Social Learning and the Shadow of the Past

Heller, Yuval and Mohlin, Erik

Bar Ilan University, Lund University

28 April 2017

Social Learning and the Shadow of the Past[*]

Yuval Heller[†] and Erik Mohlin[‡]

June 28, 2017

Abstract

In various environments new agents may base their decisions on observations of actions taken by a few other agents in the past. In this paper we analyze a broad class of such social learning processes, and study under what circumstances the initial behavior of the population has a lasting effect. Our main result shows that a population converges to the same behavior independently of the initial state, provided that the expected number of actions observed by each agent is less than one. Moreover, in any environment in which the expected number of observed actions is more than one, there is a learning rule for which the initial state has a lasting impact on future behavior.

Keywords: Social learning, steady state, unique limiting behavior, path dependence. **JEL Classification:** C73, D83.

1 Introduction

Agents must often make decisions without knowing the cost and benefits of the possible choices. In such situations an inexperienced (or "newborn") agent may learn from the experience of others, by basing his decision, on observations of a few actions taken by other agents in the past (see, e.g., the social learning models of Ellison & Fudenberg, 1993, 1995; Acemoglu *et al.*, 2011). In other environments, agents interact with random opponents, and an agent may base his choice of action on a few observations of how his current opponent behaved in the past (as first described in Rosenthal, 1979, and further developed and applied to various models of community enforcement in the Prisoner's Dilemma game in (Nowak & Sigmund, 1998; Takahashi, 2010; Heller & Mohlin, 2017)).

In this paper we analyze a broad class of social learning processes, and study under what circumstances the initial behavior of the population has a lasting influence on the population's behavior in the long run. Our main result shows that if the expected number of actions that each new agent observes is less than one, then the population converges to the same behavior independently of the initial state. Moreover, if the expected number of observed actions is more than one, then there is a rule according to which agents learn from the experience of others, such that if the agents follow this learning rule, then the environment admits multiple steady states, and the initial state determines which steady states will prevail.

Overview of the Model   We consider an infinite population of agents (a continuum of mass one). Time is discrete and in every period each agent is faced with a choice among a fixed set of alternatives. Each agent in the population is endowed with a type. The *population state* is a vector describing the aggregate distribution of actions played by agents of each type. In each period a fixed share of the agents die and are replaced with new agents. Each new agent observes a finite sequence of actions (called a *sample*) of random size. The sample may consist of either past actions of random agents in the population (as in the social learning models mentioned above) or past actions of the current random opponent (as in the community enforcement models mentioned above). An environment is a tuple that specifies all the above components.

A profile of learning rules assigns to each type in the population a rule that determines the distribution of actions played by a new agent as a function of the observed sample. The agent keeps playing the same action throughout his life. The environment and the profile of learning rules jointly induce a mapping between population states that determines a new population state for each initial state. A population state is a *steady state* if it is a fixed point of this mapping.

Main Result   Our main result (Theorem 2) fully characterizes which environments admit learning rules with multiple steady states. Specifically, it shows that an environment admits a learning rule with more than one multiple steady state if and only if the mean sample size is strictly more than one (or if agents always observe a single action). In the opposite case, each profile of learning rules admits a unique steady state, and, moreover, the population converges to the unique steady state from any initial state.

The intuition for the "only if" side is as follows. Since there is a deterministic mapping from what sample an agent observes to what distribution of actions the agent plays, the only way to make an agent choose a different distribution of actions is to show him a different sample. The probability that a sample induced by one population state is different from a sample induced by another population state is bounded by the distance between the two population states times the mean sample size. This implies that the mapping between states is a contraction mapping whenever the mean sample size is less than one.

The "if" side relies on constructing a specific learning rule, according to which agents play action a' if they observe action a' in their sample, and play action a'' otherwise. One can show that such a learning rule always admits two different steady states provided that the expected number of observed actions is greater than one. We demonstrate that this sampling rule may be consistent with Bayesian inference and the agents using best replies.

Extensions   We extend our model to deal with non-stationary environments, in which the distribution of sample sizes and the agents' learning rules depend on calendar time, and we characterize when a non-stationary environment admits a unique sequence of states, such that it converges to this sequence of states from any initial population state. We further extend the model to time-dependent common stochastic shocks that influence the learning rules of all agents, and we characterize when the initial population state may have a lasting effect in such environments.

Our results so far have not assumed anything about the agents' learning rules. Obviously, additional information on the learning rules, may allow us to achieve stronger results. Next, we present a simple notion that measures how responsive a learning rule is to different samples, and we use this notion to define the effective sample size of a learning process (which is always weakly smaller than the simple mean sample size).

Finally, we apply the notion of effective sample size to derive a tighter upper bound for learning processes that admit unique steady states.

Related Literature Various papers have studied different aspects of the question of when the initial behavior of the population has lasting effects on social learning processes. Most of this literature focuses on specific learning rules, according to which new (or revising) agents myopically best reply to the empirical frequency of the observed actions. Arthur (1989) (see related models and extensions in Arthur, 1994; Kaniovski & Young, 1995; Smith & Sorensen, 2014) studies games in which agents sequentially choose which competing technology to adopt, and he shows that social learning is path-dependent if the technologies have increasing returns.

Kandori *et al.* (1993) and Young (1993a) study models of finite large populations that are involved in a social learning process, and agents occasionally make mistakes (e.g., an agent adopts a technology that is not his myopic best reply to his sampled information). They show that the path dependency of the social learning process vanishes when infinite time horizons are considered. In many cases, when the probability of mistakes is sufficiently small the population spends almost all the time in a unique "stochastically stable state," which is independent of the initial state. A key difference between our model and theirs is that we model an infinite population, rather than a large finite population. In Section 8, we discuss the relations between the present paper and the literature on stochastic stability, and, in particular, the implications of our results for finite large populations.

Banerjee & Fudenberg (2004) study a model with a continuum of agents in which a fixed share of new agents in each period choose one of two technologies. There are two possible states of nature, and each technology has a higher quality in one of these states. Each agent, after he observes l past actions and a noisy signal about the quality of each technology, chooses the technology with the higher expected quality, conditional on the information that he has observed. Banerjee & Fudenberg show that when $l \geq 2$ the behavior of the population converges to everyone choosing the efficient technology, while if $l = 1$ the population converges to an inefficient state in which only some of the agents choose the (ex-post) better technology.

Sandholm (2001) shows that when each new agent observes k actions and the game admits a $\frac{1}{k}$-dominant action a^* (i.e., action a^* is the unique best reply against any mixed strategy assigning a mass of at least $\frac{1}{k}$ to a^*), then social learning converges to this action regardless of the initial state. Recently, Oyama *et al.* (2015) strengthened this result by extending it to iterated p-dominant actions, and by showing that global convergence is fast.

Our model differs from all the above-mentioned research in that we study general environments and arbitrary learning rules. Specifically, we ask what properties of the agents' sampling procedures imply that *any* learning rule admits a unique steady state and global convergence to this state, whereas the existing literature focuses on the dynamic behavior induced by a specific learning rule (in most of the literature, the agents myopically best reply to specific payoffs, such as those induced by competing technologies with increasing returns).

Structure We present motivating examples in Section 2. The basic model is described in Section 3. Section 4 presents our main results. Section 5 generalizes the basic model to heterogeneous populations. In Section 6 we extend the analysis to non-stationary learning processes and to common shocks. In Section 7 we define the responsiveness of a learning rule, and use it to achieve a stronger sufficient condition for an environment to admit a unique steady state. We conclude in Section 8. Technical proofs are presented in the Appendix.

2   Motivating Examples

In this section we present three motivating examples, which will be revisited further below to demonstrate the applicability of our model and the implications of our results. In all the examples the population is modeled as a continuum of mass one, and time is discrete. The main example deals with social learning with competing technologies, while the other two examples study situations in which agents are randomly matched to play a two-player game.

Example 1 (Main Motivating Example: Competing Technologies with Increasing Returns[1])**.** Consider a population in which in each period a share of $\beta \in (0,1)$ of the incumbent agents die, and are replaced with new agents. Each new agent chooses one of two competing technologies with increasing returns a' and a'', which he adopts for the rest of his life. A share of 99% of the new agents, before making their own choice, observe the technology followed by a single incumbent. We consider two cases for what the remaining 1% of the new agents observe before they choose a technology: (I) they observe no actions, and (II) they observe three actions (as summarized in Table 1). Half of the agents have a slight preference for technology a', and the remaining half have a slight preference for technology a''. These preferences are much smaller than preferences over payoff differences that are due to the fact that the technologies have increasing returns. For example, assume that if 51% of the population follows technology a', then an agent prefers technology a' regardless of his own idiosyncratic preference.

A population state $\alpha_t \in [0,1]$ describes the share of agents who use technology a' at time t. New agents are uncertain about the initial population state[2] α_1. The common prior about the initial state is that α_1 is uniformly distributed on $[30\%, 70\%]$.

The following stationary behavior of new agents constitutes a Nash equilibrium in this environment:[3]

1. An agent who does not observe any incumbent chooses a technology based on his idiosyncratic preferences.

2. An agent who observes a single incumbent chooses the technology played by the observed incumbent due to the increasing returns of the technologies. This is because conditional on the agent observing an incumbent following technology a in round t, a majority of the agents at time t are more likely to follow technology a'.

3. An agent who observes three incumbents chooses the technology played by a majority of the sample (due to the increasing returns of the technologies).

One can show that in Case (I), in which the mean sample size of a new agent is slightly less than one, the population converges to a unique steady state in which half of the agents follow each technology. By contrast, in Case (II), in which the mean sample size is slightly more than one, the initial behavior of the population has a lasting effect. Specifically, the population converges to everyone following technology a if initially a majority of the agents followed technology a' (i.e., if $\alpha_1 > 50\%$), and the population converges to everyone following technology a'' if $\alpha_1 < 50\%$.

---

[1]The example is similar to the model of Banerjee & Fudenberg (2004), except that the technologies have increasing returns, rather than having unknown inherent different qualities.

[2]As argued by Banerjee & Fudenberg (2004, p. 5), the aggregate uncertainty about the initial population state may reflect the choices of a group of "early adopters" whose preferences are uncertain even at the aggregate level.

[3]This equilibrium is unique if either each agent is sufficiently short-lived (i.e., β is sufficiently large), or if agents sufficiently discount the future.

Table 1: Summary of the Two Cases in Example 1

Case	Probability of Observing			Mean Sample Size	Convergence and steady states
	0 actions	1 action	3 actions		
I	1%	99%	-	0.99	Global convergence to 50%–50%
II	-	99%	1%	1.02	Convergence to a' if $\alpha_1 > 0.5$; convergence to a'' if $\alpha_1 < 0.5$.

Example 2 (Community Enforcement in the Prisoner's Dilemma)**.** Consider a population in which in each round each agent is randomly matched with to opponents, and plays a Prisoner's Dilemma with each of them. The population includes two types of agents. A share of 99% of the agents are of the first type. In round one, each agent of type one defects with probability α_1 in each match. In each later round, each agent of type 1 observes in each match a single action played in the previous round by his current opponent, and the agent plays this observed action in the current match. The remaining 1% of the agents are of the second type. In round one, each agent of type 2 defects with probability α_2 in each match. We consider two different cases for the observations and behavior of type 2 in each match in each later round: (I) each agent of type 2 observes no actions, and defects with probability α_2, and (II) each agent of type 2 observes the three actions played by his partner in the three matches played by the partner in the previous round, and the agent defects if and only if the partner has defected in at least two of the three observed matches.

One can show that in Case (I), in which the mean sample size of a random agent is slightly less than one, the population globally converges to a unique steady state in which in each round a share of α_2 of the agents of each type defect in each round. By contrast, in Case (II), in which the mean sample size of a random agent is slightly more than one, the initial behavior of the population has lasting effects. Specifically, the population converges to everyone defecting if the initial probability of defection of a random agent is more than half (i.e., if $\bar{\alpha} := 99\% \cdot \alpha_1 + 1\% \cdot \alpha_2 > 50\%$), and the population converges to everyone cooperating if $\bar{\alpha} < 50\%$.

Example 3 (Rock-Paper-Scissors)**.** Consider a population in which each agent is randomly matched in each round to play the rock-paper-scissors game. Each player has three pure actions (*rock*, *paper*, *scissors*), and each action is the unique best reply to the previous action (modulo 3). In the initial round $t = 1$ the aggregate distributions of actions is $\hat{\gamma} \in \Delta\left(rock, paper, scissors\right)$. In each later round, each agent observes the opponent's action in the previous round with probability p, and best replies to the observed action. With the remaining probability of $1 - p$ the agent observes no actions, and plays the mixed action $\gamma' \in \Delta\left(rock, paper, scissors\right)$. Observe that the parameter p is equal to the mean sample size of a random agent. If $p = 1$ it is immediate that the population's behavior cycles "around" permutations of the initial behavior (as is common in evolutionary models of rock-paper-scissors; see, e.g., the analysis in Cason *et al.*, 2014). Formally, let $t \in \{0, 1, 2, ...\}$:

1. In round $3 \cdot t + 1$ agents play *rock* with a probability of $\hat{\gamma}\left(rock\right)$, *paper* with a probability of $\hat{\gamma}\left(paper\right)$, and *scissors* with a probability of $\hat{\gamma}\left(scissors\right)$.

2. In round $3 \cdot t + 2$ agents play *rock* with a probability of $\hat{\gamma}\left(scissors\right)$, *paper* with a probability of $\hat{\gamma}\left(rock\right)$, and *scissors* with a probability of $\hat{\gamma}\left(paper\right)$.

3. In round $3 \cdot t + 3$ agents play *rock* with a probability of $\hat{\gamma}\left(paper\right)$, *paper* with a probability of $\hat{\gamma}\left(scissors\right)$, and *scissors* with a probability of $\hat{\gamma}\left(rock\right)$.

However, when $p < 1$, one can show that the population converges to the following unique behavior (regardless of the initial behavior α_0):

$$\Pr(rock) = \frac{\gamma'(rock) + p \cdot \gamma'(scissors) + p^2 \cdot \gamma'(paper)}{1 + p + p^2},$$

$$\Pr(paper) = \frac{\gamma'(paper) + p \cdot \gamma'(rock) + p^2 \cdot \gamma'(scissors)}{1 + p + p^2}.$$

$$\Pr(scissors) = \frac{\gamma'(scissors) + p \cdot \gamma'(paper) + p^2 \cdot \gamma'(rock)}{1 + p + p^2},$$

Note that when p is close to one, the unique behavior is close to the uniform mixed profile that assigns a probability of $\frac{1}{3}$ to each action.[4]

3 Basic Model

Throughout the paper we restrict attention to distributions with a finite support. Given a (possibly infinite) set X, let $\Delta(X)$ denote the set of distributions over this set that have a finite support. With a slight abuse of notation we use $x \in X$ to denote the degenerate distribution $\mu \in \Delta(X)$ that assigns probability one to x (i.e., we write $\mu \equiv x$ if $\mu(x) = 1$). We use $\mathbb{N}$ to denote the set of natural numbers including zero.

Population state. Consider an infinite population of agents.[5] More precisely, the population consists of continuum of agents with mass one. Time is discrete and in every period (or "round") each agent is faced with a choice among a fixed set of alternatives A. Let A be a finite set of at least two actions (i.e., $|A| \geq 2$).

The *population state* (or *state* for short) is identified with the aggregate distribution of actions played in the population, denoted $\gamma \in \Delta(A)$. Let Γ denote the set of all population states.

New/Revising agents. In each period, a share of $0 < \beta \leq 1$ of the agents exit the population and are replaced with new agents, while the remaining $1 - \beta$ share of the agents play the same action as they played in the past (see, e.g., Banerjee & Fudenberg, 2004). Each new agent chooses an action based on a sample of a few actions of incumbents. The agent then keeps playing this chosen action throughout his active life, possibly because the initial choice requires a substantial action-specific investment, and it is too costly for an agent to reinvest in a different action later on. The model can also be interpreted as describing a fixed population in which each agent reevaluates his action only every $\frac{1}{\beta}$ periods (under this interpretation, when the sample is non-empty, the first observed action might be interpreted as the revising agent's own past action).

Sample. Each new agent observes a finite sequence of actions (or *sample*). The size of the observed sample is a random variable with a distribution $\nu \in \Delta(\mathbb{N})$. Let M denote the set of all feasible samples, i.e.,

$$M = \cup_{l \in supp(\nu)} A^l,$$

---

[4]Observe that the analysis depends only on the ordinal best-reply correspondence over pure actions, and is independent of the cardinal payoffs. The uniform mixed profile is not necessarily the Nash equilibrium of the underlying game, as the latter depends on the cardinal payoffs.

[5]In Section 8, we discuss the implications of our results for large finite populations.

where $A^0 = \{\emptyset\}$ is a singleton consisting of the empty sample $\emptyset$. Let $\bar{l} = \max\left(supp\left(\nu\right)\right) < \infty$ be the maximal sample size. Note that M is finite in virtue of the finite-support assumption. For each sample size $l \in \mathbb{N}$, let $\psi_l : \Gamma \to \Delta\left(A^l\right)$ denote the distribution of samples observed by each agent in the population (*or sampling rule* for short), conditional on the sample having size l. A typical sample of size l is represented by the vector $\overrightarrow{a} = (a_1, ..., a_l)$.

We assume that each agent independently samples different agents, and observes a random action played by each of these agents. This kind of sampling is common in models of social learning (see, e.g., Ellison & Fudenberg, 1995; Banerjee & Fudenberg, 2004). Formally, we define for each sample size $l \in \mathbb{N}$, each state $\gamma \in \Gamma$, and each sample $(a_1, ..., a_l)$,

$$\psi_{l,\gamma}\left(a_1, ..., a_l\right) = \prod_{1 \leq i \leq l} \gamma\left(a_i\right). \tag{1}$$

Environment. An *environment* is a tuple

$$E = (A, \beta, \nu)$$

that includes the three components described above: a finite set of actions A, a fraction of new agents at each stage β, and a distribution of sample sizes ν.

Given environment $E = (A, \beta, \nu)$, let μ_l denote the *mean sample size*, i.e., the expected number of actions observed by a random agent in the population. Formally:

$$\mu_l = \sum_{l \in supp(\nu)} \nu\left(l\right) \cdot l.$$

Learning rule and stationary learning process. Each new agent chooses his action in the new population state by following a stationary (i.e., time-independent) *learning rule* $\sigma : M \to \Delta\left(A\right)$. That is, a new agent who observes sample $m \in M$ plays action a with probability $\sigma_m\left(a\right)$. The remaining $1 - \beta$ agents play the same action as in the previous round.

A stationary learning process (or *learning process* for short) is a pair

$$P = (E, \sigma) = (A, \beta, \nu, \sigma)$$

consisting of an environment and a learning rule.

Population dynamics. An initial state and a learning process uniquely determine a new state. To see this note that since the number of messages M, and actions A are finite, whereas the population is a continuum, the *probability* that an agent observes a message m and switches to an action a is equal to the *fraction* of agents who observe a message m and switch to an action a. For this reason we say that the learning process is *deterministic*, despite the fact that the choice of an individual agent may be stochastic.

Time is discrete in our model. Let $f_P : \Gamma \to \Gamma$ denote the mapping between states induced by a single step of the learning process P. That is, $f_P\left(\hat{\gamma}\right)$ is the new state induced by a single step of the process P, given an initial state $\hat{\gamma}$. Similarly, for each $t > 1$, let $f_P^t\left(\hat{\gamma}\right)$ denote the state induced after t steps of the learning process P, given an initial state $\hat{\gamma}$ (e.g., $f_P^2\left(\hat{\gamma}\right) = f_P\left(f_P\left(\hat{\gamma}\right)\right)$, $f_P^3\left(\hat{\gamma}\right) = f_P\left(f_P\left(f_P\left(\hat{\gamma}\right)\right)\right)$, etc.).

We say that γ^* is a *steady state* with respect to the stationary learning process P, if it is a fixed point of the induced mapping f_P, i.e., if $f_P(\gamma^*) = \gamma^*$.

Steady state γ^* is a *global attractor*, if the population converges to γ^* from any initial state, i.e., if

$$\lim_{t \longrightarrow \infty} f_P^t(\hat{\gamma}) = \gamma^* \text{ for all } \hat{\gamma} \in \Gamma.$$

L_1**-distance.** We measure distances between probability distributions with the L_1-distance (norm). Specifically, let the L_1-distance between two distributions of samples $\psi_{l,\gamma}, \psi_{l,\gamma'} \in \Delta(A^l)$ of size l, be defined as follows:

$$\|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 = \sum\nolimits_{m \in A^l} |\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m)|.$$

Similarly the L_1-distance between two distributions of actions $\gamma, \gamma' \in \Delta(A)$ is defined as follows:

$$\|\gamma - \gamma'\|_1 = \sum_{a \in A} |\gamma(a) - \gamma'(a)|.$$

We conclude this section by demonstrating how the model captures motivating Examples **1** and **3**.

Example 1 (Competing Technologies revisited). The environment in which agents choose to adopt one of two competing technologies with increasing returns is modeled by a learning process

$$P = (\{a', a''\}, \beta, \nu, \sigma),$$

in which $\{a', a''\}$ is the set of competing technologies, $\beta \in (0,1)$ is the share of new agents that join the population in each round. The learning rule of the agent is defined as follows:

$$\sigma(\overrightarrow{a}) = \begin{cases} 0.5 \cdot a' + 0.5 \cdot a'' & \overrightarrow{a} = \emptyset \\ a' & \overrightarrow{a} \in \{a', (a', a', a'), (a'', a', a'), (a', a'', a'), (a', a', a'')\} \\ a'' & \text{otherwise.} \end{cases}$$

The initial population state is given by $\hat{\gamma}(a') = \alpha$. Finally the distribution of sample size is given by:

$$\text{Case I: } \nu(l) = \begin{cases} 1\% & l = 0 \\ 99\% & l = 1 \end{cases} \quad \text{Case II: } \nu(l) = \begin{cases} 1\% & l = 3 \\ 99\% & l = 1. \end{cases}$$

Observe that the mean sample size (μ_l) is equal to 0.99 in Case I, and is equal to 1.02 in Case II.

Example **3** (Rock-Paper-Scissors revisited) The environment in which agents play the rock-paper-scissors game is modeled by a learning process

$$P = (\{rock,\ paper,\ scissors\}, \beta = 1, \nu, \sigma),$$

where the distribution of the sample size is given by

$$\nu\left(l\right) = \begin{cases} 0 & 1-p \\ 1 & 0, \end{cases}$$

the initial population state is given by $\hat{\gamma}$. Finally, the learning rule is given by

$$\sigma\left(a\right) = \begin{cases} rock & a = scissors \\ paper & a = rock \\ scissors & a = paper. \end{cases} \qquad \sigma\left(\emptyset\right) = \gamma'.$$

Observe that the expected number of actions is equal to p (i.e., $\mu_l = p$).

4 Main Results

4.1 Upper Bound on the Distance between New States

Our first result shows that the distance between two new states is at most $(1 - \beta \cdot (1 - \mu_l))$ times the distance between the two initial states. Formally,

Theorem 1. *Let $P = (A, \beta, \nu, \sigma)$ be a stationary learning process, and let $\gamma \neq \gamma' \in \Gamma$ be two population states. Then:*

$$\left\| f_P\left(\gamma\right) - f_P\left(\gamma'\right) \right\|_1 \leq (1 - \beta \cdot (1 - \mu_l)) \cdot \left\| \gamma - \gamma' \right\|_1,$$

with a strict inequality if there exists an $l > 1$ such that $\nu\left(l\right) > 0$.

(Sketch of proof. Formal proof is presented for the more general result of Theorem 3.)
The distance between the final population states is bounded as follows:

$$\left\| \left(f_P\left(\gamma\right)\right) - \left(f_P\left(\gamma'\right)\right) \right\|_1 \leq \beta \cdot \sum_{l \in \mathbb{N}} \nu\left(l\right) \cdot \left\| \psi_{l,\gamma} - \psi_{l,\gamma'} \right\|_1 + (1 - \beta) \cdot \left\| \gamma - \gamma' \right\|_1. \tag{2}$$

The intuition of this inequality is as follows. The first part of the RHS of Eq. (2) reflects the actions played by the β new agents. The social learning stage may induce different behaviors for new agents who observe samples of size l only if they observe different samples. Thus, taking the weighted average of the distances between samples yields the bound on how much the aggregate behaviors of the new agents may differ (i.e., $\sum_{l \in \mathbb{N}} \nu\left(l\right) \cdot \left\| \psi_{l,\gamma} - \psi_{l,\gamma'} \right\|_1$). Finally, the mixed average of this expression and the behavior of the incumbents, gives the total bound on the difference between the final population states.

Next, observe that the distance between distributions of samples is bounded by the sample size times the distance between the distributions of actions:

$$\left\| \psi_{l,\gamma} - \psi_{l,\gamma'} \right\|_1 \leq l \cdot \left\| \gamma - \gamma' \right\|_1,$$

with a strict inequality if $l > 1$. This is so because the event that two samples of size l differ is (a non-disjoint) union of the l events: the first action in the samples differ, the second action in the simple differs, ..., the last

9

l^{th} action in the samples differ.

Substituting the second inequality in (2) yields:

$$\|(f_P(\gamma)) - (f_P(\gamma'))\|_1 \le \beta \cdot \sum_{l \in \mathbb{N}} \nu(l) \cdot l \cdot \|\gamma - \gamma'\|_1 + (1 - \beta) \cdot \|\gamma - \gamma'\|_1 =$$

$$\left(\beta \cdot \left(\sum_{l \in \mathbb{N}} \nu(l) \cdot l \right) + (1 - \beta) \right) \cdot \|\gamma - \gamma'\| = (\beta \cdot \mu_l + 1 - \beta) \cdot \|\gamma - \gamma'\| = (1 - \beta \cdot (1 - \mu_l)) \cdot \|\gamma - \gamma'\|_1 ,$$

with a strict inequality if there exists an $l > 1$ such that $\nu(l) > 0$. $\qquad\square$

Observe that $(1 - \beta \cdot (1 - \mu_l)) < 1$ iff $\mu_l < 1$. Recall that mapping f is a *weak contraction* (or shrinking) if $\|(f(\gamma)) - (f(\gamma'))\| < \|\gamma - \gamma'\|$ for each $\gamma \ne \gamma'$. Theorem 1 implies that f_P is a weak contraction mapping if either (1) $\mu_l < 1$, or (2) $\mu_l = 1$ and[6] $\nu(1) < 1$. The fact that the mapping f_P is a weak contraction mapping implies that f_p admits a global attractor.[7] Formally:

Corollary 1. *Let $P = (A, \beta, \nu, \sigma)$ be a stationary learning process satisfying (1) $\mu_l < 1$, or (2) $\mu_l = 1$ and $\nu(1) < 1$. Then f_P is a weak contraction mapping, which implies that (1) f_P admits a unique steady state γ^*, and (2) this unique steady state γ^* is a global attractor (i.e., $\lim_{t \to \infty} f_P^t(\hat{\gamma}) = \gamma^*$ for each $\hat{\gamma} \in \Gamma$).*

4.2 Full Characterization of Environments that Admit Multiple Steady States

Our main result fully characterizes which environments admit learning rules for which the past casts a long shadow. Specifically, it shows that an environment admits a learning rule with more than one multiple steady state iff $\mu > 1$ (alternatively if all agents sample exactly one action). In the opposite case ($\mu \le 1$) each learning rule admits a unique steady state, and, moreover, the population converges to the unique steady state from any initial state. Formally:

Theorem 2. *Let $E = (A, \beta, \nu)$ be an environment. The following two conditions are equivalent:*

1. $\mu_l > 1$, or $\nu(1) = 1$.

2. There exists a learning rule σ, such that the learning process (E, σ) admits two different steady states.

Proof. Corollary 1 immediately implies that $\neg 1 \Rightarrow \neg 2$. We are left with the task of showing that $1 \Rightarrow 2$.

Case A: Assume that $\nu(1) = 1$ (i.e., each new agent in the population observes a single action). Consider the learning rule in which each agent plays the action that he observed, i.e., $\sigma^*(a) = a$. Let γ be an arbitrary population state. Observe that γ is a steady state of the learning process (E, σ^*) because:

$$(f_P(\gamma))(a) = \gamma(a).$$

Case B: Assume that $\mu_l > 1$. Let a and a' be different actions ($a \ne a' \in A$). Let σ^* be a learning rule according to which each agent plays action a^* if he has observed action a^* at least once, and plays action a'

---

[6]Note that $\mu_l = 1$ and $\nu(1) < 1$ jointly imply that there exists $l > 1$ such that $\nu(l) > 0$.

[7]See Pata (2014, Theorem 1.7) for a formal proof that any weak contraction mapping on a compact metric space admits a global attractor (see also the sketch of the proof in Munkres, 2000, Section 28, Exercise 7). We thank Xiangqian Yang for kindly referring us to these proofs.

otherwise, that is,

$$\sigma^*\left(a^l\right) = \begin{cases} a^* & \exists i, \text{ s.t., } a_i^l = a^* \\ a' & \text{otherwise.} \end{cases}$$

It is immediate that the population state in which all agents play action a' (i.e., $\gamma\left(a'\right) = 1$) is a steady state of the learning process (E, σ^*). We now show that there exists $x > 0$, such that the population state γ^x in which all agents play action a^* with probability x, and play action a' with the remaining probability of $1 - x$ (i.e., $\gamma^x\left(a^*\right) = x$ and $\gamma^x\left(a'\right) = 1 - x$) is another steady state of the learning process (E, σ^*). Observe that the state γ^x is consistent with the learning process (E, σ^*) if and only if

$$\left(f_P\left(\gamma^x\right)\right)\left(a^*\right) = \sum_{l \in supp(\nu)} \nu\left(l\right) \cdot \frac{1}{|A|^l} \cdot \sum_{\vec{a} \in A^l} \mathbf{1}_{(\exists i \text{ s.t., } a_i = a^*)} = \sum_{l \in supp(\nu)} \nu\left(l\right) \cdot \sum_{a^i \in A^l \text{ s.t.,}} 1 - \left(1 - x\right)^l \equiv g\left(x\right). \quad (3)$$

In order to demonstrate the existence of multiple steady states we need to show that there is x^* such that $g\left(x^*\right) = x^*$. Observe that $g\left(x\right)$ (defined in (3) above) is a continuous increasing function of x, and that $g\left(1\right) \leq 1$. Below we show that $g\left(x\right) > x$ for $x << 1$, which implies by standard a continuity argument that there is $0 < x^* \leq 1$ such that $g\left(x^*\right) = x^*$, and hence γ^{x^*} is a steady state of the learning process (E, σ^*). We conclude the proof by showing that $g\left(x\right) > x$ for $x << 1$. Observe that when $x << 1$ is close to 0, then $1 - \left(1 - x\right)^l$ can be (Taylor-)approximated by

$$1 - \left(1 - x\right)^l = 1 - \left(1 - l \cdot x + O\left(x^2\right)\right) = l \cdot x + O\left(x^2\right).$$

This implies that when $x << 1$, $\left(f_P\left(\gamma^x\right)\right)\left(a^*\right)$ can be approximated by:

$$\left(f_P\left(\gamma^x\right)\right)\left(a^*\right) = \sum_{l \in supp(\nu_l)} \nu\left(l\right) \cdot \sum_{a^i \in A^l \text{ s.t.,} \exists i, \text{ s.t., } a_i^l = a^*} \left(l \cdot x + O\left(x^2\right)\right) =$$

$$x \cdot \sum_{l \in supp(\nu_l)} \nu\left(l\right) \cdot \sum_{a^i \in A^l \text{ s.t.,} \exists i, \text{ s.t., } a_i^l = a^*} l + O\left(x^2\right) = x \cdot \mu_l + O\left(x^2\right) > x.$$

$\square$

Remark 1. We note that the learning rules constructed in the proof above can be consistent with Bayesian inference and best-replying in plausible environments. The learning rule in Case A (playing the observed action) induces a Nash equilibrium in a setup with competing technologies with increasing returns and uncertainty about the initial population state, such as the setup presented in Example 1.

The learning rule in Case B induces a Nash equilibrium in the following setup of two competing technologies with uncertainty about their quality. There are two states of the world. In State 1 technology a^* has a higher quality, and in state 2 technology a' has a higher quality. The technology with the higher quality yields a payoff of one to an agent who follows it, and the technology with the lower quality yields a payoff of zero. State 1 has a prior probability of 60%. In state 1, 10% of the agents follow technology a^* in the first period, and the remaining agents follow technology a. In state 2, all agents follow technology a' in period one (i.e., the setup has a payoff-determined initial popularity à la Banerjee & Fudenberg, 2004). Observe that the unique Nash equilibrium in this setup is for an agent to play a^* when observing a^* at least once (as in this case the

11

agent knows for sure that action a^* has a higher quality), and to play a' otherwise (as in this case the posterior probability that action a' has a higher quality is at least 60%).

5 Heterogeneous Population

The basic model assumes that all agents share the same distribution of sample sizes, and the same learning rule. In many applications the population might be heterogeneous, i.e., the population includes various groups that differ in their sampling procedures and learning rules. A few examples of such models with heterogeneous populations can be found in:

1. Ellison & Fudenberg (1993), who study competing technologies where each technology is better for some of the players and these different tastes induce different learning rules (see also Munshi, 2004).

2. Young (1993b), who studies social learning in a bargaining model in which agents differ in the size of their samples.

3. Heller & Mohlin (2017), who in a companion paper analyze community enforcement in which the population includes several types of agents, and each type uses a different strategy.

5.1 Extended Model

In what follows we introduce heterogeneous populations that include different types, and we redefine the notions of population state, environment, and learning process to deal with this heterogeneity.

Population state. Let Θ denote a finite set of types with a typical element θ. Let λ_θ denote the mass of agents of type θ (or θ-*agents*). For simplicity, we assume that λ has full support. We redefine a *population state* (or *state* for short) to be a vector $\gamma = (\gamma_\theta)_{\theta \in \Theta}$, where each $\gamma_\theta \in \Delta(A)$ denotes the aggregate distribution of actions played by θ-agents. Let $\bar{\gamma} \in \Delta(A)$ denote the average distribution of actions in the population (i.e., $\bar{\gamma}(a) = \sum_\theta \lambda_\theta \gamma_\theta(a)$ for each action $a \in A$). A population state is *uniform* if all types play the same aggregate distribution of actions, i.e., if $\gamma_\theta(a) = \bar{\gamma}(a)$ for each type $\theta \in \Theta$ and action $a \in A$. We redefine Γ to denote the set of all populations with heterogeneous types.

New/Revising agents. As in the basic model, in each period, a share of $0 < \beta \leq 1$ of the agents of each type die and are replaced with new agents (or, alternatively, are randomly selected to reevaluate their choice), while the remaining $1 - \beta$ share of the agents of each type play the same action as they played in the past.

Sample. Each new agent observes a finite sequence of actions (or *sample*). The size of the sample observed by type θ is a random variable with a distribution $\nu_\theta \in \Delta(\mathbb{N})$. Let M, the set of all feasible samples, be redefined as follows

$$M = \cup_{\theta \in \Theta} \cup_{l \in supp(\nu_\theta)} A^l,$$

Let $\bar{l} = \max_{l \in} (\cup_{\theta \in \Theta} supp(\nu_\theta)) < \infty$ be the maximal sample size. For each sample size $l \in \mathbb{N}$, let $\psi_l : \Gamma \to \Delta(A^l)$ denote the distribution of samples observed by each agent in the population (or *sampling rule* for short), conditional on the sample having size l. A typical sample of size l is represented by the vector $\overrightarrow{a} = (a_1, ..., a_l)$.

We analyze two kinds of sampling methods in heterogeneous populations:

1. *Observing different random agents*: Each agent independently samples different agents, and observes a random action played by each of these agents. This kind of sampling is a common modeling choice in situations in which an agent's payoff depends not on the behavior of a specific sub-group of opponents, but on the agent's own action, the state of nature, and, possibly, the aggregate behavior of the population (see, e.g., Ellison & Fudenberg, 1995; Banerjee & Fudenberg, 2004). Formally, we define for each sample size $l \in \mathbb{N}$, each state $\gamma \in \Gamma$, and each sample $(a_1, ..., a_l)$,

$$\psi_{l,\gamma}(a_1, ..., a_l) = \prod_{1 \leq i \leq l} \bar{\gamma}(a_i). \tag{4}$$

2. *Observing a single random type*: Each agent randomly draws a type $\bar{\theta}$, and then the agent samples different agents of type $\bar{\theta}$, and observes a random action played by each of these $\bar{\theta}$-agents. This kind of observation is relevant to models in which the agent is randomly matched with an opponent, and may sample some actions played in the past by agents with the same type as the opponent. Formally, we define for each size $l \in \mathbb{N}$, each state $\gamma \in \Gamma$, and each sample $(a_1, ..., a_l)$,

$$\psi_{l,\gamma}(a_1, ..., a_l) = \sum_{\theta \in \Theta} \lambda_\theta \cdot \prod_{1 \leq i \leq l} \gamma_\theta(a_i). \tag{5}$$

   In the case of $\beta = 1$, this sampling method has another interpretation that is common in models of strategic interactions among randomly matched agents (e.g., Rosenthal, 1979; Nowak & Sigmund, 1998; Heller & Mohlin, 2017). According to this interpretation, each agent is involved in $n \geq \bar{l}$ interactions in each period. In each of these interactions the agent is randomly matched with a different opponent, and the agent observes a sample of random actions played by the opponent in the previous round. The random type of the opponent is distributed according to λ_θ, and each of the actions played by the opponent of type θ in the previous round is distributed according to γ_θ.

Observe that both cases, i.e., (4) and (5), coincide in two special setups: (1) when the population state is uniform (as in the basic model), or (2) when agents observe at most one action (i.e., $\bar{l} = 1$).

Remark 2. All our results work also in a setup in which some types use the first sampling method (i.e., observing different random agents), while other types use the second sampling method (i.e., observing a single random type).

Environment. We redefine an *environment* as a tuple

$$E = \left(A, \Theta, \beta, \psi_l, (\lambda_\theta, \nu_\theta)_{\theta \in \Theta}\right)$$

that includes the six components described above: a finite set of actions A, a finite set of types Θ, a fraction of new agents at each stage β, a sampling rule ψ_l (satisfying either (4) or (5)), a distribution over the set of types λ, and a profile of distributions of sample sizes $(\nu_\theta)_{\theta \in \Theta}$.

   Given environment $E = \left(A, \Theta, \beta, \psi_l, (\lambda_\theta, \nu_\theta)_{\theta \in \Theta}\right)$, let μ_l, the *mean sample size,* be redefined as the expected

number of actions observed by a random agent in the population. Formally:

$$\mu_l = \sum_{\theta \in \Theta} \lambda_\theta \sum_{l \in supp(\nu_\theta)} \nu_\theta(l) \cdot l.$$

Learning rule and stationary learning process. Each new θ-agent chooses his action in the new population state by following a stationary (i.e., time-independent) *learning rule* $\sigma_\theta : M \to \Delta(A)$. That is, a new θ-agent who observes sample $m \in M$ plays action a with probability $\sigma_{\theta,m}(a)$. The remaining $1 - \beta$ incumbent agents play the same action as in the previous round. A profile of learning rules $(\sigma_\theta)_{\theta \in \Theta}$ is *uniform* if all types use the same learning rule, i.e., if $\sigma_\theta = \sigma_{\theta'}$ for each type $\theta, \theta' \in \Theta$.

A stationary learning process (or *learning process* for short) is a pair

$$P = \left(E, (\sigma_\theta)_{\theta \in \Theta}\right) = \left(A, \Theta, \beta, \psi_l, (\lambda_\theta, \nu_\theta, \sigma_\theta)_{\theta \in \Theta}\right),$$

consisting of an environment and a learning rule.

As in the basic model, let $f_P : \Gamma \to \Gamma$ denote the mapping between states induced by a single step of the learning process P.

L_1**-distance.** Each population state $\gamma \in \Gamma$ corresponds to a distribution $q_\gamma \in (\Theta \times A)$ as follows: $q_\gamma(\theta, a) = \lambda_\theta \cdot \gamma_\theta(a)$. We define the distance between two population states $\gamma, \gamma' \in \Gamma$ as the L_1-distance between the corresponding distributions $q_\gamma, q_{\gamma;} \in (\Theta \times A)$:

$$\|\gamma - \gamma'\|_1 = \|q_\gamma - q_{\gamma'}\|_1 = \sum_{\theta \in \Theta} \sum_{a \in A} |\lambda_\theta \cdot \gamma_\theta(a) - \lambda_\theta \cdot \gamma'_\theta(a)| = \sum_{\theta \in \Theta} \lambda_\theta \cdot \|\gamma_\theta - \gamma'_\theta\|_1.$$

We demonstrate how the extended model formalizes the motivating Example 3, in which there are two types that differ in their sample sizes.

Example. 2 (Indirect Reciprocity revisited). The environment in which agents play the Prisoner's Dilemma and use rules representing indirect reciprocity is modeled by a learning process

$$P = \left(\{c, d\}, \Theta = \{\theta', \theta''\}, \beta = 1, \psi_l, (\lambda_\theta, \nu_\theta, \sigma_\theta)_{\theta \in \Theta}\right)$$

in which $A = \{c, d\}$ is the set of actions in the Prisoner's Dilemma, all agents revise their actions in each round (i.e., $\beta = 1$), and ψ_l is given by Eq. (5). The distribution of type is $\lambda_{\theta'} = 99\%$ and $\lambda_{\theta''} = 1\%$. The distribution of the sample size of each type is as follows: $\nu_{\theta'} \equiv 1$, and $\nu_{\theta'} \equiv 0$ in Case A and $\nu_{\theta'} \equiv 3$ in Case B. The initial population state is given by $\hat\gamma_{\theta'} = \alpha_1 \cdot d + (1 - \alpha_1) \cdot c$ and $\hat\gamma_{\theta''} = \alpha_2 \cdot d + (1 - \alpha_2) \cdot c$. Finally, the profile of learning rules is as follows: $\sigma_{\theta'}(a) = a$, and $\sigma_{\theta''}(\emptyset) = \alpha_2 \cdot c + (1 - \alpha_2) \cdot d$ in Case A, and

$$\sigma_{\theta''}(\overrightarrow{a}) = \begin{cases} c & \overrightarrow{a} \in \{(c, c, c), (c, c, d), (c, d, c), (d, c, c)\} \\ d & \text{otherwise} \end{cases}$$

in Case B. Observe that the expected number of actions is equal to 0.99 in Case A ($\mu_l = 0.99$), and $\mu_l = 1.02$ in Case B.

5.2 Generalizing the Main Results for Heterogeneous Populations

Our next result generalizes Theorem 1 to heterogeneous populations. It shows that the distance between two new states is at most $(1 - \beta \cdot (1 - \mu_l))$ times the distance between the two initial states. Formally,

Theorem 3. *(Generalization of Theorem 1) Let* $P = \left(A, \Theta, \beta, \psi_l, (\lambda_\theta, \nu_\theta, \sigma_\theta)_{\theta \in \Theta}\right)$ *be a stationary learning process, and let* $\gamma \neq \gamma' \in \Gamma$ *be two population states. Then:*

$$\|f_P(\gamma) - f_P(\gamma')\|_1 \leq (1 - \beta \cdot (1 - \mu_l)) \cdot \|\gamma - \gamma'\|_1,$$

with a strict inequality if there exist a type θ *and an* $l > 1$ *such that* $\nu_\theta(l) > 0$.

The intuition is similar to Theorem 1. The proof is presented in Appendix A.

Similarly to Section 4, Theorem 1 implies that f_P admits a global attractor if either (1) $\mu_l < 1$, or (2) $\mu_l = 1$ and there is a type $\theta \in \Theta$ such that $\nu_\theta(1) < 1$. Formally:

Corollary 2. *Let* $P = \left(A, \Theta, \beta, \psi_l, (\lambda_\theta, \nu_\theta, \sigma_\theta)_{\theta \in \Theta}\right)$ *be a stationary learning process satisfying (1)* $\mu_l < 1$, *or (2)* $\mu_l = 1$ *and there is a type* $\theta \in \Theta$ *such that[8]* $\nu_\theta(1) < 1$. *Then* f_P *is a weak contraction mapping, which implies that (1)* f_P *admits a unique steady state* γ^*, *and (2) this unique steady state* γ^* *is a global attractor (i.e.,* $\lim_{t \to \infty} f_P^t(\hat{\gamma}) = \gamma^*$ *for each* $\hat{\gamma} \in \Gamma$).

The following result generalizes Theorem 2 to the setup of heterogeneous populations. The proof is analogous to the proof of Theorem 2 and is omitted for brevity.

Theorem 4. *(Generalization of Theorem 2) Let* $E = \left(A, \Theta, \beta, \psi_l, (\lambda_\theta, \nu_\theta)_{\theta \in \Theta}\right)$ *be an environment. The following two conditions are equivalent:*

1. $\mu_l > 1$, *or* $\nu_\theta(1) = 1$ *for each type* $\theta \in \Theta$.

2. *There exists a profile of learning rules* $(\sigma_\theta)_{\theta \in \Theta}$, *such that the learning process* $\left(E, (\sigma_\theta)_{\theta \in \Theta}\right)$ *admits two different steady states.*

6 Non-Stationary Learning Process and Common Shocks

6.1 The Non-Stationary Process

In this section we adapt the model of Section 5 to deal with non-stationary deterministic learning processes, in which the process explicitly depends on calendar time. We extend our results to this setup.

Adaptations to the model. For each period $t \geq 1$, let $\beta_t \in [0, 1]$ denote the random share of agents who revise their actions in period t. For each type $\theta \in \Theta$ and period $t \geq 1$, let $\nu_\theta^t \in \Delta(\mathbb{N})$ denote the distribution of sample sizes of type θ in period t. To simplify the notation we assume that the support of the sample sizes of each type is independent of the period, i.e., $supp\left(\nu_\theta^{t_1}\right) = supp\left(\nu_\theta^{t_2}\right) := supp(\nu_\theta)$ for each type $\theta \in \Theta$ and

---

[8]Note that $\mu_l = 1$ and the existence of a type $\theta \in \Theta$ such that $\nu_\theta(1) < 1$, implies that there is a type $\theta \in \Theta$ and an $l > 1$ such that $\nu_\theta(l) > 0$

periods $t_1, t_2 \geq 1$. As in the basic model, let M denote the set of all feasible sample sizes. A *non-stationary environment* is a tuple

$$E = \left(A, \Theta, (\beta_t)_{t \in \mathbb{N}}, \psi_l, (\lambda_\theta)_{\theta \in \Theta}, \left(\nu_\theta^t\right)_{\theta \in \Theta, t \geq 1} \right).$$

Given a non-stationary environment, let μ_l^t denote the expected number of actions observed in period t, i.e., $\mu_l^t = \sum_{\theta \in \Theta} \lambda_\theta \sum_{l \in supp(\nu_\theta)} \nu_\theta^t(l) \cdot l$.

Given a non-stationary environment E, let $\bar{\mu}_l$ be the upper limit of the geometric mean of $1 - \beta^t \cdot (1 - \mu_l^t)$ as t goes to to infinity, i.e.,

$$\bar{\mu}_l = \mathrm{limsup}_{\hat{t} \to \infty} \sqrt[\hat{t}]{\prod_{t \leq t_0} (1 - \beta^t \cdot (1 - \mu_l^t))}.$$

For each type $\theta \in \Theta$ and period $t \geq 1$, let $\sigma_\theta^t : M \to \Delta(A)$ denote the non-stationary learning rule of new θ-agents in period t. A *non-stationary learning process* is a pair consisting of a non-stationary environment and a non-stationary learning rule, i.e.,

$$P = \left(E, \left(\sigma_\theta^t\right)_{\theta \in \Theta, t \geq 1} \right) = \left(A, \Theta, (\beta_t)_{t \geq 1}, \psi_l, (\lambda_\theta)_{\theta \in \Theta}, \left(\nu_\theta^t, \sigma_\theta^t\right)_{\theta \in \Theta, t \geq 1} \right).$$

As as in the basic model, a non-stationary learning process P and an initial state uniquely determine a new state in each period t. Let $f_p^t(\hat{\gamma}) \in \Gamma$ denote the state induced after t stages of the non-stationary learning process P.

A sequence of states $(\gamma_t^*)_{t \in N}$ is a *global attractor* of the non-stationary learning process P, if

$$\lim_{t \longrightarrow \infty} \left\| f_P^t(\hat{\gamma}) - \gamma_t^* \right\|_1 = 0$$

for each initial state $\hat{\gamma} \in \Gamma$.

Adapted results. Minor adaptations to the proof of Theorem 3 and a simple inductive argument immediately imply that the distance between two states at time t_o is at most $\prod_{t \leq t_0} (1 - \beta^t \cdot (1 - \mu_l^t))$ the initial distance. Formally:

Corollary 3. *Let* $P = \left(A, \Theta, (\beta_t)_{t \geq 1}, \psi_l, (\lambda_\theta)_{\theta \in \Theta}, (\nu_\theta^t, \sigma_\theta^t)_{\theta \in \Theta, t \geq 1} \right)$ *be a non-stationary learning process, let* $\hat{\gamma}, \hat{\gamma}' \in \Gamma$ *be two population states, and let* $\hat{t} \geq 1$. *Then:*

$$\left\| f_p^{\hat{t}}(\hat{\gamma}) - f_p^{\hat{t}}(\hat{\gamma}') \right\|_1 \leq \|\hat{\gamma} - \hat{\gamma}'\|_1 \cdot \prod_{t \leq \hat{t}} \left(1 - \beta^t \cdot \left(1 - \mu_l^t \right) \right) \cdot$$

This, in turn, immediately implies that in any non-stationary environment in which $\bar{\mu}_l < 1$, any profile of non-stationary learning rules admits a global attractor. Formally:

Corollary 4. *Let* $E = \left(A, \Theta, (\beta_t)_{t \geq 1}, \psi_l, (\lambda_\theta)_{\theta \in \Theta}, (\nu_\theta^t)_{\theta \in \Theta, t \geq 1} \right)$ *be a non-stationary environment satisfying* $\bar{\mu}_l < 1$. *Then for any profile of non-stationary learning rules* $(\sigma_\theta^t)_{\theta \in \Theta, t \geq 1}$, *the non-stationary learning process* $P = \left(E, (\sigma_\theta^t)_{\theta \in \Theta, t \geq 1} \right)$ *admits a global attractor.*

The example presented in Case A of the proof of Theorem 2 demonstrates that the above bound of $\bar{\mu}_l < 1$ is binding in the sense that there is an environment with $\bar{\mu}_l = 1$ that admits a profile of learning rules with multiple steady states.

16

6.2 Process with Common Shocks

In this section we further extend our model to deal also with common stochastic shocks to the learning rules, and we extend our results to this setup.

Additional adaptations to the model. In what follows we further adapt the model of Section 6.1 by allowing common stochastic shocks to the learning rules of the agents.

Let $(\Omega, \mathcal{F}, p)$ be an arbitrary probability space. Each element $\omega \in \Omega$ represents the state of nature, which determines the realizations of all common shocks to the learning rules in all periods. For each type $\theta \in \Theta$ and period $t \in \mathbb{N}$, let $\sigma_\theta^t : \Omega \times M \to \Delta(A)$ denote the state-dependent learning rule of new θ-agents in period t (which also depends on the state of nature). A *learning process* with common shocks is a pair consisting of a non-stationary environment and a state-dependent learning rule, i.e., $P = \left(E, (\sigma_\theta^t)_{\theta \in \Theta, t \geq 1}\right) = \left(A, \Theta, (\beta_t)_{t \geq 1}, \psi_l, (\lambda_\theta)_{\theta \in \Theta}, (\nu_\theta^t, \sigma_\theta^t)_{\theta \in \Theta, t \geq 1}\right)$.

Learning processes with commons shocks are important in modeling situations in which there are stochastic factors that influence the learning rules of all new agents in period t. For example , Ellison & Fudenberg (1995) model a situation in which new agents in period t choose between two agricultural technologies, and each such new agent observes a noisy signal about the expected payoff of each technology conditional on the weather in period t (which is common to all agents), where the (unknown) state of nature determines the weather in all periods.

The state of nature, the learning process, and the initial population state uniquely determine the population state in each period. Let $f_p^t(\omega)(\hat{\gamma}) \in \Gamma$ denote the population state induced after t stages of the non-stationary learning process P, given an initial population state $\hat{\gamma}$, and state of nature $\omega \in \Omega$.

We say that a sequence of state-dependent population states $(\gamma_t^*)_{t \geq 1}$, where $\gamma_t^* : \Omega \to \Gamma$, is a *state-dependent global attractor* of the learning process with commons shocks P if, for each $\omega \in \Omega$, $\lim_{t \to \infty} f_P^t(\omega)(\hat{\gamma}) = \gamma_t^*(\omega)$ for each initial state $\hat{\gamma} \in \Gamma$.

Example 4 below demonstrates how to apply the extended model to a social learning process with competing technologies with common shocks:

Example 4 (Competing Technologies with Common Shocks)**.** Consider a stochastic environment in which there are two possible regimes $\{1, 2\}$. There are two technologies: a_1 and a_2. Technology a_1 is advantageous in regime 1, while technology a_2 is advantageous in regime 2. There is a uniform common prior about the regime in round 1. In each subsequent round, the regime is the same as in the previous round with probability 99%, and it is a new regime with probability 1%. In each round, a share of 25% of the incumbents die, and are replaced with new agents. Each new agent observes the action of a single random incumbent and a noisy signal about the current regime, and based on these observations, the agent chooses one of the two technologies. Assume that the learning rule used by the agents implies that each new agent plays action a_1:

1. with a probability of 95% after observing action a_1 in regime 1;

2. with a probability of 80% after observing action a_1 in regime 2;

3. with a probability of 20% after observing action a_2 in regime 1;

4. with a probability of 5% after observing action a_2 in regime 2.

One can show that the environment admits a unique steady state that is a state-dependent global attractor. The induced aggregate behavior of the population converges towards playing action a_1 with an average probability of 80% in regime 1, and it converges towards playing action a_1 with an average probability of 20% in regime 2.

This learning process with common shocks is modeled as

$$P = \left(\{a_1, a_2\}, \{\theta\}, (\beta_t \equiv 25\%)_{t \in \mathbb{N}}, \psi_l, \lambda_\theta, \left(\nu_\theta^t \equiv 1, \sigma_\theta^t \right)_{t \geq 1} \right).$$

The set of states of nature $\Omega = \left\{ (\omega_n)_{n \in N} \right\}$ is the set of infinite binary sequences, where each $\omega_n \in \{1, 2\}$ describes the regime in round n. The definition of $(\mathcal{F}, p)$ is derived from the Markovian process determining the regime in each round in a standard way. Given state $\omega = (\omega_n)_{n \in N}$, let the learning rule be defined as follows:

$$\sigma_\theta (a_1, \omega) = \begin{cases} 95\% & a = a_1 \text{ and } \omega_t = 1 \\ 80\% & a = a_1 \text{ and } \omega_t = 2 \\ 20\% & a = a_2 \text{ and } \omega_t = 1 \\ 5\% & a = a_2 \text{ and } \omega_t = 2. \end{cases}$$

Adapted Results. Minor adaptations to the proof of Theorem 3 and a simple inductive argument immediately imply that the distance between two states at time $\hat{t}$ is at most $\prod_{t \leq \hat{t}} (1 - \beta^t \cdot (1 - \mu_l^t))$ the initial distance. Formally:

Corollary 5. *Let $P = \left(A, \Theta, (\beta_t)_{t \geq 1}, \psi_l, (\lambda_\theta)_{\theta \in \Theta}, (\nu_\theta^t, \sigma_\theta^t)_{\theta \in \Theta, t \geq 1} \right)$ be a learning process with commons shocks, let $\hat{\gamma}, \hat{\gamma}' \in \Gamma$ be two population states, and let $\hat{t} \in \mathbb{N}$. Then, for each $\omega \in \Omega$,*

$$\left\| f_p^{\hat{t}} (\omega) (\hat{\gamma}) - f_p^{\hat{t}} (\omega) (\hat{\gamma}') \right\|_1 \leq \| \hat{\gamma} - \hat{\gamma}' \|_1 \cdot \prod_{t \leq t_0} \left(1 - \beta^t \cdot \left(1 - \mu_l^t \right) \right).$$

An immediate corollary of Corollary 5 is that any environment with common shocks in which $\bar{\mu}_l < 1$, given any profile of learning rules, admits a stochastic global attractor. That is, in the long run, the population's behavior depends only on the state of nature, but it is independent of the initial population state in time zero. Formally:

Corollary 6. *Let $E = \left(A, \Theta, (\beta_t)_{t \geq 1}, \psi_l, (\lambda_\theta)_{\theta \in \Theta}, (\nu_\theta^t)_{\theta \in \Theta, t \geq 1} \right)$ be an environment satisfying $\bar{\mu}_l < 1$. Then for any profile of stochastic learning rules $(\sigma_\theta^t)_{\theta \in \Theta, t \geq 1}$, the learning process with common shocks $P = \left(E, (\sigma_\theta^t)_{\theta \in \Theta, t \geq 1} \right)$ admits a global stochastic attractor.*

7   Responsiveness and Effective Sample Size

Our results so far have not assumed anything about the agents' learning rules. Obviously, additional information on the profile of learning rules, may allow us to achieve stronger results. In this section, we present simple notions of responsiveness and expected effective sample size, and use them to derive a tighter upper bound for processes that admit global attractors.

7.1 Additional Definitions

Fix a stationary learning process $P = \left(A, \Theta, \beta, \psi_l, (\lambda_\theta, \nu_\theta, \sigma_\theta)_{\theta \in \Theta} \right)$. For each type θ, each sample size $l \in supp(\nu)$, and each action $a \in A$, let $\underline{\sigma}_{\theta,l}(a)$ $(\overline{\sigma}_{\theta,l}(a))$ be the minimal (maximal) probability that learning rule σ assigns to action a after observing a sample of size l, i.e.,

$$\underline{\sigma}_{\theta,l}(a) = \min_{m \in A^l} \sigma_{\theta,m}(a) \qquad (\overline{\sigma}_{\theta,l}(a) = \max_{m \in A^l} \sigma_{\theta,m}(a)).$$

Let $r_{\theta,l}$ denote the maximal *responsiveness* of new θ-agents to changes in observed samples of size l, which is defined as follows:

$$r_{\theta,l} = \min \left(1, \ \frac{1}{2} \cdot \sum_{a \in A} \left(\overline{\sigma}_{\theta,l}(a) - \underline{\sigma}_{\theta,l}(a) \right) \right), \tag{6}$$

and let $r_{\theta,0} = 0$. The responsiveness effectively limits the maximal influence of different samples of length l on the behavior of θ-agents to be at most $r_{\theta,l} \leq 1$. Observe that when there are two actions (i.e., $A = \{a, b\}$), then $r_{\theta,l}$ is simply the difference between the maximal and minimal probability assigned to each action, i.e.,

$$r_{\theta,l} = \overline{\sigma}_{\theta,l}(a) - \underline{\sigma}_{\theta,l}(a) = \overline{\sigma}_{\theta,l}(b) - \underline{\sigma}_{\theta,l}(b) \qquad (A = \{a, b\}). \tag{7}$$

When there are more than two actions, $\frac{1}{2} \cdot \sum_{a \in A} \left(\overline{\sigma}_{\theta,l}(a) - \underline{\sigma}_{\theta,l}(a) \right)$ may be larger than one. We bound $r_{\theta,l}$ from above by one in Eq.(6) because, any change of sample cannot affect an agent's mixed behavior by more than one (as measured by the L_1-distance over the set of mixed actions).

We call the product of the sample size and the responsiveness, $r_{\theta,l} \cdot l$ the *effective sample size*. For each type θ, let $\mu_\theta^e \in \mathbb{R}^+$ denote the *expected effective sample size of θ-agents*, i.e.,

$$\mu_\theta^e = \sum_{l \in supp(\nu_\theta)} \nu_\theta(l) \cdot r_{\theta,l} \cdot l.$$

Finally, let $\mu \in \mathbb{R}^+$ denote the *effective sample size* in the population, i.e.,

$$\mu_l^e = \sum_{\theta \in \Theta} \lambda_\theta \cdot \mu_\theta^e.$$

It is immediate that the effective sample size is always weakly smaller than the expected sample size in the population; i.e., $\mu_l^e \leq \mu_l$ for each stationary learning process s $P = \left(A, \Theta, \beta, \psi_l, (\lambda_\theta, \nu_\theta, \sigma_\theta)_{\theta \in \Theta} \right)$.

7.2 A Tighter Bound on the Distance between New States

Our main result in this section shows that the distance between two new states is at most $(1 - \beta \cdot (1 - \mu_l^e))$ times the distance between the two initial states. This bound is tighter than the one presented in Theorem 1, as we replace expected sample size μ_l with the (weakly) smaller effective sample size μ_l^e. Formally,

Theorem 5. *Let* $P = \left(A, \Theta, \beta, \psi_l, (\lambda_\theta, \nu_\theta, \sigma_\theta)_{\theta \in \Theta} \right)$ *be a stationary learning process, and let* $\gamma \neq \gamma' \in \Gamma$ *be two population states. Then:*

$$\| f_P(\gamma) - f_P(\gamma') \|_1 \leq (1 - \beta \cdot (1 - \mu_l^e)) \cdot \| \gamma - \gamma' \|_1,$$

where the inequality is strict if there exist a type θ and an $l > 1$ such that $\nu_\theta(l) > 0$.

Proof. The key step of the proof is to show the following inequality:

$$\|(f_P(\gamma)) - (f_P(\gamma'))\|_1 \leq \beta \cdot \sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in \mathbb{N}} \nu_\theta(l) \cdot r_{\theta,l} \cdot \|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 + (1-\beta) \cdot \|\gamma - \gamma'\|_1. \qquad (8)$$

Inequality (8) is the same as (2) in the proof of Theorem 1, except for the factor of $r_{\theta,l} \leq 1$ on the RHS. All other arguments of the proof of Theorem 1 remain the same. We prove (8) in Lemma 6 in Appendix B.

Observe that $(1 - \beta \cdot (1 - \mu_l^e)) < 1$ iff $\mu_l^e < 1$, and in this case f_P is a contraction mapping, which implies that f_P admits a global attractor. This allows us to strengthen Corollary 1 as follows. $\qquad \square$

Corollary 7. *Let* $P = \left(A, \Theta, \beta, \psi_l, (\lambda_\theta, \nu_\theta, \sigma_\theta)_{\theta \in \Theta}\right)$ *be a learning process satisfying (1) $\mu_l^e < 1$, or (2) $\mu_l^e = 1$ and there is a type $\theta \in \Theta$ such that $\nu_\theta(1) < 1$. Then f_P is a contraction mapping, which implies that (1) f_P admits a unique steady state γ^*, and (2) this unique steady state γ^* is a global attractor (i.e., $\lim_{t \to \infty} f_P^t(\hat{\gamma}) = \gamma^*$ for each $\hat{\gamma} \in \Gamma$).*

We demonstrate the implications of Corollary 7 by revisiting Example **1.**

Example. 1 (Competing Technologies revisited). Recall, that each agent observes a single action, which implies that the expected number of observed actions is one (i.e., $\mu_l = 1$). Hence the results of the previous sections do not imply the that the learning process admits a unique steady state. To simplify notation, assume that $\overline{\alpha} > \underline{\alpha}$. Observe that the effective number of observations, μ_l^e, is equal to:

$$\mu_l^e = \mu_\theta^e = r_{\theta,l=1} \cdot 1 = \frac{1}{2} \cdot \sum_{a \in A} \left(\overline{\sigma}_{\theta,l}(a) - \underline{\sigma}_{\theta,l}(a)\right) = \frac{1}{2} \cdot \left((\overline{\alpha} - \underline{\alpha}) + ((1 - \underline{\alpha}) - (1 - \overline{\alpha}))\right) = \overline{\alpha} - \underline{\alpha},$$

which is strictly less than one if $\overline{\alpha} < 1$ or $\underline{\alpha} > 0$. Corollary 7 implies that the learning process converges to a global attractor (which is the unique steady state) whenever $\overline{\alpha} < 1$ or $\underline{\alpha} > 0$.

Our final result demonstrates that our bound of the effective sample size being less than one is tight. Specifically, it shows that given any environment in which the expected sample size $\mu_l > 1$, and any number $1 < y \leq \mu_l$, there is a learning rule with an effective sample size of $\mu_l^e = y$ with multiple steady states. Formally:

Proposition 1. *Let* $E = \left(A, \Theta, \beta, \psi_l, (\lambda_\theta, \nu_\theta)_{\theta \in \Theta}\right)$ *be an environment satisfying $\mu_l > 1$. Let $1 < y \leq \mu_l$. Then there exists a profile of learning rules $(\sigma_\theta)_{\theta \in \Theta}$, such that the learning process $\left(E, (\sigma_\theta)_{\theta \in \Theta}\right)$ admits two different steady states, and satisfies $\mu_l^e = y$.*

Proof. Let a and a' be different actions ($a \neq a' \in A$). Let $(\sigma_\theta^*)_{\theta \in \Theta}$ be a uniform learning process according to which each agent plays action a^* with a probability of $\frac{y}{\mu_l}$ if he has observed action a^* at least once, and plays action a' otherwise, that is,

$$\sigma_\theta^*(a^l) = \begin{cases} \frac{y}{\mu_l} \cdot a^* + \left(1 - \frac{x}{\mu_l}\right) \cdot a' & \text{if } \exists i, \text{ s.t., } a_i^l = a^* \\ a' & \text{otherwise.} \end{cases}$$

Observe that the responsiveness of $\left(E, (\sigma_\theta^*)_{\theta \in \Theta}\right)$ is equal to x because:

$$\mu_l^e = \sum_{\theta \in \Theta} \lambda_\theta \cdot \mu_\theta^e = \sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in supp(\nu_\theta)} \nu_\theta(l) \cdot r_{\theta,l} \cdot l = \sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in supp(\nu_\theta)} \nu_\theta(l) \cdot \frac{1}{2} \cdot \sum_{a \in A} \left(\overline{\sigma}_{\theta,l}(a) - \underline{\sigma}_{\theta,l}(a)\right) \cdot l =$$

20

$$\sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in supp(\nu_\theta)} \nu_\theta\left(l\right) \cdot \frac{1}{2} \cdot \left(\left(\frac{y}{\mu_l} - 0\right) + \left(1 - \left(1 - \frac{y}{\mu_l}\right) + 0 + ... + 0\right)\right) \cdot l =$$

$$\sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in supp(\nu_\theta)} \nu_\theta\left(l\right) \cdot \frac{y}{\mu_l} \cdot l = \frac{y}{\mu_l} \cdot \sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in supp(\nu_\theta)} \nu_\theta\left(l\right) \cdot l = \frac{y}{\mu_l} \cdot \mu_l = y.$$

It is immediate that the uniform population state in which all agents play action a' (i.e., $\gamma_\theta\left(a'\right) = 1$ for each type $\theta \in \Theta$) is a steady state of the learning process $\left(E, (\sigma_\theta^*)_{\theta \in \Theta}\right)$. An analogous argument to the one presented in Case B of the proof of Theorem 2 shows that there exists $x > 0$ such that the uniform population state γ^x in which all agents play action a^* with probability x, and play action a' with the remaining probability of $1 - x$, is another steady state of the learning process $\left(E, (\sigma_\theta^*)_{\theta \in \Theta}\right)$. $\qquad\square$

8 Concluding Remarks

Repeated Interactions without Calendar Time. In many real-life situations agents are randomly matched within a community, and these interactions have been going on since time immemorial. Modelling such situations as repeated games with a definite starting point and strategies that can be conditioned on calendar time may be a problematic modelling choice, as it seems implausible that agents would be aware of the the exact time that has transpired since the starting point, and aware of the very distant history of play of other agents. An alternative approach, is to model behavior in such situations as steady states of environments without a calendar time (see, e.g., (Rosenthal, 1979; Okuno-Fujiwara & Postlewaite, 1995; Heller & Mohlin, 2017), and the working paper version of Phelan & Skrzypacz, 2006).

An interesting question about such environments is whether the distribution of strategies used by the players to choose their actions as a function of their observations is sufficient to uniquely determine the steady states, or whether the same distribution of rules may admit multiple steady states. Our main result shows that the former is true whenever the expected number of observed actions is less than one, while if the expected number of observed actions is more than one, then there is always a distribution of rules with multiple steady states.

Multiple Locally Stable Steady States. Recall that steady state γ^* is locally (asymptotically) stable if a population starting from any sufficiently close initial state converges to γ^*. Our main result (Theorem 2) shows that an environment admits a profile of learning rules with multiple steady states if and only if $\mu_l \leq 1$ and some agents sometimes observe more than one action. The construction we present to demonstrate the existence of a learning rule with multiple steady states for any environment in which $\mu_l > 1$ has one unappealing property. The construction shows that both γ_0 (the state in which everyone plays action a') and γ_x for some $x > 0$ (the state in which agents on average play action a^* with a probability of x, and play action a' otherwise) are steady states. However, one can see that state γ_x is locally stable. By contrast, any small perturbation in the population's state will move the population away from γ_0 (and towards γ_x).

An interesting problem, which we leave for future research, is to characterize necessary and sufficient conditions for when an environment admits a learning rule with multiple *locally stable* states. Preliminary analysis suggests that the expected number of observations might have to be significantly above 1 to allow multiple locally stable steady states. In particular, for environments in which each agent chooses between two

actions, it can be shown that: (1) when agents observe at most two actions, at most one steady state is locally stable, and (2) when agents observe 3 actions with probability p and no actions otherwise, the environment may admit multiple locally stable steady states only if $p \geq \frac{2}{3}$ (i.e., if $\mu_l = 3 \cdot p \geq 2$).

Large Finite Populations. Our model studies infinite populations, and it is important to know what are the implications of our results for large finite populations. The key difference between an infinite and a finite population, is that in the former, the law of large numbers implies that the new state of the population is a deterministic function of the initial state and the learning rule (assuming that the environment does not have common stochastic shocks). By contrast, in finite populations the new population state is a random variable. If the finite population is sufficiently large then we expect the resulting stochastic process to be close to the deterministic process over finite time horizons. However, when time goes to infinity, rare random events will occasionally take the population away from one (locally stable) steady state towards another steady state.[9]

When dealing with finite large populations, one should therefore interpret our main result (Theorem 2) as follows. In environments in which $\mu_l < 1$, all learning processes admit a unique globally stable state γ^*. The population will quickly converge to state γ^*, and will almost always remain very close to this state. A rare event in which the realized observations of many agents substantially differ from their expected values, may take the population temporarily away from γ^*, but with a very high probability the population will quickly converge back to γ^*.

In environments in which $\mu_l > 1$, there are learning rules that admit multiple steady states. The fact that the population is finite and that the new population state is a random variable will typically quickly take the population away from steady states that are not locally stable. If the environment admits multiple locally stable steady states, then the initial state will determine which of these locally stable states the population will converge to in the medium run. Moreover the process will likely stay there for a significant amount of time.

The literature on stochastic evolutionary game theory (starting with the pioneering works of Foster & Young, 1990; Kandori *et al.*, 1993; Young, 1993a; see Young, 2015, for a recent survey) studies situations the long-run behavior in environments with multiple locally stable steady states, and in which there is a small level of noise in the agents' behavior. We think that it would be interesting to extend the methodology of this literature in order to apply it to the setup analyzed in this paper. It might be that such future research can characterize various cases in which, if the population size is sufficiently large, in the long run the population will spend almost all of the time in one of these locally stable states.

Observations of Action Profiles. In Heller & Mohlin (2017) we investigate environments in which an agent may observe action profiles played in past interactions between his current opponent and her past opponents. All of our results can be extended to this setup, with relatively minor adjustments to the proofs. Specifically one should count an observation of an action profile (in a two-player game) as two actions when calculating the expected number of observed actions μ_l. Our main result still holds in this setup: an environment admits a profile of learning rules with multiple steady states, essentially, if and only if $\mu_l \leq 1$.

---

[9]For a textbook overview of the deterministic approximation of stochastic evolutionary process we refer to Sandholm (2011).

A   Proof of Theorem 3

The distance between the final population states is bounded as follows (where the second inequality is strict if $\nu_\theta(l) > 0$ for some $\theta \in \Theta$ and $l \geq 2$):

$$\|(f_P(\gamma))_\theta - (f_P(\gamma'))_\theta\|_1 \leq \beta \cdot \sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in \mathbb{N}} \nu_\theta(l) \cdot \|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 + (1-\beta) \cdot \|\gamma - \gamma'\|_1 \leq$$

$$\beta \cdot \sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in \mathbb{N}} \nu_\theta(l) \cdot l \cdot \|\gamma - \gamma'\|_1 + (1-\beta) \cdot \|\gamma - \gamma'\|_1 =$$

$$\left(\beta \cdot \left(\sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in \mathbb{N}} \nu_\theta(l) \cdot l\right) + (1-\beta)\right) \cdot \|\gamma - \gamma'\| = (\beta \cdot \mu_L + 1 - \beta) \cdot \|\gamma - \gamma'\| = (1 - \beta \cdot (1 - \mu_l)) \cdot \|\gamma - \gamma'\|_1.$$

The first inequality is proven in Lemma 1. The second inequality (is strict if $\nu_\theta(l) > 0$ for some $\theta \in \Theta$ and $l \geq 2$) is implied by the inequality

$$\|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 \leq l \cdot \|\gamma - \gamma'\|_1 \ \text{ (with a strict inequality if } l \geq 2),$$

which is proven in Lemma 4.

Proofs of the various Lemmas used in the Proof of Theorem 3

Lemma 1. *For each learning environment E and states $\gamma \neq \gamma' \in \Gamma$,*

$$\|(f_P(\gamma)) - (f_P(\gamma'))\|_1 \leq \beta \cdot \sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in \mathbb{N}} \nu_\theta(l) \cdot \|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 + (1-\beta) \cdot \|\gamma - \gamma'\|_1.$$

Proof.

$$\|(f_P(\gamma)) - (f_P(\gamma'))\|_1 = \sum_{\theta \in \Theta} \lambda_\theta \cdot \|(f_P(\gamma))_\theta - (f_P(\gamma'))_\theta\|_1 \leq$$

$$\sum_{\theta \in \Theta} \lambda_\theta \cdot \left(\beta \cdot \sum_{l \in \mathbb{N}} \nu_\theta(l) \cdot \|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 + (1-\beta) \cdot \|\gamma_\theta - \gamma'_\theta\|_1\right) =$$

$$\beta \cdot \sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in \mathbb{N}} \nu_\theta(l) \cdot \|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 + (1-\beta) \cdot \sum_{\theta \in \Theta} \lambda_\theta \cdot \|\gamma_\theta - \gamma'_{\theta'}\|_1 =$$

$$\beta \cdot \sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{l \in \mathbb{N}} \nu_\theta(l) \cdot \|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 + (1-\beta) \cdot \|\gamma - \gamma'\|_1,$$

where the inequality is due to Lemma 2. $\qquad\square$

Lemma 2. *For each social learning environment E, type $\theta \in \Theta$, and each two states $\gamma \neq \gamma' \in \Gamma$:*

$$\|(f_P(\gamma))_\theta - (f_P(\gamma'))_\theta\|_1 \leq \beta \cdot \sum_{l \in \mathbb{N}} \nu_\theta(l) \cdot \|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 + (1-\beta) \cdot \|\gamma_\theta - \gamma'_\theta\|_1.$$

Proof.

$$\|(f_P(\gamma))_\theta - (f_P(\gamma'))_\theta\|_1 = \sum_{a \in A} |(f_P(\gamma))_\theta(a) - (f_P(\gamma'))_\theta(a)| =$$

23

$$\sum_{a \in A} \left| \left(\sum_{l \in supp(\nu_\theta)} \beta \cdot \nu_\theta(l) \sum_{m \in A^l} \psi_{l,\gamma}(m) \cdot \sigma_{\theta,m} + (1-\beta) \cdot \gamma_\theta \right) (a) \right.$$

$$\left. - \left(\beta \cdot \sum_{l \in supp(\nu_\theta)} \nu_\theta(l) \cdot \sum_{m \in A^l} \psi_{l,\gamma'}(m) \cdot \sigma_{\theta,m} + (1-\beta) \cdot \gamma'_\theta \right) (a) \right| =$$

$$\sum_{a \in A} \left| \beta \cdot \sum_{l \in supp(\nu_\theta)} \nu_\theta(l) \cdot \sum_{m \in A^l} (\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m)) \cdot \sigma_{\theta,m}(a) + (1-\beta) \cdot (\gamma_\theta(a) - \gamma'_\theta(a)) \right| \leq \qquad (9)$$

$$\sum_{a \in A} \left(\beta \cdot \sum_{l \in supp(\nu_\theta)} \nu_\theta(l) \cdot \left| \sum_{m \in A^l} (\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m)) \cdot \sigma_{\theta,m}(a) \right| + (1-\beta) \cdot |\gamma_\theta(a) - \gamma'_\theta(a)| \right) =$$

$$\beta \cdot \sum_{l \in supp(\nu_\theta)} \nu_\theta(l) \cdot \sum_{a \in A} \left| \sum_{m \in A^l} (\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m)) \cdot \sigma_{\theta,m}(a) \right| + (1-\beta) \cdot \sum_{a \in A} |\gamma_\theta(a) - \gamma'_\theta(a)| \leq \qquad (10)$$

$$\beta \cdot \sum_{l \in supp(\nu_\theta)} \nu_\theta(l) \cdot \|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 + (1-\beta) \cdot \|\gamma_\theta - \gamma'_{\theta'}\|_1,$$

where the (9) is a triangle inequality, and (10) is due to Lemma 3. $\qquad \square$

Lemma 3. *For each social learning environment E, each size $l \in \mathbb{N}$, each type $\theta \in \Theta$, and any two states $\gamma \neq \gamma' \in \Gamma$:*

$$\sum_{a \in A} \left| \sum_{m \in A^l} (\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m)) \cdot \sigma_{\theta,m}(a) \right| \leq \|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1.$$

Proof.

$$\sum_{a \in A} \left| \sum_{m \in A^l} (\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m)) \cdot \sigma_{\theta,m}(a) \right| \leq \sum_{a \in A} \sum_{m \in A^l} |\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m)| \cdot \sigma_{\theta,m}(a)$$

$$= \sum_{m \in A^l} |\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m)| \cdot \sum_{a \in A} \sigma_{\theta,m}(a)$$

$$= \sum_{m \in A^l} |\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m)| \cdot 1,$$

where the inequality is a triangle inequality. $\qquad \square$

Lemma 4. *For each social learning environment E, type $\theta \in \Theta$, sample size $l \in \mathbb{N}$, and states $\gamma \neq \gamma' \in \Gamma$*

$$\|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 \leq l \cdot \|\gamma - \gamma'\|_1,$$

with a strict inequality if $l > 1$.

Proof. Case I - Observing different random agents:

$$\|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 = \sum_{\overrightarrow{a} \in A^l} |\psi_{l,\gamma}(\overrightarrow{a}) - \psi_{l,\gamma'}(\overrightarrow{a})| = \tag{11}$$

$$\sum_{\overrightarrow{a} \in A^l} \left| \prod_{1 \le i \le l} \bar{\gamma}(a_i) - \prod_{1 \le i \le l} \bar{\gamma}'(a_i) \right| = \tag{12}$$

$$\sum_{\overrightarrow{a} \in A^l} \left| \sum_{1 \le i \le l} (\bar{\gamma}(a_i) - \bar{\gamma}'(a_i)) \cdot \prod_{i < j \le l} \bar{\gamma}(a_j) \cdot \prod_{1 \le k < i} \bar{\gamma}'(a_k) \right| \le \quad (< \text{ if } l > 1) \tag{13}$$

$$\sum_{\overrightarrow{a} \in A^l} \left(\sum_{1 \le i \le l} |\bar{\gamma}(a_i) - \bar{\gamma}'(a_i)| \cdot \prod_{i < j \le l} \bar{\gamma}(a_j) \cdot \prod_{1 \le k < i} \bar{\gamma}'(a_k) \right) =$$

$$\sum_{1 \le i \le l} \left(\sum_{\overrightarrow{a} \in A^l} |\bar{\gamma}(a_i) - \bar{\gamma}'(a_i)| \cdot \prod_{i < j \le l} \bar{\gamma}(a_j) \cdot \prod_{1 \le k < i} \bar{\gamma}'(a_k) \right) =$$

$$\sum_{1 \le i \le l} \left(\sum_{a_i \in A} |\bar{\gamma}(a_i) - \bar{\gamma}'(a_i)| \right) \cdot \left(\sum_{(a_{i+1},\dots,a_l) \in A^{l-i}} \prod_{i < j \le l} \bar{\gamma}(a_j) \right) \cdot \left(\sum_{(a_1,\dots,a_{i-1}) \in A^{i-1}} \prod_{1 \le k < i} \bar{\gamma}'(a_k) \right) = \tag{14}$$

$$\sum_{1 \le i \le l} \left(\sum_{a_i \in A} |\bar{\gamma}(a_i) - \bar{\gamma}'(a_i)| \right) \cdot 1 \cdot 1 = \sum_{1 \le i \le l} (\|\bar{\gamma} - \bar{\gamma}'\|_1) = l \cdot \|\bar{\gamma} - \bar{\gamma}'\|_1 \le l \cdot \|\gamma - \gamma'\|.$$

Eq. (11) is due to the independence of different observations. Eq. (12) is implied by adding to the sum elements that cancel out. Specifically, let $b_i = \bar{\gamma}(a_i)$ and $c_i = \bar{\gamma}'(a_i)$; then due to a "telescoping series" argument (in which each new element appears once with a positive sign and once with a negative sign):[10]

$$\prod_{1 \le i \le l} \bar{\gamma}(a_i) - \prod_{1 \le i \le l} \bar{\gamma}'(a_i) = \prod_{1 \le i \le l} b_i - \prod_{1 \le i \le l} c_i =$$

$$(b_1 \cdot \dots \cdot b_l - c_1 \cdot b_2 \cdot \dots \cdot b_l) + (c_1 \cdot b_2 \cdot \dots \cdot b_l + c_1 \cdot c_2 \cdot b_3 \cdot \dots \cdot b_l) - c_1 \cdot c_2 \cdot b_3 \cdot \dots \cdot b_l + \dots + c_1 \cdot \dots \cdot c_l =$$

$$(b_1 - c_1) \cdot b_2 \cdot \dots \cdot b_l + (b_2 - c_2) \cdot b_3 \cdot \dots \cdot b_l \cdot c_1 + (b_3 - c_3) \cdot b_4 \cdot \dots \cdot b_l \cdot c_1 \cdot c_2 \dots + (b_l - c_l) \cdot c_2 \cdot \dots \cdot c_l =$$

$$= \sum_{1 \le i \le l} \left((b_i - c_i) \cdot \prod_{i < j \le l} b_j \cdot \prod_{1 \le j < i} c_j \right) = \sum_{1 \le i \le l} (\bar{\gamma}(a_i) - \bar{\gamma}'(a_i)) \cdot \prod_{i < j \le l} \bar{\gamma}(a_j) \cdot \prod_{1 \le k < i} \bar{\gamma}'(a_k).$$

Eq. (13) is a triangle inequality, and it is strict if $l > 1$ because the sum inside the "$\left| \; \right|$" in (13) includes both positive and negative elements. Eq. (14) holds because each sum adds the probabilities of disjoint and exhausting events. The final inequality is implied by Lemma 5.

---

[10]We use the convention that a product of an empty set (e.g., $\prod_{1 \le j < 1}$) is equal to one.

Case II - Observing a single random type:

$$\|\psi_{l,\gamma} - \psi_{l,\gamma'}\|_1 = \sum_{\overrightarrow{a} \in A^l} |\psi_{l,\gamma}(\overrightarrow{a}) - \psi_{l,\gamma'}(\overrightarrow{a})| = \tag{15}$$

$$\sum_{\overrightarrow{a} \in A^l} \left| \sum_{\theta \in \Theta} \lambda_\theta \cdot \left(\prod_{1 \leq i \leq l} \gamma_\theta(a_i) - \prod_{1 \leq i \leq l} \gamma'_\theta(a_i) \right) \right| = \tag{16}$$

$$\sum_{\overrightarrow{a} \in A^l} \left| \sum_{\theta \in \Theta} \lambda_\theta \cdot \left(\sum_{1 \leq i \leq l} (\gamma_\theta(a_i) - \gamma'_\theta(a_i)) \cdot \prod_{i < j \leq l} \gamma_\theta(a_j) \cdot \prod_{1 \leq j < i} \gamma'_\theta(a_j) \right) \right| \leq \quad (< \text{ if } l > 1) \tag{17}$$

$$\sum_{\overrightarrow{a} \in A^l} \sum_{\theta \in \Theta} \lambda_\theta \cdot \left(\sum_{1 \leq i \leq l} |\gamma_\theta(a_i) - \gamma'_\theta(a_i)| \cdot \prod_{i < j \leq l} \gamma_\theta(a_j) \cdot \prod_{1 \leq j < i} \gamma'_\theta(a_j) \right) =$$

$$\sum_{1 \leq i \leq l} \sum_{\theta \in \Theta} \lambda_\theta \cdot \left(\sum_{\overrightarrow{a} \in A^l} |\gamma_\theta(a_i) - \gamma'_\theta(a_i)| \cdot \prod_{i < j \leq l} \gamma_\theta(a_j) \cdot \prod_{1 \leq j < i} \gamma'_\theta(a_j) \right) =$$

$$\sum_{1 \leq i \leq l} \sum_{\theta \in \Theta} \lambda_\theta \cdot \left(\sum_{a_i \in A} |\gamma_\theta(a_i) - \bar{\gamma}_\theta{}'(a_i)| \right) \cdot \left(\sum_{(a_{i+1},\ldots,a_l) \in A^{l-i}} \prod_{i < j \leq l} \gamma_\theta(a_j) \right) \cdot \left(\sum_{(a_i,\ldots,a_{i-1}) \in A^{i-1}} \prod_{1 \leq j < i} \gamma'_\theta(a_j) \right) = \tag{18}$$

$$\sum_{1 \leq i \leq l} \sum_{\theta \in \Theta} \lambda_\theta \cdot \left(\sum_{a_i \in A} |\gamma_\theta(a_i) - \bar{\gamma}_\theta{}'(a_i)| \right) \cdot 1 \cdot 1 = \sum_{1 \leq i \leq l} \sum_{\theta \in \Theta} \lambda_\theta \cdot \|\gamma_\theta - \gamma'_\theta\|_1 = \sum_{1 \leq i \leq l} \|\gamma - \gamma'\|_1 = l \cdot \|\gamma - \gamma'\|_1.$$

Eq. (15) is due to the different observations being independent conditional on the observed type θ. Eq. (16) is implied by adding to the sum elements that cancel out (i.e., a "telescoping series"). Eq. (17) is a triangle inequality, and it is strict if $l > 1$ because the sum inside the "$\left| \ \right|$" in (17) includes both positive and negative elements. Eq. (18) holds because each sum adds the probabilities of disjoint and exhausting events. $\qquad \square$

Lemma 5. $\|\bar{\gamma} - \bar{\gamma}'\|_1 \leq \|\gamma - \gamma'\|_1$ for each two states $\gamma \neq \gamma' \in \Gamma$.

Proof.

$$\|\gamma - \gamma'\|_1 = \sum_{\theta \in \Theta} \lambda_\theta \cdot \|\gamma_\theta - \gamma'_\theta\|_1 = \sum_{\theta \in \Theta} \lambda_\theta \cdot \sum_{a \in A} |\gamma_\theta(a) - \gamma'_\theta(a)| = \sum_{a \in A} \sum_{\theta \in \Theta} \lambda_\theta \cdot |\gamma_\theta(a) - \gamma'_\theta(a)| \geq$$

$$\sum_{a \in A} \left| \sum_{\theta \in \Theta} \lambda_\theta (\gamma_\theta(a) - \gamma'_\theta(a)) \right| = \sum_{a \in A} \left| \sum_{\theta \in \Theta} \lambda_\theta \gamma_\theta(a) - \sum_{\theta \in \Theta} \lambda_\theta \gamma'_\theta(a) \right| = \sum_{a \in A} |\bar{\gamma}(a) - \bar{\gamma}'(a)| = \|\bar{\gamma} - \bar{\gamma}'\|_1 \,,$$

where the various equalities are immediately implied by the definitions on the L_1-norm and $\bar{\gamma}$, and the inequality is a triangle inequality. $\qquad \square$

B Lemma Required for the Proof of Theorem 5

Lemma 6. *For each social learning environment E, each size $l \in \mathbb{N}$, each type $\theta \in \Theta$, and any two states $\gamma \neq \gamma' \in \Gamma$:*

$$\sum_{a \in A} \left| \sum_{m \in A^l} \left(\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m) \right) \cdot \sigma_{\theta,m}(a) \right| \leq r_{\theta,l} \cdot \| \psi_{l,\gamma} - \psi_{l,\gamma'} \|_1 .$$

Proof. We begin with a preliminary definition. Let $A^l_{\gamma > \gamma'} \subseteq A^l$ be the set of samples that have higher probabilities given state γ than given state γ', i.e.,

$$A^l_{\gamma > \gamma'} = \left\{ m \in A^l | \psi_{l,\gamma}(m) > \psi_{l,\gamma'}(m) \right\}.$$

We now prove the lemma:

$$\sum_{a \in A} \left| \sum_{m \in A^l} \left(\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m) \right) \cdot \sigma_{\theta,m}(a) \right| =$$

$$\sum_{a \in A} \left| \sum_{m \in A^l_{\gamma > \gamma'}} \left(\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m) \right) \cdot \sigma_{\theta,m}(a) - \sum_{m \in A^l_{\gamma' > \gamma}} \left(\psi_{l,\gamma'}(m) - \psi_{l,\gamma}(m) \right) \cdot \sigma_{\theta,m}(a) \right| \leq$$

$$\sum_{a \in A} \left| \sum_{m \in A^l_{\gamma > \gamma'}} \left(\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m) \right) \cdot \overline{\sigma}_{\theta,l}(a) - \sum_{m \in A^l_{\gamma' > \gamma}} \left(\psi_{l,\gamma'}(m) - \psi_{l,\gamma}(m) \right) \cdot \underline{\sigma}_{\theta,l}(a) \right| =$$

$$\sum_{a \in A} \left| \overline{\sigma}_{\theta,l}(a) \cdot \sum_{m \in A^l_{\gamma > \gamma'}} \left(\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m) \right) - \underline{\sigma}_{\theta,l}(a) \cdot \sum_{m \in A^l_{\gamma' > \gamma}} \left(\psi_{l,\gamma'}(m) - \psi_{l,\gamma}(m) \right) \right| = \qquad (19)$$

$$\sum_{a \in A} \left| \left(\overline{\sigma}_{\theta,l}(a) - \underline{\sigma}_{\theta,l}(a) \right) \cdot \sum_{m \in A^l_{\gamma > \gamma'}} \left(\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m) \right) \right| =$$

$$\sum_{a \in A} \left(\overline{\sigma}_{\theta,l}(a) - \underline{\sigma}_{\theta,l}(a) \right) \cdot \sum_{m \in A^l_{\gamma > \gamma'}} \left(\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m) \right) =$$

$$\sum_{a \in A} \left(\overline{\sigma}_{\theta,l}(a) - \underline{\sigma}_{\theta,l}(a) \right) \cdot 0.5 \cdot \left(\sum_{m \in A^l} |(\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m))| \right) =$$

$$0.5 \cdot \sum_{a \in A} \left(\overline{\sigma}_{\theta,l}(a) - \underline{\sigma}_{\theta,l}(a) \right) \cdot \| \psi_{l,\gamma} - \psi_{l,\gamma'} \|_1 .$$

Equality (19) is implied by the fact that $\psi_{l,\gamma}$ and $\psi_{l,\gamma'}$ are both distributions, and the sum of the differences in the probabilities that they assign to samples of size l must be equal to zero. Thus we have shown that

$$\sum_{a \in A} \left| \sum_{m \in A^l} \left(\psi_{l,\gamma}(m) - \psi_{l,\gamma'}(m) \right) \cdot \sigma_{\theta,m}(a) \right| \leq 0.5 \cdot \sum_{a \in A} \left(\overline{\sigma}_{\theta,l}(a) - \underline{\sigma}_{\theta,l}(a) \right) \cdot \| \psi_{l,\gamma} - \psi_{l,\gamma'} \|_1 , \qquad (20)$$

which together with Lemma 3 implies that the LHS of (20) is weakly smaller than $r_{\theta,l} \cdot \| \psi_{l,\gamma} - \psi_{l,\gamma'} \|_1$. $\qquad \square$

References

Acemoglu, Daron, Dahleh, Munther A., Lobel, Ilan, & Ozdaglar, Asuman. 2011. Bayesian learning in social networks. *The Review of Economic Studies*, **78**(4), 1201–1236.

Arthur, W. Brian. 1989. Competing technologies, increasing returns, and lock-in by historical events. *The economic journal*, **99**(394), 116–131.

Arthur, W. Brian. 1994. *Increasing returns and path dependence in the economy*. University of michigan Press.

Banerjee, Abhijit, & Fudenberg, Drew. 2004. Word-of-mouth learning. *Games and Economic Behavior*, **46**(1), 1–22.

Cason, Timothy N., Friedman, Daniel, & Hopkins, Ed. 2014. Cycles and instability in a rock–paper–scissors population game: A continuous time experiment. *The Review of Economic Studies*, **81**(1), 112–136.

Ellison, Glenn, & Fudenberg, Drew. 1993. Rules of thumb for social learning. *Journal of Political Economy*, **101**, 612–643.

Ellison, Glenn, & Fudenberg, Drew. 1995. Word-of-mouth communication and social learning. *The Quarterly Journal of Economics*, **110**(1), 93–125.

Foster, Dean, & Young, Peyton. 1990. Stochastic evolutionary game dynamics. *Theoretical Population Biology*, **38**(2), 219–232.

Heller, Yuval, & Mohlin, Erik. 2017. *Observations on Cooperation*. Mimeo.

Kandori, Michihiro, Mailath, George J., & Rob, Rafael. 1993. Learning, mutation, and long run equilibria in games. *Econometrica*, **61**(1), 29–56.

Kaniovski, Yuri M., & Young, H. Peyton. 1995. Learning dynamics in games with stochastic perturbations. *Games and Economic Behavior*, **11**(2), 330–363.

Munkres, James R. 2000. *Topology*. 2nd edn. Prentice Hall.

Munshi, Kaivan. 2004. Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of development Economics*, **73**(1), 185–213.

Nowak, Martin A., & Sigmund, Karl. 1998. Evolution of indirect reciprocity by image scoring. *Nature*, **393**(6685), 573–577.

Okuno-Fujiwara, Masahiro, & Postlewaite, Andrew. 1995. Social norms and random matching games. *Games and Economic Behavior*, **9**(1), 79–109.

Oyama, Daisuke, Sandholm, William H., & Tercieux, Olivier. 2015. Sampling best response dynamics and deterministic equilibrium selection. *Theoretical Economics*, **10**(1), 243–281.

Pata, Vittorino. 2014. *Fixed point theorems and applications*. Mimeo.

Phelan, Christopher, & Skrzypacz, Andrzej. 2006. *Private monitoring with infinite histories*. Tech. rept. Federal Reserve Bank of Minneapolis.

Rosenthal, Robert W. 1979. Sequences of games with varying opponents. *Econometrica*, **47**(6), 1353–1366.

Sandholm, William H. 2001. Almost global convergence to p-dominant equilibrium. *International Journal of Game Theory*, **30**(1), 107–116.

Sandholm, William H. 2011. *Population Games and Evolutionary Dynamics*. Cambridge, MA: MIT Press.

Smith, Lones, & Sorensen, Peter Norman. 2014. Rational social learning by random sampling.

Takahashi, Satoru. 2010. Community enforcement when players observe partners' past play. *Journal of Economic Theory*, **145**(1), 42–62.

Young, H. Peyton. 1993a. The evolution of conventions. *Econometrica*, **61**(1), 57–84.

Young, H. Peyton. 1993b. An evolutionary model of bargaining. *Journal of Economic Theory*, **59**(1), 145–168.

Young, H. Peyton. 2015. The evolution of social norms. *Annual Reviews of Economics*, **7**(1), 359–387.