Learning in Society*

Braz Camargo†

March 2009

Abstract

In the canonical learning model, the multi–armed bandit, 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 both 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 players where each one of them faces a multi–armed bandit and in each period a player observes the action choice of another random player. We obtain two results. First, that social conformity always happens in the long–run. Second, that if initial beliefs are sufficiently heterogeneous, then the fraction of players who choose the superior arm always converges to one.

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 Andrew Postlewaite, Jan Eeckhout, and various conference and seminar participants for their comments and suggestions.

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

1 Introduction

In an experimentation problem, actions not only provide rewards, but also information that is relevant for future decisions. There is then a tradeoff between the maximization of current rewards and the acquisition of information in such problems. When a decision maker can only learn from his own experience, this tradeoff implies that an optimal experimentation strategy sometimes leads the best alternative to be dropped altogether, the so–called "Rothschild effect".[1,2] Many situations of interest, however, have individuals learning both from their own experience and the experience of others. Examples include consumers learning about product quality and doctors learning about different treatments for the same disease. In this paper, we investigate whether social learning can overturn the Rothschild effect.

We study a discrete time economy populated by a continuum of infinitely lived and anonymous players where each one of them faces a multi–armed bandit with independent arms. The players are homogeneous in the sense that the unknown stochastic payoffs to each of the available actions are the same for all of them. The players, however, may have heterogeneous prior beliefs about the true stochastic payoffs. We assume that every so often a player observes the action choices of a finite random sample from the population.[3] More precisely, in every period a player observes the current action choice of an individual selected randomly from the population.[4] We refer to these observations as the observations in society.

As in a standard multi–armed bandit, the flow payoff of a player depends only on his action choice. The action choices of the other individuals in the population affect continuation payoffs, though. Indeed, in each period, the likelihood that a player observes someone

[1]Rothschild (1974) is the first paper in Economics to draw attention to this fact. The term Rothschild effect was introduced in Kihlstrom et al. (1984).

[2]Banks and Sundaram (1992) and Brezzi and Lai (2000) show this result for multi–armed bandits with independent arms. Easley and Kiefer (1988) shows the same result for a different class of experimentation problems that includes multi–armed bandits with correlated arms as a special case.

[3]In the case of a consumer learning about product quality this happens, for example, when the consumer goes to a store and learns about the purchase decisions of other costumers, either by observing them directly or by asking the staff about them.

[4]It is easy to adapt our analysis to the case where the size of the random sample an individual observes is greater than one or is random (with finite maximum sample size).

choosing a particular action is determined by the behavior of the other players. Since the latter depends (indirectly) on the true payoffs of the action choices, so do the observation probabilities. Hence, the observations in society reveal payoff relevant information. Thus, there is both informational and strategic interaction between the players. Formally, our setting is a game of strategic experimentation with non–atomic players.

The first main result of the paper is that in equilibrium there is always social conformity in the long–run. In other words, regardless of the true stochastic payoff of each action and the distribution of initial beliefs in the population, almost all players end up choosing the same action in the long–run in every equilibrium of the game. We say that social learning is complete if the fraction of individuals who choose the best action always converges to one. The second main result of the paper is a condition on the distribution of initial beliefs that if satisfied implies that social learning is complete in every equilibrium. Roughly speaking, this condition requires that prior beliefs be heterogeneous enough.

It is straightforward to adapt our analysis to the case where the observations in society include not only actions, but also their outcomes. The reason we assume that only actions can be observed is twofold. First, to emphasize that this alone provides information about payoffs. Second, to show that just observing action choices can be enough to overturn the Rothschild effect. Besides, the assumption that outcomes are observable is not always adequate in many settings that involve social learning.

Most of the literature on social learning only allows for informational interactions among individuals. For instance, Ellison and Fudenberg (1993,1995), Juang (2001), and Smallwood and Conlisk (1979) consider models of social learning where agents follow simple rules of behavior. Bala and Goyal (1998) and Gale and Kariv (2003) study social learning in networks with myopic agents. The literature on informational cascades, see Banerjee (1992), Bikhchandani et al. (1992), and Smith and Sorensen (2000), consider models where agents make irreversible decisions in an exogenously given order.[5]

[5] Caplin and Leahy (1994) and Chamley and Gale (1994) consider models of herd behavior where the timing of decisions is endogenous.

The paper on strategic experimentation that is most closely related to ours is Aoyagi (1998). He analyzes a game with a finite number of players where action choices, but not their outcomes, are public and shows that in every Nash equilibrium all players eventually settle on the same action (arm), not necessarily the best one. Rosenberg et al. (2006) study social conformity in a large class of finite–player games of strategic experimentation. Their main result is that social conformity need not obtain in the long–run.[6]

In the next section we discuss the main ideas of the paper in the context of a simple example. We present the model in Section 3. In Section 4 we define strategy profiles and equilibria. In Section 5 we study the individual learning problem, the experimentation problem the players face when they take the behavior of the other players as given. We prove the result on social conformity in Section 6 and the result on complete social learning in Section 7. In Section 8 we discuss some issues related to the model. We establish the existence of an equilibrium in Section 9 and conclude in Section 10. Appendices A to D contain omitted proofs and details. Appendix E sketches an extension of the model.

2 Example

We illustrate the main ideas of the paper by considering informally the case where there are two actions, a_1 and a_2, the outcome of an action is either a success or a failure, and the probability of success for each action is unknown and can assume one of two values.

Let $p_k \in \{p_k^L, p_k^H\}$, with $p_k^L < p_k^H$, be the probability that a_k yields a success. We assume that $0 < p_1^L < p_2^L < p_1^H < p_2^H < 1$. A state of the world is a value of the pair (p_1, p_2). We denote the set of states of the world by $\Theta = \{LL, LH, HL, HH\}$, where $\theta = \alpha\beta$ is the state where $p_1 = p_1^\alpha$ and $p_2 = p_2^\beta$. Prior beliefs may be heterogeneous. We assume that a mass one of players has a full support prior.

[6]See also Bolton and Harris (1999), Keller et al. (2005), and Rosenberg et al. (2007). All three papers consider finite–player games of strategic experimentation. The first two assume that outcomes are public, so that there is no asymmetric information. The focus of these papers, however, is not on the characterization of long–run equilibrium behavior.

Index time by $t \geq 1$. To each strategy profile F there is associated a sequence $\{m_t\}$, where m_t is a map from $\{a_1, a_2\} \times \Theta$ into $[0, 1]$, such that if F is under play, then $m_t(a_k, \theta)$ is the mass of players who choose a_k in period t when the state of the world is θ. Since players are non–atomic, $m_t(a_k, \theta)$ is also the chance that a player's period–t observation in society is a_k when the state of the world is θ. The anonymity of the players implies that the probabilities $m_t(a_k, \theta)$ contain all the information provided by the observations in society.

We begin by showing how the observations in society lead to the long–run convergence of behavior in any equilibrium of the game. Notice that this result would be straightforward if the observations in society included not only action choices, but also their outcomes. Indeed, suppose that outcomes are observable and consider a strategy profile where both actions are chosen by a positive fraction of the population in the long–run. In this case, every player observes the outcome of both actions an infinite number of times, and so learns which of the two is better (see Footnote 7). But then, there is a positive mass of players who are not behaving optimally, and so this cannot be an equilibrium.

Suppose first that $p_1 = p_1^L$, i.e., either $\theta = LL$ or $\theta = LH$. In this case, the mass of players who choose a_1 should converge to zero in the long–run. Otherwise, a positive mass of players choose a_1 infinitely many times and learn that $p_1 = p_L$.[7] Such players are not behaving optimally, though, since any player who is convinced enough that $p_1 = p_1^L$ should choose a_2 instead of a_1.

Suppose now that $\theta = HL$ and assume that the mass of players who choose each action is bounded away from zero. This implies, in particular, that a positive mass of players choose a_2 infinitely often and learn that $p_2 = p_2^L$. However, these players also learn that $\theta \notin \{LL, LH\}$, since they observe that there is no convergence of behavior in the long–run. Thus, they learn that $p_1 = p_1^H$, which means that they are not behaving optimally, as any player who is convinced enough that $p_1 = p_1^H$ and $p_2 = p_2^L$ should choose a_1 instead of a_2.

[7] By the Strong Law of Large Numbers, a player with a full support prior who chooses a_1 infinitely often learns the true value of p_1 with probability one if he only takes into account the outcome of his action choices. This individual, however, also learns about p_1 from his observations in society. Lemma 1 in Aoyagi (1998) shows that this extra information cannot overturn the information he obtains from choosing a_1 infinitely many times. Hence, this individual learns the true value of p_1 with probability one.

Finally, suppose that $\theta = HH$ and once more assume that there is no convergence of behavior in the long–run. Then, a mass one of players learn that $\theta = HH$, since players know that behavior converges in the long–run in every other state of the world. This implies that the mass of players who choose a_1 should converge to one in the long–run, a contradiction. Thus, in equilibrium there is always long–run social conformity.

We now consider social learning. First, notice that social learning need not be complete. Indeed, suppose that all players have a prior assigning a high enough probability to $p_1 = p_1^L$ that even if they believe that $p_2 = p_2^L$, it is still optimal to choose a_2. In this case, all players choose a_2 in the first period. Since the observations in society are uninformative in period 1, all players still choose a_2 in period 2, as their beliefs about a_1 remain the same. But then, the observations in society in period 2 reveal no information about a_1, and so all players choose a_2 in period 3 as well. Continuing with this reasoning, it is easy to see that the mass of players who choose a_2 is always one no matter the state of the world.

Let $a(\theta)$ be the best action when the state of the world is θ. Since there is convergence of behavior in the long–run, a sufficient condition for social learning to be complete is that for each θ, the mass of players who choose $a(\theta)$ is bounded away from zero. Thus, social learning is complete if for all θ there is a positive mass of players for which the probability that they choose $a(\theta)$ infinitely many times is positive. Let us see how this is possible.

Fix a state of the world θ and suppose that there is a positive mass of players with a prior π such that $\pi(\theta)$, the probability that π assigns to θ, is so large that even if they were to learn the true state of the world by choosing the other action, they would still prefer to choose $a(\theta)$.[8] Take such a player. If the true state of the world is θ, then it cannot be that

[8] For example, if we fix the payoff from a success to 1 and the payoff from a failure to 0, the lifetime payoff to a player with belief π and discount factor δ who learns the value of θ if he chooses a_2 is

$$\overline{v}_2 = [\pi(LL)+\pi(HL)]p_2^L+[\pi(LH)+\pi(HH)]p_2^H+\delta(1-\delta)^{-1}\left\{\pi(LL)p_2^L + \pi(HL)p_1^H + [\pi(LH) + \pi(HH)]p_2^H\right\}.$$

On the other hand, the player's lifetime payoff if he always chooses a_1 is

$$\underline{v}_1 = (1-\delta)^{-1}\{[\pi(LL) + \pi(LH)]p_1^L + [\pi(HL) + \pi(HH)]p_1^H\}.$$

Now observe that $\underline{v}_1 > \overline{v}_2$ if $\pi(HL)$ is close to one. Since $\overline{v}_2$ is an upper bound for the lifetime payoff to the player if he chooses a_2, it is then always optimal for him to choose a_1 if $\pi(HL)$ is high enough for $\underline{v}_1 > \overline{v}_2$.

the player's belief that the state of the world is θ falls on average. But then, there exists a positive measure set of sample paths such that along any path in this set the player's belief that the state of the world is θ is not smaller than $\pi(\theta)$ when t is large enough.[9] This implies that when the state of the world is θ there is a positive probability that the player only chooses $a(\theta)$ when t is large enough. Consequently, there exists a positive fraction of players who choose $a(\theta)$ infinitely often when the state of the world is θ.

Hence, social learning is complete if for each θ there exists a positive mass of players with prior π such that $\pi(\theta)$ is high enough. The restriction on $\pi(\theta)$ described above, although intuitive, depends on the players' discount factor (see Footnote 8) and is quite demanding when players are patient in the sense that it requires some players to have extreme beliefs. In Section 7 we show that a much weaker, discount–factor–free, condition suffices for social learning to be complete.

3 The Model

Time is discrete and indexed by $t \geq 1$. The economy is populated by a continuum of mass one of anonymous and infinitely lived players that we identify with $([0,1], \mathcal{I}, \lambda)$, where $\mathcal{I}$ is the set of Lebesgue measurable subsets of $[0,1]$ and λ is the Lebesgue measure. We denote players by either i or j. All players discount future payoffs at the same rate $\delta \in [0,1)$.

There are $K \geq 2$ possible actions, that we label a_1 to a_K. Let $A = \{a_1, \ldots, a_K\}$ denote the action set. Actions have stochastic outcomes, which are determined independently for each player. The set of possible outcomes is Y and we take it to be finite. All the results in the paper, except the result on the existence of an equilibrium, can be extended to the case where Y is a complete separable metric space. The payoff to a player who chooses action a and observes the outcome y is $u(a, y)$.

[9]The proof of this result has two parts. First, when the state of the world is θ, the likelihood ratio $\ell(\theta', \theta)$ of any state $\theta' \neq \theta$ to θ is a supermartingale, and so converges almost surely to a random variable with expectation not greater than one. Then, by Egoroff's theorem, there exists a set A of sample paths that has positive measure when the state of the world is θ with the property that in any element of this set, $\ell(\theta', \theta) \leq 1$ for all $\theta' \neq \theta$ when t is large enough. This leads to the desired result.

The timing in a period is as follows. First, players choose actions and observe their outcomes. Then, each player observes the current action choice of another random player.

The outcome of each action a depends on a parameter θ_a that is the same for all players. The set Θ_a of possible values of θ_a is finite. We refer to the value of θ_a as the (true) type of action a and to $\Theta = \times_{a \in A} \Theta_a$, with typical element θ, as the set of states of the world. The value of θ is unknown to the players. To each pair (a, θ_a) is associated a probability distribution $\eta_a(\theta_a)$ on Y that governs the realization of outcomes when a is chosen and its type is θ_a. We assume that the maps $\theta_a \mapsto \eta_a(\theta_a)$ are one–to–one, for otherwise we can redefine Θ_a to exclude any redundant types.

Let $g : Y \times A \times \Theta \to \mathbb{R}$ be such that $g(y|a, \theta)$ is the probability of the outcome y when action a is chosen and the state of the world is θ. The function g is the *likelihood function* for the multi–armed bandit problem at hand. We assume that for each $y \in Y$ there exist $a \in A$ and $\theta \in \Theta$ such that $g(y|a, \theta) > 0$, for otherwise we can redefine Y to exclude the outcomes that can never be observed.

Let $r(a, \theta) = \sum_{y \in Y} u(a, y) g(y|a, \theta)$ be the expected reward from action a when the state of the world is θ. Notice that $r(a, \theta)$ only depends on the type of a; that is, if θ and θ' are two states of the world such that $\theta_a = \theta'_a$, then $r(a, \theta) = r(a, \theta')$. We make the following three assumptions about the expected rewards.

Assumption A1. *For every $\theta \in \Theta$, $r(a, \theta) \neq r(a', \theta)$ for all $a \neq a'$ in A.*

Assumption A2. *For each $a \in A$, $r(a, \theta) \neq r(a, \theta')$ for all $\theta, \theta' \in \Theta$ such that $\theta_a \neq \theta'_a$.*

Assumption A3. *For each $a \in A$, there exists $\theta \in \Theta$ with $r(a, \theta) = \text{argmax}_{\widehat{a} \in A} r(\widehat{a}, \theta)$.*

Assumption A1 implies that for every state of the world the expected flow payoffs from any two different actions are not the same. Assumption A2 implies that for every action a, the expected rewards of any of its two types are different. We make this assumption for simplicity; the results of the paper hold without it. Assumption A3 is without loss, for if $a' \in A$ is such that for all $\theta \in \Theta$ there exists $a \in A$ with $r(a, \theta) > r(a', \theta)$, then no player who behaves optimally ever chooses a'.

Let $\Pi = \Delta(\Theta)$ be the set of beliefs about the state of the world, beliefs for short. Denote a typical element of Π by π and the probability that π assigns to θ by $\pi(\theta)$. Now let Π^d be the set of uncorrelated beliefs. Almost every player begins with a full support (non–dogmatic) prior belief π_1 in Π^d. Priors can be heterogeneous. More precisely, there exists a Lebesgue measurable function $\Phi : I \to \Pi$ such that $\Phi(i)$ is the prior belief of player i.

4 Strategy Profiles and Equilibria

In this section we define strategy profiles and equilibria.

1. Strategy Profiles.

We begin with some notation. For any metric space S, we denote the set of all bounded and Borel functions from S into $\mathbb{R}^n$ endowed with the sup norm by $\mathrm{B}_b(S, \mathbb{R}^n)$ ($\mathrm{B}_b(S)$ if $n = 1$) and the closed ball of radius M in $\mathrm{B}_b(S, \mathbb{R}^n)$ by $\mathrm{B}_b^M(S, \mathbb{R}^n)$ ($\mathrm{B}_b^M(S)$ if $n = 1$). When S is finite, we identify $\mathrm{B}_b(S, \mathbb{R}^n)$ with $\mathbb{R}^{n|S|}$. For any complete finite measure space (Ω, Σ, μ), we denote the set of equivalence classes of integrable functions from Ω into $\mathbb{R}^n$ endowed with the norm $||f||_1 = \int ||f(\omega)||\mu(d\omega)$ by $\mathrm{L}_1(\mu, \mathbb{R}^n)$ ($\mathrm{L}_1(\mu)$ if $n = 1$).

A player's experience in a given period is a list (a, y, a'), where a is his action choice, y is the outcome of a, and a' is his observation in society. Let $X = Y \times A$ be the set of possible outcome–observation pairs. The set of period–t histories is $H_t = (A \times X)^{t-1}$ and the set of infinite histories is $H_\infty = (A \times X)^\infty$. Denote a typical element of X by $x = (y, a')$ and typical element of H_t by h^t. A behavior strategy is a sequence $\sigma = \{\sigma_t\}$, where $\sigma_t : H_t \to \Delta_{K-1}$ is the map that describes how period–t histories are mapped into mixed actions; the set Δ_{K-1} is the unit simplex in $\mathbb{R}^K$. By convention, the kth coordinate of an element of Δ_{K-1} denotes the probability that a_k is chosen.

Let Σ_t be the set of functions in $\Gamma_t = \mathrm{B}_b(H_t, \mathbb{R}^K)$ that have range in Δ_{K-1}. The set of behavior strategies is then $\Sigma = \times_{t=1}^{\infty} \Sigma_t$. Now, denote a typical element of $\mathrm{L}_1(\lambda, \Gamma_t)$ by F_t and define, in an abuse of notation, $\mathrm{L}_1(\lambda, \Sigma_t)$ to be set of F_t in $\mathrm{L}_1(\lambda, \Gamma_t)$ such that $F_t(i) \in \Sigma_t$ for λ–almost all $i \in [0, 1]$.

Definition: *The set of strategy profiles is the set* $\mathfrak{S} = \times_{t=1}^{\infty} \mathrm{L}_1(\lambda, \Sigma_t)$.[10]

Let $\mathfrak{M}$ be the set of all (infinite) sequences of maps from $A \times \Theta$ into $[0,1]$. To every strategy profile F there is associated an element $m = \{m_t\}$ of this set, where $m_t(a,\theta)$ is the mass of players who choose action a in period t when the state of the world is θ if F is the strategy profile under play. We refer to m_t as the period–t *observation likelihood* and denote the map that takes a strategy profile F into its corresponding sequence of observation likelihoods by M. See Appendix A for the construction of this map.

2. Equilibria.

The difference between the individual learning problem and a standard multi–armed bandit is that in the former the players obtain information about the state of the world not only from the outcomes of their action choices, but also from their observations in society. Because players are anonymous, the information from the latter is described by a sequence of observation likelihoods. In what follows we use $\mathrm{ILP}(\pi_1, m)$ to denote the individual learning problem of a player with prior π_1 when the sequence of observation likelihoods is m.

Let $\mu_\theta(\sigma, m)$ be the probability distribution on H_∞ induced by the behavior strategy σ and the sequence of observation likelihoods m when the state of the world is θ and $\mu(\sigma, \pi, m)$ be the probability distribution on $\Theta \times H_\infty$ induced by the prior π and the probability distributions $\mu_\theta(\sigma, m)$. Denote the period–t action choice and outcome by a_t and y_t, respectively, and let $R = \sum_{t=1}^{\infty} \delta^{t-1} u(a_t, y_t)$. The objective in $\mathrm{ILP}(\pi_1, m)$ is to find $\sigma^* \in \Sigma$ such that

$$\mathrm{E}_{\mu(\sigma^*, \pi_1, m)}[R] = \sum_{\theta \in \Theta} \pi_1(\theta) \mathrm{E}_{\mu_\theta(\sigma^*, m)}[R] = \sup_{\sigma \in \Sigma} \mathrm{E}_{\mu(\sigma, \pi_1, m)}[R], \tag{1}$$

where $\mathrm{E}_\mu[R]$ denotes the expectation of R with respect to μ. We refer to a behavior strategy σ^* that satisfies (1) as an optimal experimentation strategy for $\mathrm{ILP}(\pi_1, m)$.[11]

[10] As is standard in the literature on non–atomic games, the set of strategy profiles is not the cartesian product of the set of behavior strategies for each player. The measurability restriction in the definition of $\mathfrak{S}$ is necessary for aggregate behavior to be well–defined.

[11] Formally, $\mathrm{ILP}(\pi_1, m)$ is equivalent to a non–stationary multi–armed bandit with correlated arms where the outcome space is X and the period–t likelihood function is $\widetilde{g}_t(x|a, \theta) = g(y|a, \theta) m_t(a', \theta)$. Easley and Kiefer (1988) study long–run behavior in a large class of infinite–horizon experimentation problems that

The equilibrium notion we use generalizes to our environment the notion of a Nash equilibrium for non–atomic games introduced by Schmeidler (1973).

Definition: *An equilibrium is a pair* (m^*, F^*) *such that* $F^*(i)$ *is an optimal experimentation strategy for* $\text{ILP}(\Phi(i), m^*)$ *for* λ*–almost all* $i \in [0,1]$ *and* $m^* = \mathsf{M}(F^*)$*.*[12]

5 The Individual Learning Problem

In this section we study the individual learning problem. We first establish a partial characterization result for optimal experimentation strategies and then consider individual behavior in the long–run. In what follows, we let π_t be a player's period–t belief and $\pi_t(h^t)$ be this belief after the history $h^t \in H_t$. Notice that π_t depends on the player's prior belief and on the sequence m of observation likelihoods.

1. Characterization

Let $v(a,\pi) = \sum_{\theta\in\Theta} r(a,\theta)\pi(\theta)$ be the expected reward from action a when the belief is π and $r(\theta) = \max_{a\in A} r(a,\theta)$ be the highest flow payoff possible when the state of the world is θ. Now let $D_a : \Pi \to \mathbb{R}$ be given by

$$D_a(\pi) = (1-\delta)^{-1}v(a,\pi) - \max_{a'\neq a}\left[v(a',\pi) + \delta(1-\delta)^{-1}\sum_{\theta\in\Theta} r(\theta)\pi(\theta)\right].$$

Notice that if π is such that $D_a(\pi) > 0$, then $v(a,\pi) > v(a',\pi)$ for all $a' \neq a$. This, in turn, implies that $D_{a'}(\pi) < 0$ for all $a' \neq a$. The proof of the next result is in Appendix B.

Lemma 1. *Let* σ^* *be an optimal experimentation strategy for* $\text{ILP}(\pi_1, m)$*. If* h^t *has positive probability under* $\mu(\sigma^*, \pi_1, m)$ *and* $D_a(\pi_t(h^t)) > 0$*, then* $\sigma_t^*(h^t)$ *assigns probability one to* a*.*

The strength of Lemma 1 lies on the fact that the functions D_a do not depend on the sequence m of observation likelihoods. Hence, this result places restrictions on the behavior

includes multi–armed bandits with correlated arms as a special case. A key difference between our setting and theirs is that the likelihood functions in the individual learning problem are endogenous.

[12]Notice that players need not know the distribution of prior beliefs in the population; they only need to correctly anticipate the sequence m of observation likelihoods.

of the players that hold uniformly for all equilibria of the game. We make use of this result in our discussion of complete social learning in Section 7.

To understand the meaning of the condition $D_a(\pi) > 0$, consider the hypothetical situation in which a player learns the true value of θ if he chooses any action other than a. If his current belief is π, then his expected lifetime payoff from choosing $a' \neq a$ is given by $v(a', \pi) + \delta(1-\delta)^{-1} \sum_{\theta \in \Theta} r(\theta)\pi(\theta)$. Since, in reality, learning about θ does not happen immediately (if it ever happens), the above payoff is an upper bound for the player's expected lifetime payoff when he chooses a'. Likewise, $(1-\delta)^{-1} v(a, \pi)$ is a lower bound for the player's expected lifetime payoff when he chooses a. He obtains this payoff when he settles down on a. Thus, it is optimal for a player to choose a if his belief π is such that $D_a(\pi) > 0$.

Notice that the above intuition holds regardless of how many action choices a player observes in each period. So, Lemma 1 is still true in the more general setting where in every period each player observes the action choices of a random sample from the population, where the size of the sample can be any (even infinite). We return to this point at the end of Section 7.

2. Long–Run Behavior

Consider a player with a full support prior who chooses an action a an infinite number of times. Standard results show that he learns the true type of a with probability one if he only observes the outcomes of his action choices.[13] It turns out that this is also true in the presence of the observations in society. The proof of this result is a straightforward modification of the proof of Lemma 1 in Aoyagi (1998).

Lemma 2. *A player with a full support prior who chooses an action a infinitely often learns its true type with probability one.*

Let $v_*(\pi) = \max_a v(a, \pi)$ be the expected reward to a player who chooses a myopically optimal action when his belief is π and $N(\varepsilon) = |\{t \geq 1 : v(a_t, \pi_t) \leq v^*(\pi_t) - \varepsilon\}|$, with $\varepsilon > 0$,

[13] See, for instance, Section 10.5 in DeGroot (1970). The assumption of a non–dogmatic prior is sufficient, but not necessary, for this result. It is enough that the player's prior belief assigns positive probability to the true state of the world.

be the number of periods a player chooses an action that is myopically suboptimal by at least ε. Theorem 2.1 in Rosenberg et al. (2006) shows that if σ^* is an optimal experimentation strategy for ILP(π_1, m), then $\mathrm{E}_{\mu(\sigma^*,\pi_1,m)}[N(\varepsilon)] < \infty$. The intuition for this result is roughly as follows. A player is willing to sacrifice his current reward by ε only if he expects his payoff in the subsequent periods to increase, on average, by $(1-\delta)\varepsilon/\delta$. Because payoffs are bounded, it then must be that $\mathrm{E}_{\mu(\sigma^*,\pi_1,m)}[N(\varepsilon)](1-\delta)\varepsilon/\delta < \infty$, which implies that $\mathrm{E}_{\mu(\sigma^*,\pi_1,m)}[N(\varepsilon)]$ is of the order of $\delta/\varepsilon(1-\delta)$, and so finite.

Since the number of actions is finite, a straightforward consequence of this last result is that in the long–run, the probability that a player who follows an optimal experimentation strategies chooses a myopically suboptimal action is zero. More precisely, let A_∞ be the set of limit actions, i.e., the actions that are chosen infinitely many times, and denote an element of this set by a_∞. Then the following result, which is Theorem 2.4 in Rosenberg et al. (2006) adapted to our framework, is true.

Lemma 3. *The probability that $v(a_\infty, \pi_t) = v_*(\pi_t)$ for t large enough is one for a player who follows an optimal experimentation strategy.*

An immediate corollary of Lemmas 2 and 3 is that in any optimal experimentation strategy, the probability that only one action is chosen infinitely often is one.[14]

Lemma 4. *The probability that a player who follows an optimal experimentation strategy chooses two or more actions infinitely many times is zero.*

By Assumption A1, there exists $a^1 \in A$ and $\Theta_a^1 \subset \Theta_a$ for all $a \neq a^1$ with the property that if θ is such that $\theta_a \in \Theta_a^1$, then $r(a, \theta) < r(a^1, \theta)$. In other words, if a's type is in Θ_a^1, then the expected reward from a is smaller than the expected reward from a^1 regardless of a^1's type. Thus, by Lemmas 2 and 3, if a's type is in Θ_a^1, then with probability one a player who follows an optimal experimentation strategy chooses a only a finite number of times.

Lemma 5. *Suppose that action a's type lies in Θ_a^1. The probability that a player who follows an optimal experimentation strategy chooses a infinitely many times is zero.*

[14]This result follows immediately from Lemmas 1 and 2 if the number of actions is two.

6 Social Conformity

Here we prove that in equilibrium almost all players choose the same action in the long–run. We start with a preliminary result.

Lemma 6. *Let* (m^*, F^*) *be an equilibrium. Then,* $m^*_t(a, \theta)$ *is convergent for all* $a \in A$ *and* $\theta \in \Theta$. *Moreover, when* $m^*_t(a, \theta)$ *does not converge to zero, the mass of players who choose action* a *infinitely many times when the state of the world is* θ *is positive.*

Both results in Lemma 6 are intuitive. If the mass of players who choose some action a does not converge in the long–run, this means that a positive fraction of the population switches into a and away from it infinitely many times. Since there are finitely many actions, this implies that a positive mass of players chooses a and some other action infinitely often, which is not possible by Lemma 4. Moreover, also by Lemma 4, the mass of players who choose a infinitely many times is zero only if the mass of players who stop choosing a after a finite number of periods is one. A formal proof of Lemma 6 is in Appendix C.

Theorem 1. *Let* (m^*, F^*) *be an equilibrium. For each* $\theta \in \Theta$, *there exists* $a \in A$ *such that* $m^*_t(a, \theta)$ *converges to one.*

In the remainder of this section we show how we can bootstrap the reasoning used in the example of Section 2 to prove Theorem 1 when the number of actions is two. The argument is similar to the one used in the proof of the main result of Aoyagi (1998). One important difference is that in our setting the arms are independent, while in Aoyagi's setting they are correlated. The proof of the general case is in Appendix C.

Let $r_a(\theta_a)$ be the expected reward from a when its true type is θ_a. For any two elements θ_a and θ'_a of Θ_a, we say that θ_a is smaller than θ'_a if $r_a(\theta_a) < r_a(\theta'_a)$. Order the elements of Θ_{a_1} and Θ_{a_2} from lowest to highest. Then, $\Theta_{a_1} = \{\theta^1_{a_1}, \ldots, \theta^{M_1}_{a_1}\}$ and $\Theta_{a_2} = \{\theta^1_{a_2}, \ldots, \theta^{M_2}_{a_2}\}$, where we assume, without any loss, that $r_{a_1}(\theta^1_{a_1}) < r_{a_2}(\theta^1_{a_2})$. There are two cases to consider: either $r_{a_1}(\theta^{M_1}_{a_1}) > r_{a_2}(\theta^{M_2}_{a_2})$ or $r_{a_1}(\theta^{M_1}_{a_1}) < r_{a_2}(\theta^{M_2}_{a_2})$. We consider the second case only. It is straightforward to adapt the argument that follows to the first case.

For any $A \subseteq \Theta_a$ and $B \subseteq \Theta_{a'}$, we write $A < B$ when $r_a(\theta_a) < r_{a'}(\theta_{a'})$ for all $\theta_a \in A$ and $\theta_{a'} \in B$. Since $r_{a_2}(\theta_{a_2}^{M_2}) > r_{a_1}(\theta_{a_1}^{M_1})$, there exist $M \in \mathbb{N}$ and sets $\Theta_{a_k}^n$, with $k \in \{1, 2\}$ and $n \in \{1, \ldots, M\}$, such that: (i) $\Theta_{a_k} = \bigcup_{n=1}^{M} \Theta_{a_k}^n$, with $\Theta_{a_k}^1 < \cdots < \Theta_{a_k}^M$; (ii) $\Theta_{a_1}^n < \Theta_{a_2}^n$ for all n. Let $\mathfrak{r}_k : \Theta_{a_k} \to \{1, \ldots, M\}$ be such that $\mathfrak{r}_k(\theta_{a_k}) = n$ if $\theta_{a_k} \in \Theta_{a_k}^n$. Then, $\mathfrak{r}_1(\theta_{a_1}) \leq \mathfrak{r}_2(\theta_{a_2})$ if, and only if, $r_{a_1}(\theta_{a_1}) < r_{a_2}(\theta_{a_2})$. Finally, let $\Upsilon : \Theta \to \{1, \ldots, 2M\}$ be such that

$$\Upsilon(\theta) = \begin{cases} 2\mathfrak{r}_1(\theta_{a_1}) - 1 & \text{if} \quad \mathfrak{r}_1(\theta_{a_1}) \leq \mathfrak{r}_2(\theta_{a_2}) \\ 2\mathfrak{r}_2(\theta_{a_2}) & \text{if} \quad \mathfrak{r}_2(\theta_{a_2}) < \mathfrak{r}_1(\theta_{a_1}) \end{cases}.$$

Let (m^*, F^*) be an equilibrium. Lemmas 5 and 6 imply that $m_t^*(a_1, \theta)$ converges to zero if $\Upsilon(\theta) = 1$. Suppose then, by induction, that there exists $\overline{\Upsilon} < 2M$ such that if $\Upsilon(\theta) \leq \overline{\Upsilon}$, then $m_t^*(a, \theta)$ converges to one for some action a. We only consider the case where $\overline{\Upsilon}$ is odd, as the argument is the same when $\overline{\Upsilon}$ is even. Let the state of the world be $\widehat{\theta}$ such that $\Upsilon(\widehat{\theta}) = \overline{\Upsilon} + 1$. Since $\overline{\Upsilon} + 1$ is even, $r_{a_1}(\widehat{\theta}_{a_1}) > r_{a_2}(\widehat{\theta}_{a_2})$ and $\mathfrak{r}_2 = (\overline{\Upsilon} + 1)/2$. Suppose, by contradiction, that neither $m_t^*(a_1, \widehat{\theta})$ nor $m_t^*(a_2, \widehat{\theta})$ converge to zero. By Lemma 6, for each action there is a positive mass of players who choose this action an infinite number of times. Consider a player who chooses a_2 infinitely often. The induction hypothesis implies that he learns with probability one that θ is such that $\Upsilon(\theta) \geq \overline{\Upsilon} + 1$—we prove this (intuitive) fact in Appendix C when we analyze the general case. He also learns with probability one that $\mathfrak{r}_2 = (\overline{\Upsilon} + 1)/2$, for he learns the true type of a_2 with probability one by Lemma 2. Thus, this player learns with probability one that $\mathfrak{r}_1 > \mathfrak{r}_2$, since $\Upsilon(\theta) = 2\mathfrak{r}_1 - 1 \leq 2\mathfrak{r}_2 - 1 = \overline{\Upsilon}$ if $\mathfrak{r}_1 \leq \mathfrak{r}_2$. We then have that a positive fraction of the population chooses a_2 infinitely many times even though it learns that a_2 is worse than a_1, a contradiction by Lemma 3.

7 Complete Social Learning

As in Section 2, let $a(\theta)$ be the best action when the state of the world is θ. Assumption A1 implies that $a(\theta)$ is well–defined for all $\theta \in \Theta$. We say that social learning is complete in the equilibrium (m^*, F^*) if $m_t^*(a(\theta), \theta)$ converges to one for all $\theta \in \Theta$. In this section we discuss conditions under which social learning is complete in every equilibrium.

An obvious necessary condition for social learning to be complete in (m^*, F^*) is that $m_t^*(a(\theta), \theta)$ has a positive limit for each $\theta \in \Theta$. Theorem 1 shows that this condition is also sufficient. Thus, as in Bala and Goyal (1998), a necessary and sufficient condition for social learning to be complete in an equilibrium is that: (C) for each $\theta \in \Theta$ there exists a positive mass of players for which the probability that they choose $a(\theta)$ infinitely often is greater than zero. We now present a restriction on the distribution of initial beliefs that if satisfied implies that (C) holds in every equilibrium. For this, let $\underline{d}_\theta = \inf\{d \in [0,1] : v(a(\theta), \pi) \geq v_*(\pi) \text{ if } \pi(\theta) \geq d\}$. By definition, $a(\theta)$ is the myopically optimal action for a player with belief π such that $\pi(\theta) > \underline{d}_\theta$.

(H) For all $\theta \in \Theta$ there exists a positive mass of players with prior π_1 such that $\pi_1(\theta) > \underline{d}_\theta$.

The proof that condition (H) implies that (C) holds in every equilibrium borrows from Bala and Goyal (1998). The idea is to show that if (H) is satisfied, then there is a positive mass of players for which $a(\theta)$ is the myopically optimal action in the long–run when the state of the world is θ. By Lemma 3, almost all these players must then choose $a(\theta)$ for t large enough.

Let (m^*, F^*) be an equilibrium and suppose the state of the world is θ'. Consider a player with prior π_1 such that $\pi_1(\theta') \geq \underline{d}_{\theta'}$ and let σ be his behavior strategy. Now let $q_t : H_{t+1} \times \Theta \to [0,1]$ be such that $q_t(h^{t+1}, \theta)$ is the probability that he experiences h^{t+1} when the state of the world is θ. By construction, if h^{t+1} has positive probability under $\mu(\sigma, \pi_1, m^*)$, then

$$\pi_{t+1}(h^{t+1})(\theta') = \pi_{t+1}(\theta' | h^{t+1}) = \pi_1(\theta') \left\{ \pi_1(\theta') + \sum_{\theta \neq \theta'} \pi_1(\theta) \frac{q_t(h^{t+1}, \theta)}{q_t(h^{t+1}, \theta')} \right\}^{-1}.$$

It is well–known, see Doob (1953) for example, that for each $\theta \neq \theta'$ the likelihood ratio $p_t^{\theta',\theta}(\,\cdot\,) = q_t(\,\cdot\,, \theta)/q_t(\,\cdot\,, \theta')$ is a non–negative supermartingale when the state of the world is θ'. So, there exists a function $p_\infty^{\theta',\theta}$ on H_∞ with $\mathrm{E}_{\mu_\theta(\sigma, m^*)}\left[p_\infty^{\theta',\theta}\right] \leq \mathrm{E}_{\mu_\theta(\sigma, m^*)}[p_1^{\theta',\theta}] = 1$ such that $p_t^{\theta',\theta}$ converges $\mu_{\theta'}(\sigma, m^*)$–almost surely to $p_\infty^{\theta',\theta}$. Egoroff's theorem then implies that for each $\kappa > 0$ there exists $A_\kappa \subseteq H_\infty$ with $\mu_{\theta'}(\sigma, m^*)(A_\kappa) > 0$ such that $p_t^{\theta',\theta} \leq 1 + \kappa$ in

A_κ when t is sufficiently large. So, by choosing κ appropriately, there exists $A \subseteq H_\infty$ with $\mu_{\theta'}(\sigma, m^*)(A) > 0$ such that $\pi_{t+1}(\theta'|\cdot) > \underline{d}_{\theta'}$ in A for t sufficiently large, which leads to the desired result. We have thus established the following result.

Theorem 2. *Social learning is complete in any equilibrium if* (H) *holds.*

The next example shows (not surprisingly) that (H) is not necessary for social learning to be complete.

Example 1. The number of actions is two, $Y = \{0, 1\}$, and $u(a, y) = y$. Each action a_i has two possible types, θ^i_L and θ^i_H, with the distribution of outcomes as a function of the action and its type given by the following two tables:

a_1	$y=0$	$y=1$
θ^1_L	1	0
θ^1_H	0	1

a_2	$y=0$	$y=1$
θ^2_L	3/4	1/4
θ^2_H	1/4	3/4

All players have the same prior π_1 and it satisfies $D_{a_1}(\pi_1) > 0$. Then, by Lemma 1, (almost) all players choose a_1 in the first period and learn its true type. So, from the second period on, all players choose the best alternative regardless of the state of the world. ■

In general, it is difficult to determine whether (H) implies that for each action a there exists a positive mass of players who choose this action in the first period. The next example shows that this need not be the case.[15]

Example 2. The number of actions is two, $Y = \{0, 1\}$, and $u(a, y) = y$. Action a_1 has known rewards, with both outcomes equally likely, and action a_2 is as in the Example 1.

We describe beliefs by a pair $\pi = (\pi_L, \pi_H)$, where π_k is the probability that a_2's type is θ^2_k. A positive fraction of the population, the type I players, has a prior $\pi_1 = (\pi_{1L}, \pi_{1H})$ such that $\pi_{1H} > 1/2$. The remaining fraction, the type II players, has a prior π'_1 such that

[15]This is in contrast to the result on complete social learning in Bala and Goyal (1998). The condition they consider requires that for each state of the world there exists a positive fraction of the population that chooses the optimal action in the first period, see Theorem 4.1 in their paper. The reason for this difference is that agents can be forward looking in our environment.

$\pi'_{1L} = p \in (1/2, 2/3)$. Hence, condition (H) is satisfied. We claim that, nevertheless, all players choose a_1 in period one if they are patient enough.

It is straightforward to show that since a_1 has known rewards, a player chooses a_2 as long as it is myopically optimal to do so. So, all the type I players choose a_2 in period one. Consider now a type II player. His lifetime payoff if he chooses a_2 in period one is

$$v_2 = (1-p) + \frac{\delta}{1-\delta}\left[\frac{p}{2} + (1-p)\right].$$

On the other hand, an upper bound for his lifetime payoff if he chooses a_1 in period one is

$$\overline{v}_1 = \frac{1}{2} + \delta\left\{\frac{1}{2} + \frac{\delta}{1-\delta}\left[\frac{p}{2} + (1-p)\right]\right\}.$$

Indeed, $\overline{v}_1$ is the lifetime payoff to a type II player who chooses a_1 in the first period and learns the true type of a_2 at the end of period two (so that it is optimal for him to behave myopically in period two). It is easy to see that $v_2 - \overline{v}_1 > 0$ if δ is close enough to one, and so the type II players choose a_2 in period one if they are patient enough. ■

An interesting question is whether having each action being chosen by a positive fraction of the population in the first period is enough for social learning to be complete. The next example shows that this is not the case.

Example 3. There are two actions, $Y = \{0, 1, 2\}$, and $u(a, y) = y$. Action a_1 has known rewards, with the probability of $y = 0$ equal to $1 - p$ and the probability of $y = 2$ equal to p, where $1/4 < p < 1/2$. Action a_2 has three possible types, θ_L, θ_M, and θ_H, with the distribution of outcomes as a function of a_2's type given by the following table:

	$y = 0$	$y = 1$	$y = 2$
θ_L	1/2	1/2	0
θ_M	0	1	0
θ_H	0	0	1

Thus, a_2 is better than a_1 when its type is either θ_M or θ_H. Notice, incidentally, that the likelihood function for a_2 satisfies the monotone likelihood ratio property.

We describe beliefs by a triple $\pi = (\pi_L, \pi_M, \pi_H)$, where π_s is the probability that a_2 is of type θ_s. Since $v(a_1, \pi) = 2p$ and $v(a_2, \pi) = 2\pi_H + \pi_M + \frac{1}{2}\pi_L$, we then have that

$$D_{a_1}(\pi) = \frac{2p}{1-\delta} - \left\{ 2\pi_H + \pi_M + \frac{1}{2}\pi_L + \frac{\delta}{1-\delta}[2\pi_H + \pi_M + 2p\pi_L] \right\}.$$

Notice that if π is such that $D_{a_1}(\pi_L, 0, 1-\pi_L) > 0$, then $D_{a_1}(\pi_L, \pi_M, \pi_H) > 0$. From this, it is straightforward to see that $D_{a_1}(\pi) > 0$ if

$$\pi_L > \frac{2-2p}{2 - \frac{1}{2}(1-\delta) - 2p\delta}. \tag{2}$$

A fraction $1-\varepsilon$ of the players, where $\varepsilon \in (0,1)$, has a prior $\pi_1 = (\pi_{1L}, \pi_{1M}, \pi_{1H})$ that assigns probability greater than p to θ_H. We refer to these players as the type I players. The remaining fraction of players, the type II players, has a prior $\pi'_1 = (\pi'_{1L}, \pi'_{1M}, \pi'_{1H})$ where π'_{1L} satisfies (2). As in Example 2, since a_1 has known rewards, a player chooses a_2 as long as it is myopically optimal to do so. So, in period one the type I players choose a_2 (since $v(a_2, \pi_1) > 2p$) and the type II players choose a_1 (by Lemma 1).

The observations in society are uninformative in the first period. Therefore, a type I player who observes $y = 1$ in the initial period updates his belief to $\pi_2^* = (\pi_{2L}^*, \pi_{2M}^*, 0)$, where $\pi_{2M}^* = (1 + \pi_{1L}/2\pi_{1M})^{-1}$. Since $\lim_{\pi_M \to 0} D_{a_1}(1-\pi_M, \pi_M, 0) = 2p - 1/2 > 0$, $D_{a_1}(\pi_2^*) > 0$ if π_{1L}/π_{1M} is high enough. We assume that π_1 is such that $D_{a_1}(\pi_2^*) > 0$.

Notice that all type II players choose a_1 in the second period (since they do not update their beliefs). Now notice that: (i) a type I player who observes $y = 2$ in period one learns that $\theta = \theta_H$, and so chooses a_2 from period two on; (ii) a type I player who observes $y = 0$ in period one learns that $\theta = \theta_L$, and so chooses a_1 from period two on. Finally, notice, by assumption, that a type I player who observes $y = 1$ in period one chooses a_1 in period two. Hence, $m_2(a_2, \theta_H) = 1 - \varepsilon$ and $m_2(a_2, \theta_L) = m_2(a_2, \theta_M) = 0$.

Let us now consider behavior in period three. Consider first a type I player who observes $y = 1$ in the first period (and so learns that $\theta \neq \theta_H$). Because the probability that in period two he observes someone who chooses a_2 is the same whether $\theta = \theta_L$ or $\theta = \theta_M$, his period–two observation in society does not reveal any new information about θ. So, this player still chooses a_1 in period three. In particular, the mass of players who choose a_2 in period three is the same whether $\theta = \theta_L$ or $\theta = \theta_H$, which implies that the observations in society in period three do not reveal any new information about θ to this player (and so he chooses a_1 in period four, and so on).

Consider now a type II player. If he observes someone choosing a_2 in period two, he learns that $\theta = \theta_H$, and so he chooses a_2 from period three on. If, instead, he observes someone choosing a_1 in period two, he updates his belief to $\pi'_3 = (\pi'_{3L}, \pi'_{3M}, \pi'_{3H})$, where

$$\pi'_{3L} = \frac{\pi'_{1L}}{\pi'_{1L} + \pi'_{1M} + \varepsilon\pi'_{1H}} > \pi'_{1L},$$

since $\varepsilon < 1$. So, this player chooses a_1 in period 3. Thus, $m_3(a_2, \theta_L) = m_3(a_2, \theta_M) = 0$ and $m_3(a_2, \theta_H) = m_2(a_2, \theta_H) + m_2(a_1, \theta_H)m_2(a_2, \theta_H) = 1 - \varepsilon^2$.

It is now easy to see how type I players who observe $y = 1$ in period one and type II players behave over time. The type I players who observe $y = 1$ in the first period choose a_1 from period two on. The type II players keep choosing a_1 as long as they don't observe someone who chooses a_2 (every time this happens they put more weight on $\theta = \theta_L$). The first time they observe someone who chooses a_2, they switch permanently to a_2. Consequently $m_t(a_2, \theta_H) = 1 - \varepsilon^{t-1}$ and $m_t(a_2, \theta_L) = m_t(a_2, \theta_M) = 0$ for all $t \geq 2$.

There is then a unique equilibrium and in this equilibrium the mass of agents who choose a_2 when its type is θ_M is zero after the first period.[16] It is possible to extend this example to the case where a_1 has two types, one that is better than θ_L and worse than θ_M and another that is better than θ_M and worse than θ_H. ■

We know that Lemma 1 holds regardless of how many action choices (from a random sample of the population) a player observes in each period. So, the part of Example 3 showing that in equilibrium $m_t(a_2, \theta_M) = 0$ for all $t \geq 2$ does not depend on the assumption that players only observe one other action choice in each period. Thus, social learning can fail to be complete if (H) is violated even in a setting where in each period players can observe the action choices of infinitely many other players in the population.

[16] Recall that strategy profiles are defined modulo the behavior of a mass zero of players.

8 Discussion

We assume that players are anonymous. When they are non–anonymous, the information from the observations in society is no longer described by a sequence of observation likelihoods. Instead, it is described by a sequence $\{\ell_t\}$, with $\ell_t : [0,1] \times \Theta \times A \to [0,1]$, where $\ell_t(i, \theta, a)$ is the probability that player i chooses action a in period t when the state of the world is θ. All the results in the paper can be extended to the case of non–anonymous players, with our equilibrium notion adapted in a natural way.

A natural question to ask is how long does it take for all the players to choose the best alternative when social learning is complete; or, more broadly, how long does it take for behavior to converge (whether social learning is complete or not). Example 3 in Section 7 shows that convergence of behavior need not happen after a finite number of periods. In general, an answer to this question will depend on the distribution of prior beliefs. We do note that while heterogeneity of prior beliefs works in favor of complete social learning, it can work against the convergence of behavior. An extreme example of this is as follows.

Suppose that for each $\theta \in \Theta$ there exists a fraction $1/|\Theta|$ of the population with a prior π_1 such that $D_{a(\theta)}(\pi) > 0$ for every belief π that can be reached from π_1 after one chooses $a(\theta)$ for $T \geq 1$ times (or less), ignoring the observations in society.[17] Notice that (H) holds, and so social learning is complete in any equilibrium. However, by Lemma 1, in all equilibria each player chooses the same action from periods one to T no matter the state of the world (as long as the agents' behavior does not depend on the state of the world, the observations in society are uninformative). In particular, when T is large, it takes a long time before convergence of behavior can happen.

A related question is how the speed of convergence (in behavior) is affected by the number of action choices the players observe in each period. It is easy to adapt our analysis to the case where in each period a player observes the action choices of a random sample of $N > 1$ other players. There are two opposing forces when one increases N. Holding behavior the

[17] This is possible if $g(y|a, \theta) > 0$ for all $y \in Y$, $a \in A$, and $\theta \in \Theta$.

same, the greater N is, the more information the observations in society provide. However, the players' behavior is affected by the size of the sample they observe in each period. Indeed, if players expect the information from the observations in society to be very useful (because N is large), they will not be too willing to experiment, i.e., choose a myopically suboptimal action, thus reducing the informativeness of the observations in society.

We don't have any results on how the speed of convergence is affected by N. Since Lemma 1 holds for all values of N, the example two paragraphs above nevertheless shows that no matter how large N is, the amount of time before behavior starts to converge can be arbitrarily large depending on the distribution of prior beliefs in the population.

In general, the assumption of a continuum of individuals is done in order to simplify the analysis. One might worry that in our setting this assumption may be working too hard. Let us argue (informally) that this is not the case. For this, consider the finite–agent version of our model. It is clear that all the results from Section 5 hold in this case. A straightforward modification of the argument of Section 6 shows, as in Aoyagi (1998), that in any Nash equilibrium all the players choose the same action in the long run.[18]

Suppose now that (H) is satisfied and consider players i and j with a prior π_1 such that $\pi_1(\theta) \geq \underline{d}_\theta$. The same argument of Section 7 shows that the events $A^i_{a(\theta)}$ and $A^j_{a(\theta)}$ that i and j, respectively, choose $a(\theta)$ for t large enough have positive probability when the state of the world is θ. Since the population is finite, these events are not independent, though. However, in the limit as the population grows to infinity, these events will be independent. Thus, as the population grows, the probability that social learning is complete goes to one.

9 Existence

A strategy profile F is symmetric if F is constant in $\Phi^{-1}(\pi)$ for each $\pi \in \Pi$. In other words, a strategy profile is symmetric if for every belief π almost all players with prior π

[18]Notice that the finite–population version of our environment is different from Aoyagi (1998) in two ways. First, as mentioned in Section 6, we consider a multi–armed bandit with independent arms, while Aoyagi considers a multi–armed bandit with correlated arms. Second, action choices are not public.

follow the same behavior strategy. An equilibrium (m^*, F^*) is symmetric if F^* is symmetric. Assumption A4 below implies that the set of beliefs for which a positive fraction of the population has a prior in this set is countable.

Assumption A4. *There exists a countable set* $\widehat{\Pi}$ *of full support priors with* $\lambda(\Phi^{-1}(\widehat{\Pi})) = 1$.

We now show that a symmetric equilibrium exists if A4 is satisfied.

Theorem 3. *A symmetric equilibrium exists if* A4 *is satisfied.*

Here we just outline the existence argument. The omitted details are in Appendix D. Let $\widehat{\Pi}_+ = \{\pi \in \widehat{\Pi} : \lambda(\Phi^{-1}(\pi)) > 0)\}$. This set is countable by A4. Let $\{\pi_n\}$ be an enumeration of $\widehat{\Pi}_+$ and define $(\Omega, 2^\Omega, \mu)$ to be the complete measure space where $\Omega = \{\omega \in \mathbb{N} : \pi_\omega \in \widehat{\Pi}\}$ and $\mu(E) = \sum_{\omega \in E} \lambda(\Phi^{-1}(\pi_\omega))$ for all $E \subseteq \Omega$—recall that 2^Ω is the power set of Ω. To each $F_t \in \mathrm{L}_1(\mu, \Sigma_t)$, let $\widehat{F}_t : I \to \Sigma_t$ be given by $\widehat{F}_t = \sum_{\omega \in \Omega} F_t(\omega) I_{\Phi^{-1}(\pi_\omega)}$. Since Φ is Lebesgue measurable and $\lambda(\Phi^{-1}(\widehat{\Pi})) = 1$, $\widehat{F}_t$ is Lebesgue measurable. Moreover, $\widehat{F}_t \in \mathrm{L}_1(\lambda, \Sigma_t)$ by the dominated convergence theorem. Hence, to each element of $\times_{t=1}^\infty \mathrm{L}_1(\mu, \Sigma_t)$ there is associated a symmetric strategy profile. The converse is immediate. From now on, we refer to $\times_{t=1}^\infty \mathrm{L}_1(\mu, \Sigma_t)$ as the set of symmetric strategy profiles and denote it by $\mathfrak{S}_S$.

Denote restriction of M to $\mathfrak{S}_S$ by M_S. By construction, M_S is the map that takes a symmetric strategy profile into its corresponding sequence of observation likelihoods. It is possible to show that M_S is a continuous and affine map.[19] Now let $\mathrm{BR} : \mathfrak{M} \rightrightarrows \mathfrak{S}_S$ be the correspondence such that $F \in \mathrm{BR}(m)$ if, and only if, $F(\omega)$ is an optimal experimentation strategy for $\mathrm{ILP}(\pi_\omega, m)$. It is possible to show that BR is upper hemicontinuous with convex (and non–empty) values. By Lemma 16.23 in Aliprantis and Border (1999), the composition of upper hemicontinuous correspondences is upper hemicontinuous. So, the correspondence $\Upsilon : \mathfrak{M} \to \mathfrak{M}$ such that $\Upsilon(m) = \bigcup_{F \in \mathrm{BR}(m)} \mathsf{M}_S(F)$ is upper hemicontinuous with convex (and non–empty) values. The Fan–Glicksberg Theorem then implies that Υ has a fixed point m^*. Let $F^* \in \mathrm{BR}(m^*)$. The pair (m^*, F^*) is a symmetric equilibrium.

[19] A function $f : X \to Y$, with X and Y convex, is affine if $f(\lambda x + (1 - \lambda)x') = \lambda f(x) + (1 - \lambda) f(x')$ for all $x, x' \in X$ and all $\lambda \in [0, 1]$.

10 Conclusion

This paper shows that social learning can overcome the Rothschild effect. We first show that there is long–run social conformity in equilibrium. We then use this fact to derive a condition on the distribution of prior beliefs that if satisfied implies that social learning is always complete. The condition we obtain is relatively weak: for each state of the world θ there exists a positive fraction of the population with a prior that assigns a high enough probability to θ for the best alternative in θ to be *myopically* optimal. We also show that this condition is not necessary and, more interestingly, that the alternative condition where for each θ there exists a positive mass of players who in period one choose the best alternative when the state of the world is θ is not sufficient for social learning to be complete.

An important element of our analysis is the process by which information is transmitted across the population. The assumption that every so often an individual observes the action choices of a finite random sample from the population may not be adequate in some cases. Moreover, this process of information transmission ignores non–random sources of information that individuals can have, like their social and professional circles. It is possible to show that the same results obtain in an environment in which each individual has a finite set of neighbors, any two neighbors can always observe each other's actions, *and* every so often an individual observes the action choices of a finite random sample from the population—see Appendix E (not for publication) for a sketch of the argument.

References

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

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

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

[4] A. Banerjee (1992), "A Simple Model of Herd Behavior", *Quaterly Journal of Economics*, 107(3), 797-817.

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

[6] S. Bikhchandani, D. Hirshleifer, and I. Welch (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades", *Journal of Political Economy*, 100(5), 992-1026.

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

[8] M. Brezzi and T. Lai (2000), "Incomplete Learning from Endogenous Data in Dynamic Allocation", *Econometrica*, 68(6), 1511-1516.

[9] A. Caplin and J. Leahy (1994), "Business as Usual, Market Crashes, and Wisdom After the Fact", *American Economic Review*, 84(3), 548-565.

[10] C. Chamley and D. Gale (1994), "Information Revelation and Strategic Delay in a Model of Investment", *Econometrica*, 62(5), 1065-1085.

[11] J. Doob (1953), *Stochastic Processes*. New York: John Wiley & Sons.

[12] N. Dunford and J. Schwartz (1988), *Linear Operators Part I: General Theory*, Wiley Classics Library. New York: John Wiley & Sons.

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

[14] D. Easley and N. Kiefer, "Controlling a Stochastic Process with Unkown Parameters", *Econometrica*, 56(5), 1045-1064.

[15] G. Ellison and D. Fudenberg (1993), "Rules of Thumb for Social Learning", *Journal of Political Economy*, 101(4), 612-643.

[16] G. Ellison and D. Fudenberg (1995), "Word–of–Mouth Communication and Social Learning", *Quaterly Journal of Economics*, 110(1), 93-125.

[17] D. Gale and S. Kariv (2003), "Bayesian Learning in Social Networks", *Games and Economic Behavior*, 45, 329-346.

[18] K. Hinderer (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.

[19] W. Juang (2001), "Learning from Popularity", *Econometrica*, 69(3), 735-747.

[20] G. Keller, S. Rady, and M. Cripps (2005), "Strategic Experimentation with Exponential Bandits", *Econometrica*, 73(1), 39-68.

[21] R. Kihlstrom, L. Mirman, and A. Postlewaite (1984), "Experimental Consumption and the "Rothschild Effect"", in M. Boyer and R. Kihlstrom (Eds.), *Bayesian Models in Economic Theory*. Amsterdan: North–Holland.

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

[23] D. Rosenberg, E. Solan, and N. Vielle (2006), "Informational Externalities and Converge of Behavior", mimeo.

[24] D. Rosenberg, E. Solan, and N. Vielle (2007), "Social Learning in One-Arm Bandit Problems", *Econometrica*, 75(6), 1591-1611.

[25] M. Rothschild (1974), "A Two-Armed Bandit Theory of Market Pricing", *Journal of Economic Theory*, 9, 185-202.

[26] D. Schmeidler (1973), "Equilibrium Points of Nonatomic Games", *Journal of Statistical Physics*, 7, 295-300.

[27] D. Smallwood and J. Conlisk (1979), "Product Quality in Markets Where Consumers are Perfectly Informed", *Quaterly Journal of Economics*, 93(1), 1-23.

[28] L. Smith and P. Sorensen (2000), "Pathological Outcomes of Observational Learning", *Econometrica*, 68(2), 371-398.

Appendix A: From Strategies to Aggregate Behavior

Here we construct the map M that takes strategy profiles into sequences of observation likelihoods. Let $\{e_1, \ldots, e_K\}$ be the canonical basis of $\mathbb{R}^K$ and $\langle\ ,\ \rangle$ be the usual scalar product. Notice that if $F_t \in \mathrm{L}_1(\lambda, \Gamma_t)$, then $\langle F_t(h^t), e_k\rangle \in \mathrm{L}_1(\lambda)$ for all $h^t \in H_t$, $t \in \mathbb{N}$, and $k \in \{1, \ldots, K\}$. Suppose that $F = \{F_t\}$ is the strategy profile under play. The probability that player i chooses a_k in $t = 1$ is $\ell_1(i, a_k) = \langle F_1(i)(\emptyset), e_k\rangle$. Since there is no aggregate uncertainty, the mass of players who choose a_k in $t = 1$ is then $m_1(a_k) = \int \ell_1(i, a_k)\lambda(di)$. Thus, the probability that player i experiences the period–2 history $h^2 = (a, y, a')$ when the state of the world is θ is $\tau_2(i, h^2, \theta) = \ell_1(i, a)g(y|a, \theta)m_1(a')$. We can now use $\tau_2(i, h^2, \theta)$ to construct the probability $\ell_2(i, a_k, \theta)$ that player i chooses a_k in $t = 2$ when the state of the world is θ. Aggregating the individual probabilities $\ell_2(i, a_k, \theta)$, we obtain $m_2(a_k, \theta)$.[20] A straightforward induction argument completes the construction of M.

[20] In fact, $\ell_2(i, a_k, \theta) = \sum_{h^2 \in H_2} \langle F_2(i)(h^2), e_k\rangle \tau_2(i, h_2, \theta)$. Now observe that for each state of the world θ, the event where player i's period–2 history is h^2 is independent of the event where player j's period–2 history is $h^{2\prime}$ for all $i \neq j$ and all $h^2, h^{2\prime}$ in H_2—this follows from the assumption that both the outcome of an action choice and the identity of the individual a player observes are determined independently for each player in the game. So, conditional on the state of the world, there is no aggregate uncertainty, which implies that $m_2(a_k, \theta) = \int \ell_2(i, a_k, \theta)\lambda(di)$ for each $\theta \in \Theta$.

Appendix B: Proof of Lemma 1

We start with a preliminary characterization result for the individual learning problem. Let $\gamma_t : A \times \mathfrak{M} \times \Theta \to \Delta(X)$ be given by $\gamma_t(a, m, \theta)(S) = \sum_{(y,a')\in S} g(y|a, \theta)m_t(a', \theta)$. Now let $q_t(h^t, a, m, \pi) = \sum_\theta \gamma_t(a, m, \theta)\pi_t(h^t, m, \pi)(\theta)$ and $r_t(h^t, a, m, \pi) = \sum_\theta r(a, \theta)\pi_t(h^t, m, \pi)(\theta)$, where $\pi_t(h^t, m, \pi)$ is the period–t belief of a player with history h^t and prior π when the sequence of observation likelihoods is m. Rieder (1975) shows that ILP(π_1, m) is equivalent to a non–stationary dynamic programming problem where the period–t reward function and transition probability are $r_t(\cdot, m, \pi_1)$ and $q_t(\cdot, m, \pi_1)$, respectively.[21] With this formulation, the Bellman equations for ILP(π_1, m) are

$$V_t(h^t) = \max_{a\in A}\left\{r_t(a, h^t, m, \pi_1) + \delta \int V_{t+1}(h^t, a, x)q_t(dx|h^t, a, m, \pi_1)\right\}, \quad t \in \mathbb{N}. \tag{B.1}$$

Let $\overline{R} = \overline{r}/(1 - \delta)$, where $\overline{r} = \max_{a,\theta} r(a, \theta)$. A standard contraction mapping argument shows that for each $\pi_1 \in \Pi$ and $m \in \mathfrak{M}$ there exists a unique sequence $\{V_t^*\}$ in $\times_{t=1}^\infty \mathrm{B}_b^{\overline{R}}(H_t)$ that solves (B.1). We omit the dependence of the V_t^* on π_1 and m for ease of notation. Standard dynamic programming results show that $V_t^*(h^t)$ is the optimal continuation payoff in ILP(π_1, m) to a player with history h^t. The next result follows from Theorems 15.4 and 17.7 in Hinderer (1970), where *supp*(ξ) denotes the support of the mixed action ξ.

Lemma 7. *The strategy $\sigma^* = \{\sigma_t^*\}$ is optimal for* ILP(π_1, m) *if, and only if, for all $t \geq 1$,*

$$supp(\sigma_t^*(h^t)) \subseteq \mathrm{argmax}_a \left\{r_t(a, h^t, m, \pi_0) + \delta \int V_{t+1}^*(h^t, a, x)q_t(dx|h^t, a, m, \pi_0)\right\}$$

for all $h^t \in H_t$ that happen with positive probability under σ_t^*.

The next result, that the belief is a sufficient statistic for the individual learning problem, is intuitive. Let $\nu_t : \Pi \times A \times \mathfrak{M} \to \Delta(X)$ be given by $\nu_t(\pi, a, m) = \sum_{\theta\in\Theta} \pi(\theta)\gamma_t(a, m, \theta)$ and define $\rho_t : \Pi \times A \times \mathfrak{M} \to \Delta(\Pi)$ to be such that if $S \in \mathcal{B}(\Pi)$, then

$$\rho_t(\pi, a, m)(S) = \int \mathrm{I}_S(B_t(\pi, a, x, m))\nu_t(dx|\pi, a, m),$$

[21] Formally, the individual learning problem is a stochastic dynamic decision problem with unknown transition probabilities.

where I_S is the indicator function of the set S and $B_t(\pi, a, x, m)$ is the updated belief, when the sequence of observation likelihoods is m, of a player who in period t has belief π, chooses a, and observes x. Now consider the sequence of functional equations given by

$$U_t(\pi) = \max_{a \in A} \left\{ v(a, \pi) + \beta \int U_{t+1}(\pi') \rho_t(d\pi' | \pi, a, m) \right\}, \quad t \in \mathbb{N}, \tag{B.2}$$

where $m \in \mathfrak{M}$. A standard contraction mapping argument shows that for each $m \in \mathfrak{M}$ there exists a unique sequence $\{U_t^*\}$ in $\times_{t=1}^{\infty} \mathrm{B}_b^{\overline{R}}(\Pi)$ that solves (B.2) and that

$$\max_{a \in A} (1-\delta)^{-1} v(a, \pi) \leq U_t^*(\pi) \leq (1-\delta)^{-1} \sum_{\theta \in \Theta} r(\theta) \pi(\theta) \tag{B.3}$$

for all $\pi \in \Pi$—we omit the dependence of the U_t^* on m for ease of notation.

Lemma 8. *$V_t^*(h^t) = U_t^*(\pi_t(h^t, m, \pi_1))$ for each $\pi_1 \in \Pi$ and $m \in \mathfrak{M}$.*

Proof: Let $w_t : H_t \to \mathbb{R}$ be such that $w_t(h^t) = U_t^*(\pi_t(h^t, m, \pi_1))$. Now observe, omitting the dependence of π_t on π_0 and m, that

$$\begin{aligned} r(h^t, a, m, \pi_0) + \delta \int w_{t+1}(h^t, a, x) q_t(dx | h^t, a, m, \pi_1) \\ &= v(a, \pi_t(h^t)) + \delta \int U_{t+1}^*(\pi_{t+1}(h^t, a, x)) \nu_t(dx | \pi_t(h^t), a, m) \\ &= v(a, \pi_t(h^t)) + \delta \int U_{t+1}^*(B_t(\pi_t(h^t), a, x, m)) \nu_t(dx | \pi_t(h^t), a, m) \\ &= v(a, \pi_t(h^t)) + \delta \int U_{t+1}^*(\pi') \rho_t(d\pi' | \pi_t(h^t), a, m), \end{aligned}$$

where the first equality follows from the definitions of $v(a, \pi)$ and ν_t, the second follows from the definition of π_t, and the third follows from the definition of ρ_t. Hence, $\{w_t\}$ is a sequence in $\times_{t=1}^{\infty} \mathrm{B}_b^{\overline{R}}(H_t)$ that satisfies (B.1), so that $w_t(h_t) = V_t^*(h_t, m, \pi_1)$ for all $t \geq 1$. □

The next result follows immediately from Lemmas 7 and 8. Lemma 1 is an immediate consequence of it together with the inequalities (B.3).

Corollary 1. *The strategy $\sigma^* = \{\sigma_t^*\}$ is optimal for* ILP(π_1, m) *if, and only if, for all $t \geq 1$,*

$$supp(\sigma_t^*(h^t)) \subseteq \mathrm{argmax}_a \left\{ v(a, \pi_t(h^t, m, \pi_1)) + \delta \int U_{t+1}^*(\pi') \rho_t(d\pi' | \pi_t(h^t, m, \pi_1), a, m) \right\}$$

for all $h^t \in H_t$ that happen with positive probability under σ_t^*.

Appendix C: Proofs of Lemma 6 and Theorem 1

Proof of Lemma 6: Let (m^*, F^*) be an equilibrium and $\tau : I \times \Theta \to \Delta(H_\infty)$ be such that $\tau(i,\theta)(S)$ is the probability that i's infinite history lies in $S \in \mathcal{B}(H_\infty)$ when the state of the world is θ. Let $S_{t,a}$ be the event that $a_t = a$. Then, $m_t^*(a,\theta) = \int \tau(i,\theta)(S_{t,a})\lambda(di)$.[22] Now let $S_a = \bigcap_{t=1}^\infty S_a^t$ and $\widetilde{S}_a = \bigcup_{t=1}^\infty \widetilde{S}_a^t$, where $S_a^t = \bigcup_{m=t}^\infty S_{m,a}$ and $\widetilde{S}_a^t = \bigcap_{m=t}^\infty S_{m,a}$. By construction, S_a is the event that a is chosen infinitely often and $\widetilde{S}_a \subset S_a$ is the event that a is chosen for t large enough. Since $\lim_t \tau(i,\theta)(\widetilde{S}_a^t) = \tau(i,\theta)(\widetilde{S}_a) \leq \tau(i,\theta)(S_a) = \lim_t \tau(i,\theta)(S_a^t)$ for each $i \in I$ and $\theta \in \Theta$ and $\widetilde{S}_a^t \subset S_{t,a} \subset S_a^t$ for all $t \geq 1$, we then have that

$$\int \tau(i,\theta)(\widetilde{S}_a)\lambda(di) \leq \liminf m_t^*(a,\theta) \leq \limsup m_t^*(a,\theta) \leq \int \tau(i,\theta)(S_a)\lambda(di).$$

So, $\{m_t^*(a,\theta)\}$ is convergent, as $\tau(i,\theta)(\widetilde{S}_a) = \tau(i,\theta)(S_a)$ for λ–almost all i for each $\theta \in \Theta$ by Lemma 4. Moreover, $\lambda(\{i : \tau(i,\theta)(S_a) > 0\}) > 0$ if $m_t^*(a,\theta)$ does not converge to zero. □

Proof of Theorem 1: We start with some notation. Let $\mathfrak{R} = \{r(a,\theta) : a \in A \text{ and } \theta \in \Theta\}$, with typical element $\mathfrak{r}$, be the set of possible expected rewards. Denote the smallest and greatest elements of $\mathfrak{R}$ by $\mathfrak{r}_1$ and $\mathfrak{r}_M$, respectively. For any $\mathfrak{r} < \mathfrak{r}_M$, let $\mathfrak{r}_+$ denote the successor of $\mathfrak{r}$, and for any $\mathfrak{r}' < \mathfrak{r}''$, let $[\mathfrak{r}', \mathfrak{r}''] = \{\mathfrak{r} \in \mathfrak{R} : \mathfrak{r}' \leq \mathfrak{r} \leq \mathfrak{r}''\}$. Now let $A \oplus \Theta = \bigcup_{a \in A} \{a\} \times \Theta_a$, with typical element ξ, and define $\Lambda : \mathfrak{R} \to A \oplus \Theta$ and $\aleph : A \oplus \Theta \to A$ to be such that $r(\Lambda(\mathfrak{r})) = \mathfrak{r}$ and $\aleph(\xi)$ is the first component of ξ, respectively. By construction, if $\mathfrak{R}' \subseteq \mathfrak{R}$, then $\Lambda(\aleph(\mathfrak{R}'))$ is the set of all actions a for which there exists θ with $r(a,\theta) \in \mathfrak{R}'$.

As the first step in the argument, we define the family $\{\mathfrak{R}^q, a^q, \Theta^q\}_{q \in Q}$, where Q is a finite subset of $\mathbb{N}$, and $\mathfrak{R}^q \subset \mathfrak{R}$, $a^q \in A$, and $\Theta^q \subset \Theta$ for all $q \in Q$, by the following procedure:

I) Let $\mathfrak{r}^* = \max\{\mathfrak{r} : |\aleph(\Lambda([\mathfrak{r}_1, \mathfrak{r}]))| = K - 1\}$ and set $\mathfrak{R}^1 = [\mathfrak{r}_1, \mathfrak{r}^*]$. Recall that K is the number of actions. Now let $a^1 = \aleph(\Lambda(\mathfrak{r}_+^*))$, so that a^1 is the action that does not belong to the set $\aleph(\Lambda([\mathfrak{r}_1, \mathfrak{r}]))$. Moreover, for each $a \neq a^1$, let $\Theta_a^1 = \{\theta_a \in \Theta_a : r_a(\theta_a) \in \mathfrak{R}^1\}$. Then, $\Theta^1 = \Theta_{a^1} \times (\times_{a \neq a^1} \Theta_a^1)$. By construction, the action a^1 and the sets Θ_a^1 are the same as in the last paragraph of Section 5;

[22] It is easy to show that the map $i \mapsto \tau(i,\theta)(S)$ is λ–measurable for all $\theta \in \Theta$ and $S \in \mathcal{B}(H_\infty)$.

II) Suppose there exists $\overline{q} \geq 1$ such that: (i) the triples $(\mathfrak{R}^q, a^q, \Theta^q)$ are defined for $q \leq \overline{q}$; (ii) $|\aleph(\Lambda(\mathfrak{R}^q))| = K - 1$ and $|\aleph(\Lambda(\mathfrak{R}^q \cup \{\mathfrak{r}(q)_+\}))| = K$ for all $q \in \{1, \ldots, \overline{q}\}$, where $\mathfrak{r}(q)$ is the greatest element of $\mathfrak{R}^q$. If $\mathfrak{r}(\overline{q}) = \mathfrak{r}_M$, then $Q = \{1, \ldots, \overline{q}\}$ and the process stops. Otherwise, let $\mathfrak{r}^{**} = \max\{\mathfrak{r} \in [\mathfrak{r}(\overline{q})_+, \mathfrak{r}_M] : \aleph(\Lambda([\mathfrak{r}(\overline{q})_+, \mathfrak{r}^{**}])) = 1\}$. Then, $a^{\overline{q}+1} = \aleph(\Lambda(\mathfrak{r}^{**}_+))$ and $\mathfrak{R}^{\overline{q}+1} = [\mathfrak{r}(\overline{q})_+, \mathfrak{r}^{**}] \cup (\mathfrak{R}^{\overline{q}} \setminus \{\mathfrak{r} \in \mathfrak{R}^{\overline{q}} : \aleph(\Lambda(\mathfrak{r})) = a^{\overline{q}+1}\})$. Now, for each $a \neq a^{\overline{q}+1}$, let $\Theta_a^{\overline{q}+1} = \{\theta_a \in \Theta_a : r_a(\theta_a) \in \mathfrak{R}^{\overline{q}+1}\}$, and let $\Theta^{\overline{q}+1}_{a^{\overline{q}+1}} = \{\theta_{a^{\overline{q}+1}} \in \Theta_{a^{\overline{q}+1}} : r_{a^{\overline{q}+1}}(\theta_{a^{\overline{q}+1}}) \geq \mathfrak{r}^{**}_+\}$. Then, $\Theta^{\overline{q}+1} = \Theta^{\overline{q}+1}_{a^{\overline{q}+1}} \times \left(\times_{a \neq a^{\overline{q}+1}} \Theta_a^{\overline{q}+1}\right)$. Notice, by construction, that $|\aleph(\Lambda(\mathfrak{R}^q))| = K - 1$ and that $|\aleph(\Lambda(\mathfrak{R}^q \cup \{\mathfrak{r}(q)_+\}))| = K$ for all $q \in \{1, \ldots, \overline{q}+1\}$.

Observe that if $\theta \in \Theta^q$, then $r_{a^q}(\theta_{a^q}) > r_a(\theta_a)$ for all $a \neq a^q$. Also observe that $\bigcup_{k=1}^q \mathfrak{R}^k = [\mathfrak{r}_1, \mathfrak{r}(q)]$ for all $q \in Q$.

Let (m^*, F^*) be an equilibrium. By Lemma 5, $m^*_t(a^1, \theta)$ converges to one for all $\theta \in \Theta^1$. Suppose now, by induction, that there exists $\overline{q} \geq 1$ such that if $\theta \in \Theta^q$ for some $q \in \{1, \ldots, \overline{q}\}$, then there exists $a \in A$ such that $m^*_t(a, \theta)$ converges to one. Let the state of the world be $\widehat{\theta} \in \Theta^{\overline{q}+1}$ and assume, by contradiction, that there exist $a' \neq a''$ in A such that both $\{m^*_t(a', \widehat{\theta})\}$ and $\{m^*_t(a'', \widehat{\theta})\}$ have positive limits. We know that either $a' \neq a^{\overline{q}+1}$ or $a'' \neq a^{\overline{q}+1}$. Assume that $a' \neq a^{\overline{q}+1}$. By Lemma 5, we can take a' to be such that $r(a', \widehat{\theta}) \neq \mathfrak{R}^1$. So, by the above construction, there exists an unique $\widetilde{q} \in \{1, \ldots, \overline{q}\}$ such that $a' = a^{\widetilde{q}}$.

Now observe, by Lemma 6, that a positive mass of players choose a' infinitely many times and a positive mass of players choose a'' infinitely many times. Consider a player who chooses a' infinitely often. By Lemma 2, this player learns with probability one the true type of a'. Since there is no long–run convergence of behavior, the induction hypothesis implies that he also learns with probability one that $q \geq \overline{q} + 1$. Thus, with probability one, this player learns that a' is not myopically optimal—we prove this claim below. So, by Lemma 3, a positive fraction of the players is not behaving optimally, a contradiction.

We can then conclude, by induction, that for all $\theta \in \Theta$ there exists $a \in A$ such that $m^*_t(a, \theta)$ converges to one. □

Proof of the Claim: Consider a player i with a full support prior π_1 who follows an optimal experimentation strategy and chooses a' infinitely many times. By Lemma 4, we can assume without loss that he chooses a' in every period. Let $\gamma_\theta \in \Delta(X^\infty)$ be such that $\gamma_\theta(D)$ is the probability that he observes an infinite sequence of outcome–observation pairs in $D \subset X^\infty$. Moreover, let y_t be the player's period–t outcome and let b_t be such that $b_t = 1$ if he observes someone who chooses a' in period t and $b_t = 0$ otherwise. Notice that $(y_1, b_1, y_2, b_2, \ldots)$ is a sequence of independent random variables and that $\mathrm{E}_{\gamma_\theta}[b_t] = m_t^*(a', \theta)$.

Let $f_t(b_t, \theta) = m_t^*(a', \theta)^{b_t}(1 - m_t^*(a', \theta))^{1-b_t}$. If $\pi(\cdot|\pi_1, y_1, b_1, \ldots, y_t, b_t)$ is player i's updated belief after he observes $(y_1, b_1, \ldots, y_t, b_t)$, then

$$\pi(\theta|\pi_1, y_1, b_1, \ldots, y_t, b_t) = \frac{\Pi_{n=1}^t g(y_n|a', \theta) f_n(b_n, \theta)\pi_1(\theta)}{\sum_{\theta' \in \Theta} \Pi_{n=1}^t g(y_n|a', \theta') f_n(b_n, \theta')\pi_1(\theta')}$$

when $(y_1, b_1, \ldots, y_t, b_t)$ has positive probability. Now observe that $\pi(\theta|\pi_1, y_1, b_1, \ldots, y_t, b_t)$ is bounded above by $\{1 + L_t(\theta|\pi_1, y_1, b_1, \ldots, y_t, b_t)\}^{-1}$, where

$$L_t(\theta|\pi_1, b_1, y_1, \cdots, y_t, b_t) = \frac{\Pi_{n=1}^t g(y_n|a', \widehat{\theta}) f_n(b_n, \widehat{\theta})\pi_1(\widehat{\theta})}{\Pi_{n=1}^t g(y_n|a', \theta) f_n(b_n, \theta)\pi_1(\theta)}.$$

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

$$\alpha_t(j) = \ln\left\{\frac{m_t^*(a', \widehat{\theta})^j(1 - m_t^*(a, \widehat{\theta}))^{1-j}}{m_t^*(a', \theta)^j(1 - m_t^*(a', \theta))^{1-j}}\right\}, \quad \xi_t = \ln\left\{\frac{g(y_t|a', \widehat{\theta})}{g(y_t|a', \theta)}\right\}, \text{ and } \quad \delta = \ln\left\{\frac{\pi_1(\widehat{\theta})}{\pi_1(\theta)}\right\}.$$

From now on, all almost sure statements are with respect to $\gamma_{\widehat{\theta}}$. There are two cases to consider. When θ is such that $\theta_a \neq \widehat{\theta}_a$, Lemma 2 implies that $\pi(\theta|\pi_1, y_1, b_1, \ldots)$ converges to zero almost surely. Suppose then that $\theta \in \Theta' = \{\theta \in \Theta : \theta_a = \widehat{\theta}_a \text{ and } \theta \in \Theta_{\widetilde{q}}\}$. Notice that $\xi_t \equiv 0$ in this case. By assumption, $\lim_t m_t^*(a', \widehat{\theta}) = m^* > 0$. Since $\mathrm{Var}_{\chi_\theta}(b_t) \leq 1/4$ and $\frac{1}{t}\sum_{n=1}^t m_n^*(a', \widehat{\theta}) \to m^* > 0$, Kolmogorov's SLLN implies that

$$\frac{1}{t}\sum_{n=t}^t b_n = \frac{1}{t}\sum_{n=t}^t (b_n - m_n^*(a', \widehat{\theta})) + \frac{1}{t}\sum_{n=1}^t m_n^*(a', \widehat{\theta}) \to \alpha > 0$$

almost surely. Now observe that $1 - m_t^*(a', \widehat{\theta}) \geq m_t^*(a'', \widehat{\theta})$ has a positive limit by assumption, while $m_t^*(a', \theta)(1 - m_t^*(a, \theta))$ converges to zero by the induction hypothesis. Hence, $\{\alpha_t(1)\}$ is a bounded sequence and $\alpha_t(0)$ diverges to infinity. This implies that $t^{-1} \ln L_t \to \infty$ almost surely, so that $L_t \to \infty$ almost surely as well. Therefore, $\pi(\theta|\pi_1, y_1, b_1, \ldots)$ converges to zero almost surely, which proves the claim. $\square$

Appendix D: Existence

1. The set $\mathfrak{S}_S$.

Let $\mathrm{L}_\infty(\mu, \mathbb{R}^s)$, with $s \in \mathbb{N}$, be the set of bounded functions from Ω into $\mathbb{R}^s$ endowed with the norm $||F||_\infty = \sup_{\omega\in\Omega} ||F(\omega)||$, where $||\cdot||$ is the Euclidian norm. It is well–known that the dual of $\mathrm{L}_1(\mu, \Gamma_t)$ is $\mathrm{L}_\infty(\mu, \Gamma_t)$. Notice that if $F_t \in \mathrm{L}_1(\mu, \Sigma_t)$, then $||F_t||_\infty \leq \sqrt{|H_t|/K}$; that is, $\mathrm{L}_1(\mu, \Sigma_t)$ is a norm bounded subset of $\mathrm{L}_\infty(\mu, \Gamma_t)$. Hence, $\mathrm{L}_1(\mu, \Sigma_t)$ is a weakly compact subset of $\mathrm{L}_1(\mu, \Gamma_t)$ by the Eberlein–Šmulian Theorem and Theorem 9 in page 292 of Dunford and Schwartz (1988). Thus, $\mathrm{L}_1(\mu, \Sigma_t)$ is norm compact by Corollary 13 in page 295 of Dunford and Schwartz (1988). The assumption that the outcome space Y is finite is crucial for this result. From now on we take the sets $\mathrm{L}_1(\mu, \Sigma_t)$ to be endowed with the norm topology. The next result follows from Tychonoff's Theorem and the fact the countable product of metric spaces is metrizable.

Lemma 9. *$\mathfrak{S}_S$ is a compact and metrizable.*

2. The map M_S.

In what follows we make use of the following result, which is straightforward to prove.

Lemma 10. *Let $A \subset \mathrm{L}_1(\mu, \mathbb{R}^s)$ and $A' \subset \mathrm{L}_1(\mu)$. Define $Q_k : A \times A' \to \mathbb{R}$ to be such that $Q_k(F, \tau) = \int \langle F(\omega), e_k\rangle \tau(\omega)\mu(d\omega)$. The map Q_k is jointly norm continuous if A is a norm bounded subset of $\mathrm{L}_\infty(\mu, \mathbb{R}^s)$ and A' is a norm bounded subset of $\mathrm{L}_\infty(\mu)$.*

Lemma 11. *M_S is continuous and affine.*

Proof: We first prove that M_S is continuous. For this, let $\mathsf{M}_{S,t}$ be the period–t component of M_S. By definition, $\mathsf{M}_{S,t}(F)$ is the period–t observation likelihood when the strategy profile is $F \in \mathfrak{S}_S$. We are done if we show that for each $t \geq 1$, $F \mapsto m_{S,t}(a, \theta, F) = \mathsf{M}_{S,t}(F)(a, \theta)$ is a continuous map from $\mathfrak{S}_S$ into $[0, 1]$ for all $\theta \in \Theta$ and $a \in A$. The proof is by induction.

We start with some notation. For each $F_t \in \mathrm{L}_1(\mu, \Sigma_t)$ and $h^t \in H_t$, let $F_{t,h^t} \in \mathrm{L}_1(\mu, \mathbb{R}^K)$ be such that $F_{t,h^t}(\omega) = F_t(\omega)(h^t)$. By construction, if $F = \{F_t\} \in \mathfrak{S}_S$, then $F_{t,h^t}(\omega)$ is the mixed action that a player with prior π_ω chooses in period t after the history h^t. Moreover, let $k(a)$ be the label of action a; that is, $a_{k(a)} = a$.

Notice that $m_{S,1}(a,\theta,F) = \int \langle F_{1,\emptyset}(\omega), e_{k(a)}\rangle \mu(d\omega)$, and so $m_{S,1}$ is continuous in F for each $a \in A$ and $\theta \in \Theta$ by Lemma 10. Now let $\widehat{\tau}_2 : \Omega \times H_2 \times \Theta \times \mathfrak{S}_S \to [0,1]$ be such that if $h^2 = (a, y, a')$, then $\widehat{\tau}_2(\omega, h^2, \theta, F) = \langle F_{\emptyset,1}(\omega), e_{k(a)}\rangle g(y|a,\theta) m_{S,1}(a', \theta, F)$. By construction, $\widehat{\tau}_2(\omega, h^2, \theta, F)$ is the probability that a player with prior belief π_ω experiences $h^2 \in H_2$ when the state of the world is θ and the strategy profile under play is F. Now define $\tau_2 : H_2 \times \Theta \times \mathfrak{S}_S \to \mathrm{L}_1(\mu)$ to be such that $\tau_2(h^2, \theta, F)(\omega) = \widehat{\tau}_2(\omega, h^2, \theta, F)$. We claim that τ_2 is (norm) continuous in F for each $\theta \in \Theta$ and $h^2 \in H_2$. For this, let $\{F^n\}$ be a sequence in $\mathfrak{S}_S$ with limit F and suppose that $h^2 = (a, y, a')$. Then,

$$\begin{aligned} ||\tau_2(h^2, \theta, F^n) - \tau_2(h^2, \theta, F)||_1 &= \int |\widehat{\tau}_2(h^2, \omega, \theta, F^n) - \widehat{\tau}_2(h^2, \omega, \theta, F)(h^2)|\mu(d\omega) \\ &\leq ||F^n - F||_1 + ||F||_1 \cdot |m_{S,1}(a', \theta, F^n) - m_{S,1}(a', \theta, F)|, \end{aligned}$$

which implies the desired result.

Suppose then, by induction, that for some $t \geq 1$ there exist: (i) A continuous map $\tau_{t+1} : H_{t+1} \times \Theta \times \mathfrak{S}_S \to \mathrm{L}_1(\mu)$ such that $\tau_{t+1}(h^{t+1}, \theta, F)(\omega)$ is the probability that a player with prior π_ω experiences $h^{t+1} \in H_{t+1}$ when the state of the world is θ and the strategy profile under play is F; (ii) A continuous map $m_{S,t} : A \times \Theta \times \mathfrak{S}_S \to [0,1]$ where $m_{S,t}(a, \theta, F)$ is the fraction of players who choose $a \in A$ in period t when the state of the world is θ and the strategy profile is F. Now let $m_{S,t+1} : A \times \Theta \times \mathfrak{S}_S \to [0,1]$ be given by

$$m_{S,t+1}(a, \theta, F) = \sum_{h^{t+1} \in H_{t+1}} \int \langle F_{h^{t+1},t+1}(\omega), e_{k(a)}\rangle \tau_{t+1}(h^{t+1}, \theta, F)(\omega)\mu(d\omega).$$

By construction, $m_{S,t+1}(a, \theta, F)$ is the mass of players who choose a in $t+1$ when the state of the world is θ and the strategy profile under play is F. Lemma 10 and the continuity of τ_{t+1} imply that $m_{S,t+1}$ is continuous in F for each $a \in A$ and $\theta \in \Theta$. Now define $\widehat{\tau}_{t+2} : \Omega \times H_{t+2} \times \Theta \times \mathfrak{S}_S \to [0,1]$ to be such that if $h^{t+2} = (h^{t+1}, a_{t+1}, y_{t+1}, a'_{t+1})$, then

$$\widehat{\tau}_{t+2}(\omega, h^{t+2}, \theta, F) = \tau_{t+1}(h^{t+1}, \theta, F)(\omega)\langle F_{h^{t+1},t+1}(\omega), e_{k(a_{t+1})}\rangle g(y_{t+1}|a_{t+1}, \theta) m_{S,t+1}(a'_{t+1}, \theta, F),$$

and let $\tau_{t+2} : H_{t+2} \times \Theta \times \mathfrak{S}_S \to \mathrm{L}_1(\mu)$ be such that $\tau_{t+2}(h^{t+2}, \theta, F)(\omega) = \widehat{\tau}_{t+2}(\omega, h^{t+2}, \theta, F)$. By construction, $\tau_{t+2}(h^{t+2}, \theta, F)(\omega)$ is the probability that a player with prior π_ω experiences $h^{t+2} \in H_{t+2}$ when the state of the world is θ and the strategy profile under play is F. An argument similar to the one used to prove that τ_2 is continuous shows that τ_{t+2} is continuous.

Thus, M_S is continuous by induction. The fact that M_S is affine follows immediately from its construction. $\square$

3. Best–responses.

Recall that (i) $R = \sum_{t=1}^{\infty} \delta^{t-1} r(a_t, y_t)$, where a_t and y_t are the period–t action choice and outcome, respectively; (ii) $\mu_\theta(\sigma, m)$ is the probability distribution on H_∞ induced by the behavior strategy σ when the sequence of observation likelihoods is m and the state of the world is θ.

Lemma 12. *$V_\theta(\sigma, m) = \mathrm{E}_{\mu_\theta(\sigma,m)}[R]$ is jointly continuous in σ and m for each $\theta \in \Theta$.*

Proof: Let $\underline{r} = \min_{a,\theta} r(a, \theta)$ and $\overline{r} = \max_{a,\theta} r(a, \theta)$. Now let $\underline{R}_n = \sum_{t=1}^{n} \delta^{t-1} \underline{r}$ and $R_n = \sum_{t=1}^{n} \delta^{t-1} r(a_t, y_t)$, where $n \in \mathbb{N} \cup \{\infty\}$. Then, if $V_{n,\theta}(\sigma, m) = E_{\mu_\theta(\sigma,m)}[R_n]$,

$$V_\theta(\sigma, m) - V_{n,\theta}(\sigma, m) = E_{\mu_\theta(\sigma,m)}[(R - \underline{R}_\infty) - (R_n - \underline{R}_n)] + \underline{R}_\infty - \underline{R}_n.$$

Since $(R_\infty - \underline{R}_\infty) - (R_n - \underline{R}_n) = \sum_{t=n+1}^{\infty} \delta^{t-1}[r(a_t, y_t) - \underline{r}] \geq 0$, we then have that

$$\begin{aligned} |V_\theta(\sigma, m) - V_{n,\theta}(\sigma, m)| &\leq E_{\mu_\theta(\sigma,m)}[R_\infty - R_n] - (\underline{R}_\infty - \underline{R}_n) + |\underline{R}_\infty - \underline{R}_n| \\ &\leq \sum_{t=n+1}^{\infty} \delta^{t-1}\overline{r} - (\underline{R}_\infty - \underline{R}_n) + |\underline{R}_\infty - \underline{R}_n|. \end{aligned}$$

Thus, $V_{n,\theta}(\sigma, m)$ converges to $V_\theta(\sigma, m)$ uniformly in σ and m. The desired result now follows from the fact that $V_{n,\theta}(\sigma, m)$ is jointly continuous in σ and m for each $n \geq 1$ and $\theta \in \Theta$. $\square$

Let $V(\sigma, m, \pi) = \sum_{\theta \in \Theta} \pi(\theta) V_\theta(\sigma, m)$ and $V^*(m, \pi) = \max_{\sigma \in \Sigma} V(\sigma, m, \pi)$. Note that V^* is well–defined since the set of behavior strategies is compact by Tychonoff's Theorem. Now let $\mathrm{BR}_\omega \rightrightarrows \Sigma$, with $\omega \in \Omega$, be such that $\mathrm{BR}_\omega(m) = \{\sigma \in \Sigma : V(\sigma, m, \pi_\omega) = V^*(m, \pi_\omega)\}$. By construction, $\mathrm{BR}_\omega(m)$ is the set of all optimal experimentation strategies for $\mathrm{ILP}(\pi_\omega, m)$.

Lemma 12 and the Maximum Theorem imply that BR_ω has non–empty values and is upper hemicontinuous for all $\omega \in \Omega$. It is easy to see that each BR_ω is also convex–valued. To finish, let $\mathrm{BR} : \mathfrak{M} \rightrightarrows \mathfrak{S}_S$ be given by $\mathrm{BR}(m) = \{F \in \mathfrak{S}_S : F(\omega) \in \mathrm{BR}_\omega(m) \text{ for all } \omega \in \Omega\}$.

Lemma 13. BR *is upper hemicontinuous with non–empty and convex values.*

Proof: It is immediate to see that BR has non–empty and convex values. Since $\mathfrak{S}_S$ is compact, BR is upper hemicontinuous if it has a closed graph. Let $\{m^k\}$ be a convergent sequence in $\mathfrak{M}$ with limit m and $\{F^k\}$ be a convergent sequence in $\mathfrak{S}_S$ with limit F and such that $F^k \in \mathrm{BR}(m^k)$ for each $k \in \mathbb{N}$. Since all elements of Ω have positive μ–measure, $\{F^k(\omega)\}$ converges to $F(\omega)$ in Σ for all $\omega \in \Omega$. Now observe that $F^k(\omega) \in \mathrm{BR}_\omega(m^k)$ by construction. Hence, since each BR_ω is upper hemicontinuous, $F(\omega) \in \mathrm{BR}_\omega(m)$ for all $\omega \in \Omega$. Therefore, $F \in \mathrm{BR}(m)$, and so BR is upper hemicontinuous. □

Appendix E: Extension (Not for Publication)

Here we provide a rough sketch of the argument that introducing a neighborhood structure (and keeping all else the same) does not change our results. We simplify the exposition by considering the case where each player has one neighbor only. In this case, a neighborhood structure can be described by a map $\mathrm{Q} : [0,1] \to [0,1]$ such that $\mathrm{Q}(i) \neq i$ and $\mathrm{Q}(\mathrm{Q}(i)) = i$ for almost every player i; $\mathrm{Q}(i)$ is the neighbor of i. The analysis can be extended to the case where for each player i the set of players in his network (the neighbors of his neighbors, and so on) is finite. The restriction to finite networks is due to existence problems.

As in the main text, to each strategy profile F there is associated a sequence $\{m_t\}$ such that $m_t(a, \theta)$ is the fraction of players who choose a in period t when the state of the world is θ if F is under play. Moreover, to each profile F there is also associated a sequence $\{\ell_t\}$, with $\ell_t : [0,1] \times A \times \Theta \to [0,1]$, such that $\ell_t(i, a, \theta)$ is the probability that i's neighbor chooses a in period t when the state of the world is θ if F is under play; $\ell_t(i, \cdot)$ is player i's period–t neighborhood likelihood. A player also uses his neighborhood likelihoods to update beliefs.

Let $\mathfrak{L}^*$ be the set of sequences of neighborhood likelihoods. An equilibrium is a triple (F^*, m^*, ℓ^*), where ℓ^* is a map from $[0,1]$ into $\mathfrak{L}^*$ such that: (i) $F^*(i)$ solves the individual learning problem with sequence m^* of observation likelihoods and sequence $\ell^*(i)$ of neighborhood likelihoods for almost all i; (ii) when F is under play, m^* is the sequence of observation likelihoods and $\ell^*(i)$ is the sequence of neighborhood likelihoods of player i for almost all i.

The fact that a player now observes the action choices of his neighbor does not change the fact that he learns the true type of any action he chooses infinitely often. So, Lemma 2 still holds. Theorems 2.1 and 2.4 in Rosenberg et al. (2006) still apply, and so the remaining results in Section 5 also hold in this setting. Thus, the same argument used in Section 6 shows that there is long–run social conformity in equilibrium; and the same argument used in Section 7 shows that social learning is complete in every equilibrium if (H) is satisfied. The key point is that even if the behavior of a player is not independent of the behavior of his neighbor, there is still no aggregate uncertainty, as the behavior of almost every pair of players is independent.