \times A$ in Π.*

Proof: Fix $m \in \Xi$. Let $h : \Pi \to \mathbb{R}$ be a continuous function and suppose $\{\pi_n\}$ is a sequence in Π that converges to some $\pi \in \Pi$. Then, by the definition of ρ_t,

$$\begin{aligned}
&\left| \int h(\pi')\rho_t(d\pi'|\pi_n, j, m) - \int h(\pi')\rho_t(d\pi'|\pi, j, m) \right| \\
&\qquad \leq \left| \int h(B_t(\pi_n, j, x, m))\nu(dx|\pi_n, j, m) - \int h(B_t(\pi_n, j, x, m))\nu(dx|\pi, j, m) \right| \\
&\qquad + \left| \int h(B_t(\pi_n, j, x, m))\nu(dx|\pi, j, m) - \int h(B_t(\pi, j, x, m))\nu(dx|\pi, j, m) \right|. \qquad (C.3)
\end{aligned}$$

It is straightforward to show $\nu(\pi_n, j, m)$ converges to $\nu(\pi, j, m)$ in norm.[11] Since $h \circ B_t$ is a bounded function, the first term on the right-hand side of (C.3) converges to zero. Now observe that $B_t(\pi_n, j, x, m)$ converges to $B_t(\pi, j, x, m)$ for $\nu(\pi, j, m)$-almost all $x \in X$. Hence, by the dominated convergence theorem, the second term on the right-hand side of (C.3) also converges to zero. □

Proof of Lemma 1: First notice that $\overline{V}_t^*$ is bounded. Now define $\overline{V}_{t,n} : H_t \to \mathbb{R}$, with $t \in \mathbb{N}$ and $n \in \mathbb{Z}_+$, to be such that $\overline{V}_{t,0} \equiv 0$ and $\overline{V}_{t,n} = \overline{T}_t \overline{V}_{t+1,n-1}$. Because the maps $q_t(\cdot, \pi_0, m)$ are transition probabilities from $H_t \times A$ into H_{t+1}, the operator $\overline{T}_t$ maps measurable functions into measurable functions. Hence, for each $t \in \mathbb{N}$, the elements of $\{\overline{V}_{t,n}\}_{n \in \mathbb{Z}_+}$ are measurable. Since $\{\overline{V}_{t,n}\}$ converges pointwise to $\overline{V}_t^*$ for all $t \in \mathbb{N}$, see Theorem 14.5 in Hinderer (1970), $\overline{V}_t^*$ is also measurable. To finish, observe, by Theorem 14.4 in the same book, that $\{\overline{V}_t^*\}$ solves (2). The desired result is then a consequence of Lemma 7 in Appendix A. □

Proof of Lemma 4: Consider the map $T = \{T_t\} : \mathrm{B}_b(\Pi)^\infty \to \mathrm{B}_b(\Pi)^\infty$ such that if $v = \{v_t\} \in \mathrm{B}_b(\Pi)^\infty$, then $Tv = \{T_t v_{t+1}\} = \{\max_{j \in A} T_{t,j} v_{t+1}\}$. We know, from the proof of Lemma 7, that T is a contraction in $\mathrm{B}_b(\Pi)^\infty$. Now let B_j be the subset of $\mathrm{B}_b(\Pi)^\infty$ such that if $v = \{v_t\} \in \mathrm{A}_j$, then

$$T_{t,k} v_{t+1}(\pi) - T_{t,j} v_{t+1}(\pi) \leq S_{jk} W_j^*(\pi) - S_{jj} W_j^*(\pi)$$

for all $k \neq j$, $\pi \in \Pi$, and $t \in \mathbb{N}$. This set is a closed subset of $\mathrm{B}_b(\Pi)^\infty$. If we show that T maps B_j into itself, then $\{V_t^*\} \in \mathrm{B}_j$ by a standard argument. In what follows we omit π when convenient. It is also convenient to introduce the operators $Q_{t,j}$ such that $T_{t,j} v = s_j + \beta Q_{t,j} v$.

[11] Recall that if S is a metric space and $\nu \in \mathcal{P}(S)$, then the norm of ν is the supremum of $\int f(s)\nu(ds)$ over all $f \in \mathrm{B}_b(S)$ such that $||f||_{\sup} \leq 1$.

Let $v = \{v_t\} \in \mathrm{B}_j$. First notice that

$$\begin{aligned} T_{t+1}v_{t+2} &= T_{t+1,j}v_{t+2} + \max_{k\neq j}\{\,0\,, T_{t+1,k}v_{t+2} - T_{t+1,j}v_{t+2}\} \\ &\leq T_{t+1,j}v_{t+2} + \max_{k\neq j}\{\,0\,, S_{jk}W_j^* - S_{jj}W_j^*\} \\ &= T_{t+1,j}v_{t+2} - S_{jj}W_j^* + W_j^*. \end{aligned}$$

A similar argument shows that $T_{t+1}v_{t+2} \geq T_{t+1,k}v_{t+2} - S_{jk}W_j^* + W_j^*$ if $k \neq j$. Since $Q_{t,k}Q_{t+1,j}v_{t+2} = Q_{t,j}Q_{t+1,k}v_{t+2}$ for all $k, j \in A$, we then have that $k \neq j$ implies that

$$\begin{aligned} &T_{t,k}T_{t+1}v_{t+2} - T_{t,j}T_{t+1}v_{t+2} \\ &\quad\leq T_{t,k}T_{t+1,j}v_{t+2} - T_{t,k}S_{jj}W_j^* + T_{t,k}W_j^* - T_{t,j}T_{t+1,k}v_{t+2} + T_{t,j}S_{jk}W_j^* - T_{t,j}W_j^* \\ &\quad= (1-\beta)(r_k - r_j) + \beta T_{t,k}(W_j^* - S_{jj}W_j^*) + T_{t,j}S_{jk}W_j^* - T_{t,j}W_j^*. \end{aligned} \tag{C.4}$$

Now observe that $W_j^* - S_{jj}W_j^* = -r_j + (1-\beta)W_j^*$ is convex. By Lemma 3.1 in Banks and Sundaram (1992), for each $j \in A$ and $t \in \mathbb{N}$, the operator $Q_{t,j}$ maps convex functions into convex functions. Hence, by repeated application of Jensen's inequality,

$$Q_{t,k}(W_j^* - S_{jj}W_j^*) \leq \underbrace{Q_{t,1}\cdots Q_{t,N}}_{Q_t}(W_j^* - S_{jj}W_j^*) \leq \lim_{n\to\infty}(Q_t)^n(W_j^* - S_{jj}W_j^*),$$

where $(Q_t)^n$ is the nth iterate of Q_t. By Lemma 12 below, if $v : \Pi \to \mathbb{R}$ is continuous, then $(Q_t)^n v$ converges pointwise to $\sum_{\theta\in\Theta} v(\delta_\theta)\pi(\theta)$ for all $t \in \mathbb{N}$. Therefore,

$$\begin{aligned} \beta Q_{t,k}(W_j^* - S_{jj}W_j^*) &\leq -\beta r_j + \beta(1-\beta)\sum_{\theta\in\Theta} W_j^*(\delta_\theta)\pi(\theta) \\ &= (1-\beta)S_{jk}W_j^* - (1-\beta)(r_k - r_j) - r_j. \end{aligned} \tag{C.5}$$

To finish, notice that: (A) $S_{jk}W_j^*$ is linear, and so $T_{t,j}S_{jk}W_j^* = r_j + \beta S_{jk}W_j^*$; and (B) W_j^* is convex, and so $T_{t,j}W_j^* \geq S_{jj}W_j^*$ by Jensen's inequality. From (A), (B), (C.4), and (C.5), we can then conclude that if $\{w_t\} = \{T_t v_{t+1}\}$, then

$$T_{t,k}w_{t+1}(\pi) - T_{t,j}w_{t+1}(\pi) \leq S_{jk}W_j^*(\pi) - S_{jj}W_j^*(\pi)$$

for all $k \neq j$, $\pi \in \Pi$, and $t \in \mathbb{N}$, the desired result. □

Lemma 12. *Suppose $v : \Pi \to \mathbb{R}$ is continuous. Then, for all $t \in \mathbb{N}$ and $\pi \in \Pi$, $(Q_t)^n v(\pi)$ converges to $\sum_{\theta \in \Theta} v(\delta_\theta)\pi(\theta)$ as $t \to \infty$.*

Proof: We omit the dependence of $\nu_t^\theta(j, m)$, $\rho_t(\pi, j, m)$, and $B_t(\pi, j, x, m)$ on both m and t, as these parameters play no role in the argument. Accordingly, we drop the dependence of Q_t on t.

Let $W = X^N$ and denote a typical element of this set by w. Then, W is the set of all possible observations after N action choices and meetings in society. Now let $B(\pi, w) \in \Pi$ be the updated belief when π is the prior, actions 1 to N are consecutively chosen, and $w \in W$ is observed. Moreover, let Ψ_∞^θ be the Borel measure on W^∞ that extends the measures $\Psi_n^\theta = \times_{t=1}^n \Psi^\theta$, where $\Psi^\theta = \times_{j=1}^N \nu^\theta(j)$. To finish, define $\{B_n\}_{n \in \mathbb{N}}$, with $B_n : \Pi \times W^\infty \to \Pi$, to be such that if $w_\infty = \{w_n\} \in W^\infty$, then: (i) $B_1(\pi, w_\infty) = B(\pi, w_1)$; (ii) $B_n(\pi, w_\infty) = B(B_{n-1}(\pi, w_\infty), w_n)$ for $n > 1$.[12] Then, if $v : \Pi \to \mathbb{R}$ is measurable,

$$Q^n v(\pi) = \sum_{\theta \in \Theta} \pi(\theta) \int v(B_n(\pi, w_\infty)) \Psi_\infty^\theta(dw_\infty).$$

Fix θ. Lemma 3 implies that if $\pi \in \operatorname{int} \Pi$, the interior of Π, then $\{B_n(\pi, \cdot\,)\}_{n \in \mathbb{N}}$ converges Ψ_∞^θ-almost surely to δ_θ. Fix $\pi \in \operatorname{int} \Pi$, let $\epsilon > 0$, and suppose $v : \Pi \to \mathbb{R}$ is continuous. For any $\kappa > 0$, let $N_{n,\kappa}^\theta = \{||B_n(\pi, \cdot\,) - \delta_\theta|| > \kappa\}$. Because almost sure convergence implies convergence in measure, $\Psi_\infty^\theta(N_{n,\kappa}^\theta) \to 0$ as $n \to \infty$. Therefore, there exists $k_0(\kappa, \theta) \in \mathbb{N}$ such that if $n \geq k_0(\kappa, \theta)$, then $\Psi_\infty^\theta(N_{n,\kappa}^\theta) \leq \epsilon \overline{M}/4$, where $\overline{M} = ||v||_{\sup}$. Now let $\overline{\delta} > 0$ be such that if $||x - x'|| < \overline{\delta}$, then $||v(x) - v(x')|| < \frac{\epsilon}{2}$. Since $\Psi_\infty^\theta((N_{n,\overline{\delta}}^\theta)^c) \leq 1$,

$$\begin{aligned}
\left| \int v(B_n(\pi, w_\infty)) \Psi_\infty^\theta(dw_\infty) - v(\delta_\theta) \right| &\leq \int |v(B_n(\pi, w_\infty)) - v(\delta_\theta)| \Psi_\infty^\theta(dw_\infty) \\
= \int_{N_{n,\overline{\delta}}^\theta} |v(B_n(\pi, w_\infty)) - v(\delta_\theta)| \Psi_\infty^\theta(dw_\infty) &+ \int_{(N_{n,\overline{\delta}}^\theta)^c} |v(B_n(\pi, w_\infty)) - v(\delta_\theta)| \Psi_\infty^\theta(dw_\infty) \\
\leq \int_{N_{n,\overline{\delta}}^\theta} |v(B_n(\pi, w_\infty)) - v(\delta_\theta)| \Psi_\infty^\theta(dw_\infty) &+ \frac{\epsilon}{2}.
\end{aligned}$$

We can then conclude that if $n \geq k_0(\overline{\delta}, \theta)$, then

$$\left| \int v(B_n(\pi, w_\infty)) \Psi_\infty^\theta(dw_\infty) - v(\delta_\theta) \right| \leq 2\overline{M} \Psi_\infty^\theta(N_{n,\overline{\delta}}^\theta) + \frac{\epsilon}{2} = \epsilon.$$

Since Θ is finite, $Q^n v(\pi) \to \sum_{\theta \in \Theta} \pi(\theta) v(\delta_\theta)$. To finish, notice that $Q^n v(\pi) = \sum_{\theta \in \Theta} \pi(\theta) v(\delta_\theta)$ for all $n \in \mathbb{N}$ when π belongs to the boundary of Π. □

[12] The functions B_n are measurable by Lemma 10.

Appendix D

Proof of Lemma 5: Let F^* be a perfect bayesian equilibrium. By Appendix B, there exists a sequence of maps $\tau_t : [0,1] \times \Theta \to \mathcal{P}(H_t)$ such that: (i) if $D \in \mathcal{B}(H_t)$, then $\tau_t(i,\theta)(D)$ is the probability that agent i experiences a period t history in D when the state of the world is θ and the strategy profile is F^*; (ii) $\tau_t(\,\cdot\,,\theta)$ is a transition probability from $([0,1],\Sigma)$ into H_t. By Kolmogorov's extension theorem, for each $i \in [0,1]$ and $\theta \in \Theta$, the probability measures $\tau_t(i,\theta)$ admit an unique extension to $\mathcal{P}(H_\infty)$ that we denote by $\tau(i,\theta)$. By definition, $\tau(i,\theta)(D)$ is the probability that i experiences an infinite history in $D \in \mathcal{B}(H_\infty)$ when the state of the world is θ and the strategy profile is F^*. Let $\mathcal{H}_t = \{G \subset H_\infty : G = D \times Z^{-t}, \text{ where } D \in \mathcal{B}(H_t)\}$ and $\mathcal{D}$ be the subset of $\mathcal{B}(H_\infty)$ such that if $D \in \mathcal{D}$, then $i \mapsto \tau(i,\theta)(D)$ is Lebesgue measurable. By construction, $\mathcal{D}$ contains $\bigcup_{t=1}^{\infty} \mathcal{H}_t$. Since $\mathcal{D}$ is also a monotone class, the monotone class lemma implies that $\mathcal{D} = \mathcal{B}(H_\infty)$. Hence, for each $\theta \in \Theta$, the map $\tau(\,\cdot\,,\theta)$ is a transition probability from $([0,1],\Sigma)$ into H_∞.

Let $E_{t,j} \subset H_\infty$ be the event where j is chosen in period t. By construction,

$$m_t(j,\theta,F^*) = \int \tau(i,\theta)(E_{t,j})\mu(di).$$

Now let $E_j = \bigcap_{t=1}^{\infty} E_j^t$, where $E_j^t = \bigcup_{m=t}^{\infty} E_{m,j} \supset E_{t,j}$; i.e., E_j denotes the event where j is chosen infinitely many times. Since $E_j^t \downarrow E_j$, $\tau(i,\theta)(E_j) = \lim_t \tau(i,\theta)(E_j^t)$ for all $i \in [0,1]$. Therefore, $m_t(j,\theta,F^*)$ converges to zero if $\tau(i,\theta)(E_j) = 0$ almost surely. Suppose then, by contradiction, that $m_t(j,\theta,F^*)$ does not converge to zero. This implies that there is $I' \subseteq [0,1]$ with $\mu(I') > 0$ such that $\tau(i,\theta)(E_j) > 0$ for all $i \in I'$. Consequently, a positive measure of agents chooses j an infinite number of times. Since all agents have non-dogmatic priors, Lemma 3 implies that any agent who chooses j infinitely often learns its true type with probability one. In particular, if π_t denotes the period t belief of such an agent, there is $t_0 \in \mathbb{N}$ such that if $t \geq t_0$, then $l_N(\pi_t) > 0$ with probability one, a contradiction by Theorem 1. $\square$

Proof of Lemma 6: Let F^* be a perfect bayesian equilibrium and suppose the state of the world is $\widehat{\theta} \in \Theta(j)$. We divide the argument in several small steps.

(A) For each $t \in \mathbb{N}$, let $\mathcal{X}_t = \{G \subseteq X^\infty \mid G = H \times X^{-t}, \text{ where } H \in \mathcal{B}(X^t)\}$. Recall that X is the observation space. Consider an agent who chooses j in all periods, including zero. For each $\theta \in \Theta$, the probability that he observes an infinite history of one-period observations in $G \in \bigcup_{t=1}^{\infty} \mathcal{X}_t$ can be computed from the transition probabilities $\{\nu_t^\theta\}_{t\in\mathbb{Z}_+}$. Denote the set function obtained in this way by $\eta_j(\theta)$. By Kolmogorov's extension theorem, $\eta_j(\theta)$ admits an unique extension to $\mathcal{P}(X^\infty)$ that we also denote by $\eta_j(\theta)$. To finish this step, let $\eta_j(\pi)$ be the unique probability measure on $\mathcal{B}(\theta \times X^\infty)$ such that $\eta_j(\pi)(\Theta' \times G) = \sum_{\theta\in\Theta'} \pi(\theta)\eta_j(G|\theta)$ for all $\Theta' \subseteq \Theta$ and $G \in \mathcal{B}(X^\infty)$, where π is some given prior belief.

(B) For each $\theta \in \Theta$ and $t \in \mathbb{Z}_+$, let $\pi_{\theta,t}(\pi) = \mathrm{E}_{\eta_j(\pi)}[I_{\{\theta\}\times X^\infty}|\mathcal{G}_t]$, where $\mathcal{G}_0 = \{\emptyset, \Theta \times X^\infty\}$ and $\mathcal{G}_t = \{\emptyset, \Theta\} \times \mathcal{X}_t$ for $t \geq 1$. Then, $\pi_{\theta,0}(\pi) \equiv \pi(\theta)$ and $\pi_{\theta,t}$, with $t \geq 1$, is the period t (unconditional) posterior belief that the state of the world is θ for any agent with prior π who chooses j in every period. By Levy's theorem, see Theorem 3, p. 510, in Shiryaev (1996), $\pi_{\theta,t}(\pi)$ converges $\eta_j(\pi)$-almost surely to $\pi_{\theta,\infty}(\pi) = \mathrm{E}_{\eta_j(\pi)}[I_{\{\theta\}\times X^\infty}|\mathcal{G}_\infty]$, where $\mathcal{G}_\infty = \sigma(\bigcup_{t=0}^{\infty} \mathcal{G}_t)$.

(C) We know that if $\pi \in \Pi$ puts probability one on $\widehat{\theta}$, then $l_j(\pi) = r_j(\widehat{\theta}_j) - \max_{k\neq j} r_k(\widehat{\theta}_k)$, which is greater than zero by assumption. Consequently, there exists $\alpha \in (0,1)$ such that $l_j(\pi) > 0$ if $\pi(\widehat{\theta}) \geq 1 - 2\alpha$. Consider then the sequence $\{L_t(\pi)\}_{t\in\mathbb{Z}_+}$ with $L_t(\pi) = \pi_{\widehat{\theta},t}(\pi) - (1-\alpha)$, and let $L_\infty(\pi) = \pi_{\widehat{\theta},\infty}(\pi) - (1-\alpha)$. Define $\tau(\pi)$ to be such that $\tau(\pi) = \inf\{t \in \mathbb{Z}_+ : L_t(\pi) < 0\}$, with the convention that the infimum of an empty set is ∞. Notice that $\{\tau(\pi) = t\} \in \mathcal{G}_t$ for all $t \in \mathbb{Z}_+ \cup \{\infty\}$. In particular, $\tau(\pi)$ is a stopping time with respect to the filtration $\{\mathcal{G}_t\}_{t\in\mathbb{Z}_+}$. Now let $L_\tau(\pi)$ be the stopped random variable given by

$$L_\tau(\pi) = \begin{cases} L_t(\pi) & \text{if} \quad \tau = t \\ L_\infty(\pi) & \text{if} \quad \tau = \infty \end{cases}.$$

By Proposition A.1 in Banks and Sundaram (1992), $\mathrm{E}_{\eta_j(\pi)}[L_\tau(\pi)] = \mathrm{E}_{\eta_j(\pi)}[L_0(\pi)]$, from which we can conclude that $\mathrm{E}_{\eta_j(\pi)}[L_\tau(\pi)] = \pi(\widehat{\theta}) - (1-\alpha)$.

(D) Let $\Pi_j'(\widehat{\theta}) = \{\pi \in \Pi : \pi(\widehat{\theta}) \geq 1 - \alpha \text{ and } \pi(\theta) = 0 \text{ if } \theta_k \neq \widehat{\theta}_k \text{ for } k \neq j\}$. By definition, an agent with prior belief in $\Pi_j'(\widehat{\theta})$ knows the true type of $k \neq j$ and assigns high enough probability to the event that $\theta = \widehat{\theta}$. Notice that $\Pi_j'(\widehat{\theta}) \subset \Pi^d$. By item C and the definition of $\Pi_j'(\widehat{\theta})$, $\mathrm{E}_{\eta_j(\pi)}[L_\tau(\pi)] \geq 0$ if $\pi \in \Pi_j'(\widehat{\theta})$. Hence, $\pi \in \Pi_j'(\widehat{\theta})$ implies that $\eta_j(\pi)(\{\tau(\pi) = \infty\}) > 0$. In particular, there exists $G \in \mathcal{B}(X^\infty)$ and $\widetilde{\theta} \in \Theta$ such that $\pi(\widetilde{\theta}) > 0$, $\eta_j(G|\widetilde{\theta}) > 0$, and $\tau(\pi) = \infty$ in $\{\widetilde{\theta}\} \times G$. This last result follows from the definition of $\eta_j(\pi)$. Theorem 1 then implies that the following is true for a measure one of agents: if they have a prior belief in $\Pi_j'(\widehat{\theta})$ and choose j in period zero, then they choose j in every $t \geq 1$ when the state of the world is $\widetilde{\theta}$. Suppose, by contradiction, that $\widetilde{\theta} \neq \widehat{\theta}$. By Lemma 3, an agent who chooses j infinitely many times learns its true type with probability one, regardless of the meetings in society. Since an agent with prior π in $\Pi'_j(\widehat{\theta})$ already knows the true type of $k \neq j$, $\pi_{\widehat{\theta},t}(\pi)$ converges to zero $\eta_j(\pi)$-almost surely in $\{\widetilde{\theta}\} \times G$. This, however, contradicts the fact that $\pi_{\widehat{\theta},t}(\pi) \geq 1 - \alpha$ in this set.

(E) This completes the first part of the proof. Let $M_t = \pi_{\widehat{\theta},t} - (1 - 2\alpha)$ and define $\tau^+(\pi)$ to be such that $\tau^+(\pi) = \inf\{t \in \mathbb{Z}_+ : M_t < 0\}$. Notice that $\tau^+(\pi) = \infty$ if $\tau(\pi) = \infty$. Now let $\Pi_j(\widehat{\theta})$ be the subset of $\{\pi \in \Pi^d : \pi(\widehat{\theta}) > 0\}$ with the property that if an agent has a prior π in $\Pi_j(\widehat{\theta})$, then there is $G \in \mathcal{B}(X^\infty)$ with $\eta_j(G|\widehat{\theta}) > 0$ such that $\tau^+(\pi) = \infty$ in $\{\widehat{\theta}\} \times G$. This set is non-empty by the above reasoning. In what follows we use a continuity argument to show that $\Pi_j(\widehat{\theta})$ is an open subset of Π^d. The full support assumption together with Theorem 1 and the assumption that all agents randomize in period zero then imply that there is a positive measure of agents who choose j in every $t \geq 1$.

(F) Fix $\pi \in \Pi_j(\widehat{\theta})$ and let $\{\pi_n\} \in \Pi^d$ be such that $\lim \pi_n = \pi$. Assume, without loss, that $\pi_n(\widehat{\theta}) > 0$ for all $n \in \mathbb{N}$. Moreover, let $G \in \mathcal{B}(X^\infty)$ be such that $\eta_j(G|\widehat{\theta}) > 0$ and $\tau(\pi) = \infty$ in $\{\widehat{\theta}\} \times G$. Since $\pi_{\widehat{\theta},t}(\pi) \geq 1-\alpha$ in $\{\widehat{\theta}\} \times G$ for all $t \in \mathbb{N}$, $\pi_{\widehat{\theta},\infty}(\pi) \geq 1-\alpha$ $\eta_j(\pi)$-almost surely in $\{\widehat{\theta}\} \times G$. Now observe that $\eta_j(\pi_n)$ converges to $\eta_j(\pi)$ in (variation) norm. Corollary 3.2 in Crimaldi and Pratelli (2005) then implies that $\pi_{\widehat{\theta},t}(\pi_n)$ converges to $\pi_{\widehat{\theta},\infty}(\pi)$ in $\eta_j(\pi)$-measure as $t, n \to \infty$. Hence, there exists $G' \subseteq G$ with $\eta_j(G'|\widehat{\theta}) > 0$ and $t_0, n_0 \in \mathbb{N}$ such that $\pi_{\widehat{\theta},t}(\pi_n) > 1-2\alpha$ on $\{\widehat{\theta}\} \times G'$ if $t \geq t_0$ and $n \geq n_0$. Also notice that for each $t \in \mathbb{N}$, $\pi_{\widehat{\theta},t}(\pi_n)$ converges pointwise to $\pi_{\widehat{\theta},t}(\pi)$. Therefore, by Egoroff's theorem, there exist $n_1 \in \mathbb{N}$ and $G'' \subseteq G'$ with $\eta_j(G''|\widehat{\theta}) > 0$ such that $\pi_{\widehat{\theta},t}(\pi_n) > 1-2\alpha$ in $\{\widehat{\theta}\} \times G''$ for all $t \in \{1, \ldots, t_0 - 1\}$ if $n \geq n_1$. We can then conclude that for all $t \in \mathbb{N}$, $\pi_{\widehat{\theta},t}(\pi_n) > 1 - 2\alpha$ in $\{\widehat{\theta}\} \times G''$ if $n \geq \max\{n_0, n_1\}$, the desired result. □

Lemma 13. *Suppose that if $\theta \in \Theta$ is such that $z_1(\theta_1) \leq z_2(\theta_2) = \overline{z}$, then $m_t(1, \theta, F^*)$ converges to zero in any perfect bayesian equilibrium F^*. Let the state of the world be $\overline{\theta}$ with $z_2(\overline{\theta}_2) = \overline{z}$ and suppose that F^* is a perfect bayesian equilibrium where $m_t(1, \overline{\theta}, F^*)$ converges to some $\alpha \in (0, 1)$. Then, with probability one, an agent with a non-dogmatic prior who chooses 2 infinitely many times learns that z_1 is greater than $\overline{z}$.*

Proof: Consider an agent who chooses 2 infinitely many times. We can restrict attention to an agent who follows an optimal strategy.[13] By Corollary 1, he can only choose 1 a finite number of times. Hence, without loss, we can assume that he chooses 2 in all periods. Let $\chi_\theta \in \mathcal{P}(X^\infty)$ be such that if $D \in \mathcal{B}(X^\infty)$, then $\chi_\theta(D)$ is the probability that this agent experiences an infinite history of one-period observations that belongs to D. Denote a typical element of X^∞ by $x = \{x_t\}$, where $x_t = (x_{1t}, x_{2t}) \in Y \times A$ is the period t observation. Now let $b_t : X^\infty \to \{0, 1\}$ be such that $b_t(x) = 1$ if $x_{2t} = 2$ and $b_t(x) = 0$ otherwise. By construction, $\mathrm{E}_{\chi_\theta}[b_t] = m_t(2, \theta, F^*)$. Moreover, let $y_t : X^\infty \to \mathbb{R}$ be such that $y_t(x) = x_{1t}$. Notice that the random variables $\{b_t\}_{t\in\mathbb{Z}_+}$ and $\{y_t\}_{t\in\mathbb{Z}_+}$ are independent. In what follows we omit the dependence of the matching probabilities on F^*

Suppose the agent under consideration has a non-dogmatic prior π_0 and assume, without loss, that $m_t(j, \overline{\theta}) > 0$ for all j and t. Let $f_t(b_t, \theta) = m_t(2, \theta)^{b_t} m_t(1, \theta)^{1-b_t}$. If $\pi(\,\cdot\,|\pi_0, y_0, b_0, \ldots, y_t, b_t)$ is his updated belief after he observes $(y_0, b_0, \ldots, y_t, b_t)$ in the first $t + 1$ periods, then

$$\pi(\theta|\pi_0, y_0, b_0, \ldots, y_t, b_t) = \frac{\Pi_{n=0}^t g_2(y_n, \theta_2) f_n(b_n, \theta)\pi_0(\theta)}{\sum_{\theta'\in\Theta} \Pi_{n=0}^t g_2(y_n, \theta_2) f_n(b_n, \theta')\pi_0(\theta')}.$$

Observe that $\pi(\theta|\pi_0, y_0, b_0, \ldots, y_t, b_t) \leq \{1 + L_t(\theta|\pi_0, y_0, b_0, \ldots, y_t, b_t)\}^{-1}$, where

$$L_t(\theta|\pi_0, b_0, y_0, \cdots, y_t, b_t) = \frac{\Pi_{n=0}^t g_2(y_n, \overline{\theta}_2) f_n(b_n, \overline{\theta})\pi_0(\overline{\theta})}{\Pi_{n=0}^t g_2(y_n, \theta_2) f_n(b_n, \theta)\pi_0(\theta)}.$$

Simple algebra shows that $\ln L_t = \sum_{n=0}^t b_n \gamma_n(2) + \sum_{n=0}^t (1 - b_n)\gamma_n(1) + \sum_{n=0}^t \xi_n + \delta$, where

$$\gamma_n(j) = \ln\left\{\frac{m_n(j, \overline{\theta})}{m_n(j, \theta)}\right\} \quad , \quad \xi_n = \ln\left\{\frac{g_2(y_n, \overline{\theta}_2)}{g_2(y_n, \theta_2)}\right\} \quad , \text{ and } \delta = \ln\left\{\frac{\pi_0(\overline{\theta})}{\pi_0(\theta)}\right\}.$$

Let $\Theta' = \{\theta \in \Theta : z_2(\theta_2) = \overline{z} \text{ and } z_1(\theta_1) > \overline{z}\}$, $\Theta'' = \{\theta \in \Theta : z_2(\theta) = \overline{z} \text{ and } z_1(\theta_1) \leq \overline{z}\}$, and $\widehat{\Theta} = \Theta' \cup \Theta''$. Notice that $\overline{\theta} \in \Theta'$. From now on, all almost sure statements are with respect to $\chi_{\overline{\theta}}$. There are two cases to consider. If $\theta_2 \neq \overline{\theta}_2$, Lemma 3 implies that $\pi(\theta|y_0, b_0, \ldots)$ converges to zero almost surely. Suppose then that $\theta_2 = \overline{\theta}_2$ and $\theta \in \Theta''$. Notice that $\xi_n \equiv 0$ in this case. Since $\mathrm{Var}_{\chi_\theta}(b_t) \leq 1/4$ and $\sum_{n=0}^t (1 + t)^{-1} m_n(2, \overline{\theta}) \to 1 - \alpha > 0$, Kolmogorov's SLLN implies that

$$\frac{1}{t+1}\sum_{n=0}^t b_n = \frac{1}{t+1}\sum_{n=0}^t (b_n - m_n(2, \overline{\theta})) + \frac{1}{t+1}\sum_{n=0}^t m_n(2, \overline{\theta}) \to 1 - \alpha > 0$$

almost surely. Now observe that $m_t(1, \overline{\theta}) \to \alpha > 0$ and $m_t(1, \theta) \to 0$ by assumption. Hence, $\{\gamma_t(2)\}$ is bounded while $\gamma_t(1) \to \infty$. This implies that $(1 + t)^{-1} \ln L_t \to \infty$ almost surely, and so $L_t \to \infty$ almost surely as well. Therefore, $\pi(\theta|\pi_0, y_0, b_0, \ldots) \to 0$ almost surely, the desired result. □

[13] This lemma is true without this restriction, but the proof is longer.