Diversity Fosters Learning in Environments with Experimentation and Social Learning

Cunha, Douglas and Monte, Daniel

18 April 2023

Diversity Fosters Learning in Environments with Experimentation and Social Learning*

Douglas Cunha   Daniel Monte

April 18, 2023

Abstract

We study long-lived rational agents who learn through experimentation and observing each other's actions. Experimentation and social learning, even when combined, often lead to learning failures as agents may stop experimenting due to the Rothschild effect or social conformity. We show that when there is diversity in preferences, there will be complete learning in the limit, thereby overcoming these learning failures. Our analysis demonstrates the critical interaction between experimentation, social learning, and diversity and provides a new rationale for the increasingly held view that diversity is crucial in institutions.

Keywords: Two-armed bandit; Social learning; Heterogeneous players

JEL Codes: C73, D82, D83

*Cunha: Takhe. Monte: ESOMAS Department, University of Turin, Collegio Carlo Alberto, and Sao Paulo School of Economics - FGV. E-mail: daniel.monte@unito.it. We are very thankful for conversations with Braz Camargo, Eduardo Faingold, Bruno Ferman, Ignacio Monzon, and Felipe Shalders.

1 Introduction

Agents may attempt to learn the true state of the world by experimentation and by observing the choices of others. When agents learn only from the outcomes of their own choices, as in a single agent multi-armed bandit, complete learning often fails. With positive probability, agents settle on a suboptimal choice (Rothschild [1974]). Furthermore, learning exclusively from observing the actions of others may result in informational cascades, leading to further flawed learning outcomes (Banerjee [1992] and Bikhchandani et al. [1992]).[1]

A conjecture once held was that allowing agents to learn both from their own actions and the actions of others would be sufficient for complete learning. The failure of complete learning in a single-agent environment with experimentation is because the agent eventually settles for a single choice. Once she does, she no longer learns about the other possible choices (Rothschild [1974]). However, in environments with both experimentation and social learning, agents are simultaneously learning from their past experiences and, by observing others' choices, inferring something from the actions they do not take. Thus, one might expect social learning to help overcome the so-called Rothschild effect. However, Aoyagi [1998], and later Camargo [2014], showed a surprising no-learning result: complete learning fails even in a large pool of agents learning from their own previous experiences and the actions of others. The main intuition for this failure of complete learning is that social learning leads to social conformity; agents eventually settle for the same action, and thus, the learning-from-others component halts. Importantly, the action is often not the optimal one.

We are interested in understanding how diversity may affect learning in an environment where agents learn from their own past experiences and the actions of others. Specifically, by diversity, we mean a pool of agents with sufficiently rich heterogeneity in preferences. In environments with diversity, agents look at the actions of others with the understanding that others are choosing actions by balancing their private preferences with the information they

[1]Smith and Sørensen [2000] show that suboptimal herds may persist indefinitely if and only if signals are bounded.

have accumulated about the different options. Diversity might deter social conformity since individuals with different preferences are less likely to settle on the same choice. However, diversity also means that the information externality of each agent's choice is small: an agent might have taken an action because she "likes it," irrespective of the information she has received about it.[2]

We have a continuum of rational long-lived agents who play a two-armed bandit. At every period, they simultaneously choose an arm after observing a random sample of past play and their own previous history of actions and outcomes. Analyzing learning in our environment is not straightforward since beliefs are private due to each agent's private experimentation and their private history of observation of others. Moreover, our diversity assumption implies that each agent's choice balances the idiosyncratic preference of this agent and her private belief about the expected payoffs of all actions. Thus, it becomes difficult to infer anything meaningful from any single observation from others. Nevertheless, we prove a learning result by studying the long run information aggregation.

Our main result is to show that in environments with experimentation and social learning, diversity implies that complete learning happens with probability one in the long run. The intuition is two-fold. First, agents eventually learn about the choice they settle for, a direct consequence of the law of large numbers. The second part of the intuition is an equilibrium argument. In the long run, after observing a sufficiently long sequence of others' choices, every agent will form a belief about the distribution of choices in the economy. Agents know how preferences are distributed across the pool of agents, and they know that each agent has learned about the expected payoff of the choices they have settled for. The most detailed part of the argument (our Lemma 8, together with an induction argument) is to use the results mentioned above to prove that each state of the world must induce a different asymptotic distribution of choices. Consequently, any agent that has settled for a wrong choice is likely

[2]As an illustration, consider the paper by Munshi [2004] who uses data from the Indian Green Revolution to show that rice growers respond less to neighbors' experience than wheat growers since rice-growing regions are more heterogeneous in growing conditions and rice varieties are more sensitive to unobserved farm characteristics.

to revise it based on her observed sequence of actions from others, which uniquely pins down the underlying true state of the world in the long run. Only a small fraction of agents will choose suboptimal actions, and complete learning happens with probability one in the limit.

Our result, combined with other existing results in the literature, delivers the following message: Diversity is critical for complete learning. Commonly cited reasons for pursuing diversity in institutions include the fact that diversity boosts creativity, helps overcome behavioral biases, and increases the pool of talents. Moreover, exposure to diverse groups reduces bias toward ethnic minorities and plays an important role in fostering tolerance (See Boisjoly et al. [2006]; Carrell et al. [2019]). Our result provides another rationale for the increasingly held view that diversity is crucial in institutions.[3]

Ours is the first paper that shows that diversity in preferences will lead to complete learning in an environment that combines experimentation with social learning. In pure social learning environments, diversity has been shown to lead to learning in some cases but not others. On the one hand, diversity can foster learning by breaking informational cascades, an intuition similar to the case of homogeneous preferences but unbounded signals (Smith and Sørensen [2000]). In a sufficiently rich private values environment, all actions are played infinitely often since a player with a strong preference for an action may disregard the public belief when choosing (see Goeree et al. [2006] and Wiseman [2008]). This intuition does not carry over in a direct way to environments such as ours. In our environment, the fact that all actions are taken in all states of the world is not sufficient for complete learning: indeed, in all states, all actions will be taken infinitely often. On the other hand, diversity in pure social learning environments can also be detrimental to learning—for example, a positive fraction of players that disregard public beliefs when acting could preclude information aggregation. Indeed, Monzon [2019] provides a simple example in which even a small fraction of extreme types (which may be interpreted as agents that strictly prefer an action over another, regardless of

[3]As an example, universities are actively pursuing diversity goals. For example, Harvard's diversity and inclusion commitment message includes: "Harvard's commitment to diversity in all forms is rooted in our fundamental belief that engaging with unfamiliar ideas, perspectives, cultures, and people creates the conditions for dramatic and meaningful growth."

the information received) can lead to significant deviations from complete learning. In our case, the inference from observing others' choices is always taken with caution: it is obfuscated by the fact that others might be experimenting, making it difficult to draw a meaningful conclusion from a single observation. It is, instead, the aggregation of the observation of others that is meaningful in our environment.

1.1 Related Literature

This paper contributes to the literature on social learning and experimentation. Our approach extends that of Aoyagi [1998] and Camargo [2014]. Aoyagi [1998] studies a two-armed bandit model with finite homogeneous players who observe each other's actions in every period. Camargo [2014] considers a multi-armed bandit model with a continuum of homogeneous players that observe, in each period, the action of another randomly chosen player, which he refers to as *observation in society*. Both of them show that, in equilibrium, all players eventually settle on the same alternative, although not necessarily on the best one. That is, social learning is not enough to overcome the so called "Rothschild effect",[4] when players completely drop the best alternative in the long run. Camargo [2014] goes on to derive a sufficient condition on the distribution of prior beliefs that prevents this result.

Other papers studied the importance of heterogeneity in pure social learning environments. Goeree et al. [2006] and Wiseman [2008] consider a continuum of agents' types, with continuous possible payoffs for each action that depend on agent's type. Short-lived rational individuals decide sequentially, observe the choices of previous agents and receive a private signal about the state of the world.[5] They show that society completely learns the true state of the world in the long run. With enough heterogeneity, there is always a type of agent whose decision will depend on her private signal. Our approach complements that of Goeree et al. [2006]

[4]Due to the work of Rothschild [1974].

[5]Related to the case with homogeneous players, introduced by Banerjee [1992] and Bikhchandani et al. [1992]. In a recent paper, Kartik et al. [2022] studies a general model of sequential learning over social networks, and the interplay of preferences and information. They provide necessary and sufficient conditions for learning.

and Wiseman [2008] since we consider long-lived agents, who choose optimal experimentation strategies. A key difference is that, in our environment, players' private signals depend on their own actions, and each player observes a different history of others' actions. Furthermore, we consider states with two independent components, which makes an action to be uninformative about the other.

Social learning with heterogeneous agents has been studied in other different frameworks. Ellison and Fudenberg [1993] consider, for example, that agents are heterogeneous in the sense that each one has a different (real-valued) parameter. In each period, agents can choose a technology and observe choice and payoff from those with parameters close to theirs. This parameter also affects payoffs, which makes the best technology not to be the same for everyone. Among other results, Ellison and Fudenberg [1993] show that, in general, agents do not eventually take the best alternative for themselves. A key difference from our paper is that the learning process in their environment do not follow Bayes' rule, but an exogenous rule that do not consider all the past experiences. On the other hand, Bala and Goyal [2001] consider agents that update their beliefs with Bayes' rule. One difference from our paper is that agents are myopic in their environment. Bala and Goyal [2001] consider a network in which agents can observe the actions and outcomes of the ones connected to them. Heterogeneity comes from the fact that there are two types of agents, with different preferences. Agents eventually choose the best action for themselves depending on the connections they have in the network. Ali [2018] studies a model of social learning with information acquisition and heterogeneity in the cost of acquiring information. His main message is that there will be herding only on correct actions if and only if for every interior public belief, players might have strict incentives to acquire information that can overturn the belief.

Harel et al. [2021] study a model of long-lived rational agents learning form each other's actions and from private signals. In their paper, there is no experimentation (signals are independent of actions), so their focus is to study the speed of learning from others' actions.

2 Model

A continuum of anonymous players is identified with a probability space $(I, \mathcal{I}, \lambda)$. Where I is the set of players, $\mathcal{I}$ is a σ-algebra on I, and λ is a probability measure on $\mathcal{I}$. These players are playing a two-armed bandit: they choose between two arms with stochastic payoffs, but fixed expected payoffs. That is, in every period $t \in \mathbb{N}$, each agent has to choose one action in the set $A = \{1, 2\}$. The set of states of the world is $\Theta = \Theta_1 \times \Theta_2$, where, for each arm $k \in A$, $\Theta_k = \{\theta_k^1, \theta_k^2, ..., \theta_k^{n_k}\}$ for some $n_k \in \mathbb{N}$. We define a linear order on Θ_k such that $\theta_k^1 < \theta_k^2 < ... < \theta_k^{n_k}$ and a partial order on Θ such that $\theta \geq \theta'$ if, and only if, $\theta_1 \geq \theta_1'$ and $\theta_2 \geq \theta_2'$. We say that $\theta > \theta'$ if at least one of the two inequalities is strict.

For each agent $i \in I$, let Y^i denote the set of her finitely many possible payoffs.[6] She gets payoff y with probability $g^i(y|k, \theta_k)$ when she chooses action k and the state of the world is $\theta = (\theta_1, \theta_2)$. Players are heterogeneous since Y^i and g^i may be different for each $i \in I$. Player i's expected payoff is denoted by $r_k^i(\theta_k)$, for each $k \in A$ and $\theta_k \in \Theta_k$. We assume that, for each state $\theta \in \Theta$,

$$\forall \varepsilon > 0, \lambda\{0 < r_k^i(\theta_k) - r_{k'}^i(\theta_{k'}) < \varepsilon\} > 0, \text{ for } k, k' \in A \text{ s.t } k \neq k', \tag{A1}$$

and that

$$\lambda\{r_1^i(\theta_1) = r_2^i(\theta_2)\} = 0. \tag{A2}$$

For each state $\theta \in \Theta$, assumptions A1 and A2 imply that, under complete information, there would be a strict positive mass of players arbitrarily close to the indifference between the two actions but no mass of players exactly in the indifference. Assumption A1 is important to make asymptotic distributions different depending on the state of the world, and assumption A2 is important for the convergence of the observation likelihood (Lemma 4).

[6] Although we could consider infinitely many different payoffs without changing the results, we chose finitely many payoffs for ease of notation.

We also assume that r_k^i is strictly increasing in θ_k. Stronger than that, we consider, for each $k \in A$,

$$\inf\{r_k^i(\theta_k) - r_k^i(\theta'_k) : \theta_k, \theta'_k \in \Theta_k \text{ s.t. } \theta_k > \theta'_k, i \in I\} > 0. \tag{A3}$$

Assumption A3 implies that agents are not arbitrarily close to the indifference between states θ and θ' with $\theta_k \neq \theta'_k$ when they choose $k \in A$.[7]

Let $\Pi = \Delta(\Theta)$ be the set of possible beliefs about the state of the world. Let us consider a common prior $\pi_1 \in \Pi$. Assume that $\pi_1(\theta) > 0,\ \forall \theta \in \Theta$.[8] Denote by $\pi_t^i(\theta)$ the probability player i assigns to $\theta \in \Theta$ in period t and $\pi_t^i(k, \theta_k) := \sum_{\{\theta' : \theta'_k = \theta_k\}} \pi_t^i(\theta')$ the probability she assigns to $\theta_k \in \Theta_k$.

In a given period t, a player i chooses an action $k \in A$, observes an outcome $y \in Y^i$ and, a choice $\tilde{k} \in A$ of another randomly chosen anonymous player. The set of histories in period t is $H_t^i = (A \times Y^i \times A)^{t-1}$. The set of infinite histories is $H_\infty^i = (A \times Y^i \times A)^\infty$. A strategy for player i is a sequence $\sigma = \{\sigma_t\}$ such that $\sigma_t : H_t^i \to \Delta(A)$ maps any history in H_t^i to an action (possibly mixed) in period t. Let Σ^i be the set of all possible strategies for player $i \in I$ and define $\Sigma := \cup_{i \in I} \Sigma^i$. A strategy profile $F : I \to \Sigma$ is a $\mathcal{I}$-measurable function[9] that maps each player $i \in I$ to a strategy $F(i) \in \Sigma^i$. The set of all possible strategy profiles is denoted by $\mathcal{F}$.

Given a strategy profile F, the proportion of players choosing each action $k \in A$ at period

[7]Assuming only that r_k^i is strictly increasing for each $k \in A$ and $i \in I$ is not enough. To see that, consider states θ and θ' such that $\theta_1 > \theta'_1$ and $\theta_2 = \theta'_2$. Consider that either $r_1^i(\theta'_1) - r_2^i(\theta'_2) < r_1^i(\theta_1) - r_2^i(\theta_2) < 0$ or $0 < r_1^i(\theta'_1) - r_2^i(\theta'_2) < r_1^i(\theta_1) - r_2^i(\theta_2)$ for each $i \in I$. This means that agents would like to make exactly the same choices in both states, without breaking strict monotonicity of r_k^i. Since different asymptotic distributions of actions are important in our model, we rule out this type of situation. We illustrate the importance of this assumption in figures 2 and 3.

[8]We could consider an economy with heterogeneous prior and all our results would follow. For that, let $\phi : I \to \Pi$ be such that $\phi(i) \in \Pi$ is the initial prior of player $i \in I$, which considers that state components are independent. Assume that players do not assign a probability arbitrarily close to 0 to any state in the beginning of the game:

$$\eta := \inf\{\pi_1^i(\theta) : i \in I, \theta \in \Theta\} > 0. \tag{A4}$$

[9]F must be a $\mathcal{I}$-measurable function so we can aggregate individual actions.

$t \in \mathbb{N}$ when the state of the world is $\theta \in \Theta$ is well defined. We denote it by $m_t(k,\theta)$ and define $m_t(\theta) := (m_t(1,\theta), m_t(2,\theta))$.[10] Let $\mathcal{M}$ be the set of all possible sequences $m = \{m_t\}$ from $A \times \Theta$ into $[0,1]$. We denote by $M : \mathcal{F} \to \mathcal{M}$ the map such that $m = M(F)$ gives us the proportion of players choosing each action for each state of the world and period when the strategy profile is F.

Fix a player $i \in I$ and her strategy $\sigma \in \Sigma^i$. Her (possibly mixed) action is defined by strategy σ in the first period, θ defines the outcome distribution $g^i(y|k,\theta)$, and $m_1(k,\theta)$ the probability of observing a player who plays k in period $t = 1$. Hence, σ, θ and m define a probability distribution on H_1. For each $h_1 \in H_1$, σ defines the player's action in the second period, and so on. Therefore, σ, θ and m define a probability distribution on H_∞, which we denote by $\mu(\sigma|\theta,m)$. Since i does not know θ, she considers the prior $\pi_1 \in \Pi$ with $\mu(\sigma|\theta,m)$ to find a distribution probability on $\Theta \times H_\infty$, which we denote by $\mu(\sigma|\pi,m)$.

Let y_t^i be player i's stochastic payoff at period t and denote by $R^i = \sum_{t=1}^{\infty} \delta^{t-1} y_t^i$ the expected sum of payoffs, where $\delta \in [0,1)$ is the discount factor. Given m, we denote the individual learning problem of player i under prior π_1 by $ILP^i(\pi_1,m)$. An optimal experimentation strategy σ^* for $ILP^i(\pi_1,m)$ is such that

$$\mathbb{E}_{\mu(\sigma^*|\pi_1,m)}[R^i] = \sup_{\sigma \in \Sigma^i} \mathbb{E}_{\mu(\sigma|\pi_1,m)}[R^i]. \tag{1}$$

Hence, the observation likelihood m affects optimal strategies and a strategy profile F defines the observation likelihood through $m = M(F)$. This is the idea behind the Nash equilibrium for non-atomic games that we adapt to our environment.

Definition 1 (Equilibrium)**.** *An equilibrium is a pair (m^*, F^*) such that $F^*(i)$ is an optimal experimentation strategy for $ILP^i(\pi_1, m^*)$ for λ-almost all $i \in I$ and $m^* = M(F^*)$.*

[10]Although individual behavior may be stochastic, aggregate is not. Appendix A.2 of Camargo [2014] shows how to aggregate individual behavior to find $\{m_t\}$.

3 Results

In this section we present our main result: there will be complete learning in the limit. We construct this result by presenting 8 lemmas. The first four lemmas are adapted from Aoyagi [1998], Rosenberg et al. [2009] and Camargo [2014] and we simply refer to the original work for the proofs, while the remaining four lemmas are specific to our environment and we present their proofs.

Consider a player i who chooses k infinitely many times. She will have infinitely many outcome observations and, by the Strong Law of Large Numbers (and assumption A3), will asymptotically learn the true expected payoff of action k with probability 1. This is what Lemma 1 below states. The proof is omitted since it is essentially that of Lemma 1 of Aoyagi [1998].

Let $I_k \in I$ be the set of players who play $k \in A$ infinitely many times. Let E_k^i be the event such that $i \in I_k$.

Lemma 1 (Learning about your choice)**.** *Assume θ is the true state of the world. Suppose the sequence of observation likelihoods is m. Consider a player $i \in I$, prior π_1 and strategy σ^i. If $\mu(\sigma^i|\pi_1, m)(E_k^i) > 0$, then $\mu(\sigma^i|\pi_1, m)\left(\lim_{t\to\infty} \pi_t^i(k, \theta_k) = 1|E_k^i\right) = 1$.*

Our next result states that almost all players will eventually choose actions that are myopically optimal according to their beliefs. Intuitively, players do not want to experiment forever, since they experiment when they want to acquire information for the future, giving up current expected payoff. This result is a consequence of Proposition 2.1 of Rosenberg et al. [2009], and we state without proof.

Before stating Lemma 2, we need to establish some notation. Consider a player i with belief π. Let $\mathbb{E}_\pi[r_k^i] := \sum_{j=1}^{n_k} \pi(k, \theta_k^j) r_k^i(\theta_k^j)$ be her current expected outcome when she plays k and define $BR^i(\pi)$ the set of her myopically optimal actions, that is, $k \in BR^i(\pi) \Leftrightarrow k \in \arg\max_k \mathbb{E}_\pi[r_k^i]$. Since beliefs $\{\pi_t^i\}_{t=1}^\infty$ are martingales, they converge almost surely to a (random) limit π_∞^i. Let A_∞^i be the (random) set of actions player i chooses infinitely many times.

Lemma 2 (Myopic optimal actions in the limit). *Suppose the sequence of observation likelihoods is m and σ^i is an optimal strategy. Then $\mu(\sigma^i|\pi^i_1, m)(A^i_\infty \subseteq BR^i(\pi^i_\infty)) = 1$ for each player $i \in I$.*

Consider a player i who plays both actions infinitely many times, that is, $i \in I_1 \cap I_2$. Lemma 1 implies that she asymptotically learns the true state of the world. Lemma 2 requires that $\{1, 2\} \subseteq BR(\pi^i_\infty)$ and, therefore, $r^i_1(\theta_1) = r^i_2(\theta_2)$ with probability 1. However, this is ruled out by assumption A2. Thus, we can state our next result.

Lemma 3 (Settle on one choice). *Suppose almost all players follow optimal strategies. Then the mass of players that choose two actions infinitely many times is zero.*

If the mass of players who choose $k = 1$ did not converge, there would be a positive mass of players alternating their choice infinitely many times, which cannot occur, according to Lemma 3. This result is our Lemma 4, which comes from Lemma 6 of Camargo [2014] and, therefore, we state without proof.

Lemma 4 (Long-run proportion is convergent). *Let (m^*, F^*) be an equilibrium. Then $\{m^*_t(\theta)\}$ is convergent for all $\theta \in \Theta$.*

We define $m_\infty(\theta)$, for each $\theta \in \Theta$, such that $m_t(\theta) \to m_\infty(\theta)$.

Now assume θ is the true state of the world. Then Lemma 4 asserts that the fraction of players choosing action k converges to $m_\infty(k, \theta)$. Since players make infinitely many observations in society, the fraction of players they observe choosing k must also converge to $m_\infty(k, \theta)$. It is a consequence of the Strong Law of Large Numbers. Therefore, players will asymptotically know that the true state cannot be any state θ' such that $m_\infty(k, \theta') \neq m_\infty(k, \theta)$. This is what Lemma 5 below states. For a formal proof, see Appendix A.1.

Lemma 5 (Correct shares are observed in the limit). *Let (m^*, F^*) be an equilibrium and $\theta, \theta' \in \Theta$ such that $m^*_\infty(\theta') \neq m^*_\infty(\theta)$. Then $\mu(F^*(i)|\theta, m^*)(\pi^i_\infty(\theta') = 0) = 1$, for λ-almost all $i \in I$.*

Before we proceed, we state some definitions. Suppose that, when the true state is θ, almost all players asymptotically discover that the state is not θ', and vice-versa. In this case, we say that θ is distinguishable from θ'.

Definition 2 (Distinguishable States)**.** *Let (m^*, F^*) be an equilibrium. We say θ is distinguishable from θ', and vice-versa, if*

1. $\mu(F^*(i)|\theta, m^*)(\pi^i_\infty(\theta') = 0) = 1$ *and*

2. $\mu(F^*(i)|\theta', m^*)(\pi^i_\infty(\theta) = 0) = 1,$

for λ-almost all $i \in I$.

Next we provide a stronger definition. When a state θ is distinguishable from all other states of the world, players discover, in the long run, whether or not θ is the true state of the world. In this situation, we say θ is identified.

Definition 3 (Identified States)**.** *θ is identified if θ is distinguishable from θ', for every $\theta' \neq \theta$.*

We denote by ID the set of identified states.

We now provide two different sufficient conditions so that two states are distinguishable from each other. Consider θ to be the true state. Because almost all players asymptotically learn at least one component of the true state of the world (by Lemma 1), they asymptotically discover that a state θ' cannot be the true state if both $\theta'_1 \neq \theta_1$ and $\theta'_2 \neq \theta_2$. Moreover, the same happens if the asymptotic distribution of actions in θ', $m^*_\infty(\theta')$, is different from the true one, $m^*_\infty(\theta)$ (by Lemma 5). Lemma 6 summarizes these results.

Lemma 6 (Sufficiency for distinguishable states)**.** *Assume (m^*, F^*) is an equilibrium. Let $\theta, \theta' \in \Theta$. If at least one of the following conditions is true:*

1. $\theta_k \neq \theta'_k$*, for each* $k \in A$*;*

2. $m^*_\infty(\theta) \neq m^*_\infty(\theta')$.

Then θ is distinguishable from θ'.

Denote $B^i_\infty(k, \tilde{\theta}_k)$ the event such that player i asymptotically believes the true θ_k is at least $\tilde{\theta}_k$, that is, $\pi^i_\infty(\theta') = 0$, for each θ' such that $\theta'_k < \tilde{\theta}_k$.

We state the following definition regarding the players' equilibrium asymptotic beliefs.

Definition 4 (Efficient Beliefs)**.** *Let (m^*, F^*) be an equilibrium. We say players have efficient beliefs in θ if $\mu(F^*(i)|\theta, m^*)(B^i_\infty(k, \theta_k)) = 1$, for each $k \in A$ and λ-almost all $i \in I$.*

Intuitively, players do not underestimate the effect of any action if they have efficient beliefs. We denote by EB the set of states in which players have efficient beliefs. It is straightforward to see that $ID \subseteq EB$.

If a player chooses action k infinitely often and does not underestimate the payoff of $k' \neq k$, she is certainly choosing the best action for herself in the long run: she asymptotically learns θ_k (Lemma 1) and chooses k even though she can only have overestimated the expected payoff of k'. Hence, players with efficient beliefs must eventually play the best action for themselves. Next lemma is about this and this the reason we chose the name "efficient beliefs" for the definition above.

Let $BA^i(\theta)$ denote the set of player i's best actions when the state is θ.

Lemma 7 (Optimal Action under Efficient Beliefs)**.** *Assume θ is the true state and players follow optimal strategies. If $\theta \in EB$, then $A^i_\infty \subseteq BA^i(\theta)$, for λ-almost all players $i \in I$.*

Assume θ is the true state and all states $\underline{\theta}$ such that $\underline{\theta} < \theta$ are identified.[11] It is straightforward to see that θ is distinguishable from such $\underline{\theta}$'s. But, much stronger than that, Lemma 8 below asserts that θ is also distinguishable from states $\bar{\theta} > \theta$. This means that a sufficient condition for a given state to be identified is that all states "below" are identified.[12]

[11] Recall that "$<$" (and similarly "$>$") refers to the partial ordering on the set of states such that $\theta' < \theta$ if $\theta'_1 < \theta_1$ and $\theta'_2 \leq \theta_2$ or if $\theta'_1 \leq \theta_1$ and $\theta'_2 < \theta_2$.

[12] Note that states $\tilde{\theta}$ such that neither $\tilde{\theta} < \theta$ nor $\tilde{\theta} > \theta$ are distinguishable from θ because both $\tilde{\theta}_1 \neq \theta_1$ and $\tilde{\theta}_2 \neq \theta_2$.

To prove this result, first we show $\theta \in EB$. Consider players who play $k = 1$ infinitely often. Lemma 1 implies that they asymptotically learn θ_1. Since they distinguish θ from any $\underline{\theta} < \theta$, they cannot asymptotically underestimate action $k = 2$. Analogously, players who play $k = 2$ also do not asymptotically underestimate action $k = 1$, thus $\theta \in EB$.

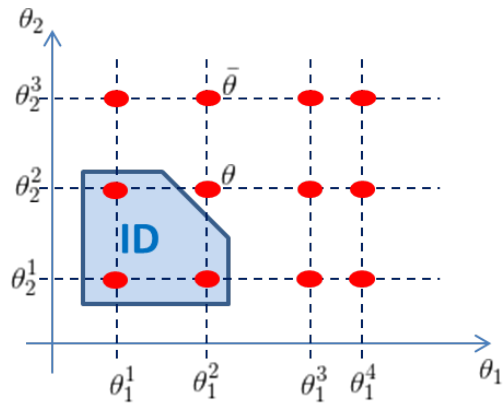

Figure 1: $\theta = (\theta_1^2, \theta_2^2)$ is the true state and $\underline{\theta} \in ID$ for all $\underline{\theta} < \theta$.

Now we argue that θ is distinguishable from any $\bar{\theta} > \theta$. Suppose, for instance, that θ is not distinguishable from $\bar{\theta}$ and is such that $\bar{\theta}_1 = \theta_1$ and $\bar{\theta}_2 > \theta_2$ (see figure 1), then $m_\infty(\bar{\theta}) = m_\infty(\theta)$ by Lemma 6. Since almost all players eventually play correctly when the state is θ (Lemma 7), a strictly positive mass of players who would be better off playing $k = 2$ when the state is $\bar{\theta}$ have to play $k = 1$ infinitely often. In other words, under $\bar{\theta}$, every agent is weakly better off, however, not all of them will have a different optimal action. Assumption A3 guarantees that at least a positive mass of agents will have her optimal action reversed: it is $k = 1$ under θ but is $k = 2$ under $\bar{\theta}$. The role that assumption A3 plays can be better understood through figures 2 and 3 below.

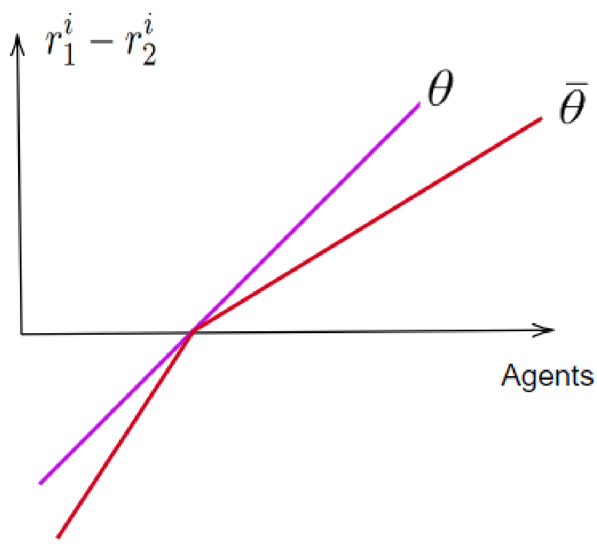

Figure 2: Violation of Assumption A3

In the example of figure 2, every agent prefers $\bar{\theta}$ to θ, but the optimal action for each agent i under θ is the same as under $\bar{\theta}$. This example violates Assumption A3. Instead, assumption A3 guarantees that there is a minimum distance between the blue line and the new curve representing the payoff distance between actions 1 and 2 in state $\bar{\theta}$. This can better be seen below:

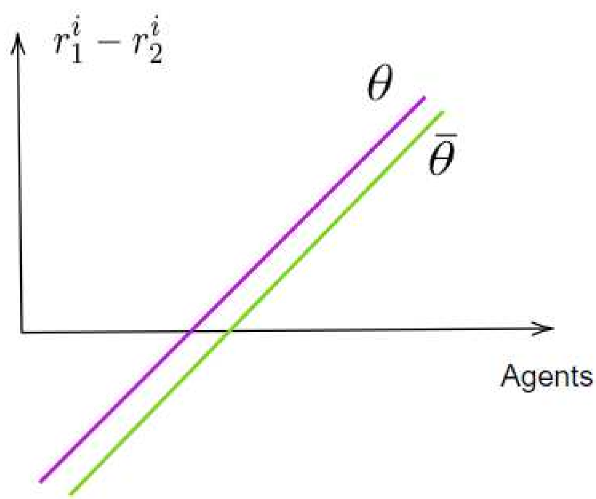

Figure 3: Payoff difference under Assumption A3

Note that now, in the example of figure 3, the outer curve has a minimum distance from the inner curve. This means that if agents found out that the true state is $\bar{\theta}$ instead of θ, there would be a positive mass of agents who would be better off by switching their actions.

When the true state is $\bar{\theta}$, players asymptotically discover the true state cannot be states $\underline{\theta} < \theta$ since they are identified. If states $\bar{\theta}$ and θ are not distinguishable, then we use the argument of the previous paragraph to show that a strictly positive mass of players (players close to indifference between $k = 1$ and $k = 2$ when the state is θ who choose $k = 1$) has to assign probability arbitrarily close to 0 to the true state $\bar{\theta}$ in the long run (otherwise, they would switch actions and the states would be distinguishable). However, since beliefs in the true state are submartingales, this cannot happen. This means that $m_\infty(2, \bar{\theta}) > m_\infty(2, \theta)$ and, therefore, θ is also distinguishable from a state $\bar{\theta} > \theta$ (Lemma 2). Hence, θ is identified. A formal proof can be found in Appendix A.2.

Lemma 8 (Sufficiency: Lower States Identified). *Consider an equilibrium and let $\theta \in \Theta$. Assume $\forall \underline{\theta} \in \Theta$, $\underline{\theta} < \theta \Rightarrow \underline{\theta} \in ID$. Then $\theta \in ID$.*

Lemma 8 implies that (θ_1^1, θ_2^1) is identified. In turn, (θ_1^1, θ_2^2) and (θ_1^2, θ_2^1) are also identified, and so on. With a simple induction argument, Lemma 8 is the key to prove Theorem 1, our main result.

Theorem 1 (Complete Learning). *In equilibrium, $\theta \in ID$, for each $\theta \in \Theta$.*

Theorem 1 shows that players asymptotically learn the true state of the world. Corollary 1 asserts that players eventually choose the best action for themselves.

Corollary 1 (Efficiency). *In equilibrium, $A_\infty^i \subseteq BA^i(\theta)$, for each $\theta \in \Theta$ and λ-almost all players $i \in I$.*

Appendix A Proofs

A.1 Proof of Lemma 5

Proof. Consider $\theta' \neq \theta$. Fix $i \in I_k$ a player with strategy σ and beliefs $\{\pi_t\}$ that plays k infinitely many times. If $\theta'_k \neq \theta_k$, Lemma 1 completes the proof. Suppose, then, that $\theta'_k = \theta_k$. Let $\{\tilde{k}_t\}$ be the sequence of observations in society. By Lemma 3, there exists T such that i plays k for all $t \geq T$. By Bayes' Rule, for $t \geq T$,

$$\frac{\pi_{t+1}(\theta')}{\pi_{t+1}(\theta)} = \frac{\pi_t(\theta')}{\pi_t(\theta)}.\frac{m_t(\tilde{k}_t, \theta')}{m_t(\tilde{k}_t, \theta)}. \tag{2}$$

The outcome y_t does not change the likelihood ratio $\frac{\pi(\theta')}{\pi(\theta)}$ since $\theta'_k = \theta_k$.
In equation 2, defining $\gamma_t := \log\left(\frac{\pi_t(\theta')}{\pi_t(\theta)}\right)$ and $\zeta_t := \log\left(\frac{m_t(\tilde{k}_t,\theta')}{m_t(\tilde{k}_t,\theta)}\right)$,

$$\gamma_{t+1} = \gamma_t + \zeta_t. \tag{3}$$

Using strict concavity of the log function, for t such that $m_t(\theta) \neq m_t(\theta')$,

$$\begin{aligned} \mathbb{E}[\zeta_t|\theta] &= m_t(\tilde{k}, \theta) \log\left(\frac{m_t(\tilde{k}_t, \theta')}{m_t(\tilde{k}_t, \theta)}\right) + (1 - m_t(\tilde{k}, \theta)) \log\left(\frac{1 - m_t(\tilde{k}_t, \theta')}{1 - m_t(\tilde{k}_t, \theta)}\right) \\ &< \log\left(m_t(\tilde{k}, \theta)\frac{m_t(\tilde{k}_t, \theta')}{m_t(\tilde{k}_t, \theta)} + (1 - m_t(\tilde{k}, \theta))\frac{1 - m_t(\tilde{k}_t, \theta')}{1 - m_t(\tilde{k}_t, \theta)}\right) = 0. \end{aligned} \tag{4}$$

Since $\{m_t\}$ converges, $\{\mathbb{E}[\zeta_t|\theta]\}$ must also converge to some ζ_∞. Because $m_\infty(\theta) \neq m_\infty(\theta')$, the previous equation guarantees that ζ_∞ is strict negative.

Consider $Z_n := \frac{1}{n}\sum_{t=T}^{T+n-1} \zeta_t$. As a consequence of the Strong Law of Large Numbers, $\{Z_n\}$ converges almost surely to $\zeta_\infty < 0$. Hence, we get that $\sum_{t=T}^{\infty} \zeta_t \overset{a.s.}{\to} -\infty$. Then $\gamma_t \overset{a.s.}{\to} -\infty$ and, thus, $\{\pi_t(\theta')\}$ converges almost surely to 0.

□

A.2 Proof of Lemma 8

Proof. Fix $\theta \in \Theta$. Assume,

$$\forall \underline{\theta} \in \Theta,\ \underline{\theta} \leq \theta \Rightarrow \underline{\theta} \in ID. \tag{5}$$

Since players can distinguish θ from any $\underline{\theta} < \theta$, Lemma 1 implies that $\theta \in EB$. Lemma 7 implies that $A^i_\infty = BA^i(\theta)$ for λ-almost all players.

Now we prove that $\theta \in ID$. Assume, by contradiction, there exists $\bar{\theta} \neq \theta$ such that θ is not distinguishable from $\bar{\theta}$. Lemma 6 implies $m_\infty(\bar{\theta}) = m_\infty(\theta)$ and either $\bar{\theta}_1 = \theta_1$ or $\bar{\theta}_2 = \theta_2$. Assume, without loss of generality, $\bar{\theta}_1 = \theta_1$. Equation 5 implies that $\bar{\theta}_2 > \theta_2$.

When the state is $\bar{\theta}$, using Lemmas 1 and 2 and the fact that states $(\theta_1, \underline{\theta}_2)$ with $\underline{\theta}_2 < \theta_2$ are either inexistent or identified,

$$i \in I_1 \text{ in } \bar{\theta} \Rightarrow r^i_1(\theta_1) = r^i_1(\bar{\theta}) \geq \mathbb{E}_{\pi^i_\infty}[r^i_2] \geq r^i_2(\theta_2),\ \text{for almost all } i \in I. \tag{6}$$

When the state is θ, $A^i_\infty = BA^i(\theta)$ for almost all players. Then equation 6 can be rewritten as

$$i \in I_1 \text{ in } \bar{\theta} \Rightarrow i \in I_1 \text{ in } \theta,\ \text{for almost all } i \in I. \tag{7}$$

Equation 7 implies that $m_\infty(1, \bar{\theta}) \leq \lambda\{i \in I : BA^i(\theta) = \{1\}\} = m_\infty(1, \theta)$. Since we assumed $m_\infty(\bar{\theta}) = m_\infty(\theta)$, the converse of equation 7 must be true.

$$i \in I_1 \text{ in } \theta \Rightarrow i \in I_1 \text{ in } \bar{\theta},\ \text{for almost all } i \in I. \tag{8}$$

Consider $I(\varepsilon) := \{i \in I : 0 \leq r^i_1(\theta_1) - r^i_2(\theta_2) < \varepsilon\} \subset \{i \in I : BA^i(\theta) = \{1\}\}$. Note that $\lambda\{I(\varepsilon)\} > 0$ according to assumption A1. When the true state is θ, almost all $i \in I(\varepsilon)$ are such that $i \in I_1$. Equation 8 implies also that almost all $i \in I(\varepsilon)$ are such that $i \in I_1$ when the true state is $\bar{\theta}$.

Consider $\bar{\theta}$ is the true state. For $i \in I(\varepsilon) \cap I_1$,

$$0 \leq r_1^i(\theta_1) - \mathbb{E}_{\pi_\infty^i}[r_2^i] \leq r_1^i(\theta_1) - [\pi_\infty^i(2, \bar{\theta}_2) r_2^i(\bar{\theta}_2) + (1 - \pi_\infty^i(2, \bar{\theta}_2)) r_2^i(\theta_2)]. \tag{9}$$

Where the last inequality considers that $r_2^i(\theta_2)$ is the worst possible expected payoff for action $k = 2$ according to equation 5. Rewriting equation 9,

$$0 \leq r_1^i(\theta_1) - r_2^i(\theta_2) - \pi_\infty^i(2, \bar{\theta}_2)(r_2^i(\bar{\theta}_2) - r_2^i(\theta_2)) \leq r_1^i(\theta_1) - r_2^i(\theta_2) < \varepsilon. \tag{10}$$

Taking ε arbitrarily small, equation 10 implies that $\pi_\infty^i(2, \bar{\theta}_2)$ must be arbitrarily small (using assumption A3), for almost all $i \in I(\varepsilon)$. We get a contradiction: beliefs on the true state are submartingales, thus cannot be arbitrarily wrong for almost all $i \in I(\varepsilon)$.

□

References

S. N. Ali. Herding with costly information. *Journal of Economic Theory*, 175:713–729, 2018.

M. Aoyagi. Mutual observability and the convergence of actions in a multi-person two-armed bandit model. *Journal of Economic Theory*, 82(2):405–424, 1998.

V. Bala and S. Goyal. Conformism and diversity under social learning. *Economic theory*, 17 (1):101–120, 2001.

A. V. Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107 (3):797–817, 1992.

S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.

J. Boisjoly, G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles. Empathy or antipathy? the impact of diversity. *American Economic Review*, 96(5):1890–1905, December 2006.

B. Camargo. Learning in society. *Games and Economic Behavior*, 87:381–396, 2014.

S. E. Carrell, M. Hoekstra, and J. E. West. The impact of college diversity on behavior toward minorities. *American Economic Journal: Economic Policy*, 11(4):159–82, November 2019.

G. Ellison and D. Fudenberg. Rules of thumb for social learning. *Journal of Political Economy*, 101(4):612–643, 1993.

J. K. Goeree, T. R. Palfrey, and B. W. Rogers. Social learning with private and common values. *Economic theory*, 28(2):245–264, 2006.

M. Harel, E. Mossel, P. Strack, and O. Tamuz. Rational groupthink. *The Quarterly Journal of Economics*, 136(1):621–668, 2021.

N. Kartik, S. Lee, T. Liu, and D. Rappoport. Beyond unbounded beliefs: How preferences and information interplay in social learning. *Working paper*, 2022.

I. Monzon. Observational learning in large anonymous games. *Theoretical Economics*, 14(1): 403–435, 2019.

K. Munshi. Social learning in a heterogeneous population: technology diffusion in the indian green revolution. *Journal of Development Economics*, 73(1):185–213, 2004.

D. Rosenberg, E. Solan, and N. Vieille. Informational externalities and emergence of consensus. *Games and Economic Behavior*, 66(2):979–994, 2009.

M. Rothschild. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9(2):185–202, 1974.

L. Smith and P. Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68 (2):371–398, 2000.

T. Wiseman. Disagreement leads to complete learning: Sequential choice with continuous types. *Economics Letters*, 100(1):53–55, 2008.