Observational Learning in Large Anonymous Games*

Ignacio Monzón†

September 28, 2017

Abstract

I present a model of observational learning with payoff interdependence. Agents, ordered in a sequence, receive private signals about an uncertain state of the world and sample previous actions. Unlike in standard models of observational learning, an agent's payoff depends both on the state and on the actions of others. Agents want both to learn the state and to anticipate others' play. As the sample of previous actions provides information on both dimensions, standard informational externalities are confounded with coordination motives. I show that in spite of these confounding factors, when signals are of unbounded strength there is learning in a strong sense: agents' actions are *ex-post* optimal given both the state of the world and others' actions. With bounded signals, actions approach ex-post optimality as the signal structure becomes more informative.

JEL Classification: C72, D83, D85

Keywords: observational learning, payoff interdependence, information aggregation, position uncertainty

*I am grateful to Bill Sandholm for his advice, suggestions and encouragement. I also thank Nageeb Ali, Andrea Gallice, Leandro Gorno, Daniel Hauser, Toomas Hinnosaar, Giorgio Martini, Alexei Parakhonyak, Lones Smith, and Aleksey Tetenov for valuable comments and suggestions. Some early ideas that led to this paper were present in "Social Learning in Games", the third chapter of my dissertation, which was joint work with Michael Rapp. I especially thank him. Online Appendix available at http://bit.ly/stratlearn.

†Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri (TO), Italy. (ignacio@carloalberto.org, http://www.carloalberto.org/people/monzon/)

1. Introduction

In several economic environments, the utility of an agent is affected both by some uncertain state of the world and by the actions of others. Consider a brand new operating system, of unknown quality. Each consumer cares not only about its quality, but also about whether others will adopt it. A consumer who considers buying the new system does not know how many after him will also adopt it, and may not know exactly how many before him have already adopted it. Alternatively, consider a farmer who must decide to plant either corn or soybeans at the start of the season but is uncertain about their relative demand by the end of the season. Even though he does not know what other farmers will choose, the choices of others affect the relative profitability of each crop: if most farmers plant corn, then the price of corn will be lower, and so it is more profitable to plant soybeans. Similar stories apply to investment in assets with unknown fundamentals, voting, contributions to public goods of uncertain quality, network congestion, and many other environments.

By observing a sample of the actions of others, an agent obtains information both about the state of the world and about how others behave. The farmer deciding between crops may have private information on how the demand will be at the end of the season. He may also observe the decisions of some of his neighbors. With these two sources of information he must form beliefs about both the future demand and about the actions of those farmers he does not observe. With payoff externalities, standard informational externalities are confounded with coordination motives. When a farmer observes that one of his neighbors plants corn, this might mean that his neighbor believes corn to be in high demand, as in standard models of observational learning. It may also mean that most farmers are planting corn.

I study the outcomes of observational learning in large games. In the standard setup of observational learning, (complete) learning has a simple definition: the fraction of adopters of the superior action must approach one. When payoffs depend on others' actions, the *right* action depends not only on the state of the world, but also on what others do. I focus then on whether realized actions are ex-post optimal. I say that *strategic* *learning* occurs when agents' actions are *ex-post* optimal given both the state of the world and the realized actions of others. The main message of this paper is simple:

Proposition 2. Strategic learning occurs, provided that the signal structure is sufficiently informative.

The notion of strategic learning is demanding: it requires that agents not only learn about the state of the world, but also that they correctly anticipate others' actions and best respond to them. In what follows I describe the framework and I present the intuition behind this result.

Agents are exogenously ordered in a sequence and are uncertain about their position in it. There are two, a priori equally likely, states of the world. Each agent receives a private signal about the underlying state of the world and observes the actions of some of his predecessors. Then, he makes a once-and-for-all decision between two actions (zero and one). The main innovation with respect to the standard setup is that an agent's payoff depends not only on his own action and the unknown state of the world, but also on the *proportion* X of agents who choose action one My framework applies to the examples described before (*coordination games*, like the adoption of a new operating system, and *anti–coordination games*, like the example of farmers). I do not impose any particular functional form on how payoffs depend on X.

Payoff interdependence adds a strategic consideration to observational learning: each agent understands that since his own action is observed by some of his successors, it partly determines *their* decisions. An agent who can affect aggregate outcomes needs to take into account the effect of his decision on others' actions. Gallice and Monzón [2016] show that this strategic component can have a strong effect on the aggregate play when there is a finite number of agents who never make mistakes.[1] However, this should intuitively be less relevant in large games. Individual farmers do not expect to be able to affect aggregate supply, and individual consumers typically do not believe that they can determine the overall adoption rate of a new operating system. In this paper I assume that agents make mistakes with arbitrarily small probability. I show that this implies that

[1] In Gallice and Monzón [2016], a finite number of agents must decide sequentially whether to contribute to a public good. Full contribution can occur in equilibrium because agents are uncertain about their positions and make no mistakes. Each individual agent can determine the realized aggregate outcome.

agents cannot individually determine the aggregate play.

The intuition behind strategic learning is simple and has two components. First, although each agent could in principle affect aggregate outcomes, in practice there are no *butterfly effects*. As the number of agents grows large, each individual's action has a smaller effect on the proportion X. Second, as each agent foresees that each action has a small effect on X, he can treat the proportion X as given. Realized payoffs depend (approximately) only on the state of the world and his own action. In this sense, I translate a game of observational learning *with* payoff externalities into a game of observational learning *without them*. Then, I use tools of standard observational learning to show that strategic learning occurs. I develop this intuition in detail in what follows.

The first main result (Proposition 1) shows that as the number of agents grows large, the proportion X converges to its expectation in each state of the world. This proposition addresses two challenges that result from the additional strategic factors associated with payoff externalities. First, each agent needs to anticipate how others will behave. Second, each agent may need to account for the effect of his own action on others' decisions.

I develop a novel approach to show convergence of the proportion X. If the equilibrium strategy profile were the same regardless of the number of agents, Proposition 1 would be straightforward. Agents make mistakes with positive probability, so a fixed strategy profile would create an irreducible and aperiodic Markov Chain over actions. Thus, a standard ergodic argument would lead to this result. However, as the number of agents grows, the game changes, so the equilibrium strategy profile varies with the number of agents. I use a coupling argument to show that *any* Markov Chain induced by a strategy profile converges to its stationary distribution. The speed of convergence has a geometric lower bound which is *independent* of the particular equilibrium strategy profile. Thus, the effect of one individual's action on the proportion X wanes as the number of agents grows, even with strategy profiles that change with the number of agents. I show through this argument that the proportion X converges to its expectation. As a direct consequence, no individual agent can affect the aggregate outcome. This result holds true for all payoff specifications.

The second main result (Proposition 2) explains why strategic learning must occur in equilibrium when signals are of unbounded strength. Since the proportion X converges to its expectation in each state of the world, each agent can anticipate the *payoffs* he would get from each action in each state of the world. Optimality considerations limit the possible combinations of proportions X and payoffs that can occur in equilibrium. To see this, consider first a long-run outcome where *in both states of the world*, the payoff from choosing action one exceeds that from choosing action zero. Any agent who chooses action zero regrets it ex-post. Intuitively, an agent could instead choose action one *always*, and obtain higher payoffs. It follows that no positive proportion of agents can choose a dominated action.

The final step in Proposition 2 deals with long-run outcomes where agents want to choose different actions in different states of the world. I provide an improvement principle that applies to environments with payoff externalities. An individual can always copy a random action from the sample he observes. Moreover, when his private signal is strong enough, he can go *against* the observed action, and do (in expected terms) strictly better than the observed agent. Then, as the number of agents grows large, it must be the case that either 1) the fraction of agents who choose the superior action approaches one, or that 2) the extra payoff from choosing the right action approaches zero. In either case, there is strategic learning.

Proposition 2 provides a unique prediction of play for games with only one Nash equilibrium (e.g. an anti–coordination game). In the farmers' example, the proportion of crops planted correctly matches the demand. If instead there are several equilibria in each state of the world, Proposition 2 does not select among them. I illustrate this point through a coordination game (Example 7).

Finally, I show that some degree of information aggregation also occurs with signals of bounded strength. Lemma 7 presents a notion of *bounded* strategic learning. Although actions may be ex-post suboptimal with bounded signals, there is a bound on how far actions can be from optimality. This bound depends on the information structure, and approaches zero as signals' informativeness increases.

1.1 Related Literature

There is a large literature that studies observational learning, starting from the seminal contributions of Bikhchandani, Hirshleifer, and Welch [1992] and Banerjee [1992]. In these papers, a set of rational agents choose sequentially between two actions. An agent's payoff depends on whether his action matches the unknown state of the world, but not on others' actions. The actions of others are relevant only because of their informational content. In Bikhchandani et al. [1992] and Banerjee [1992], each agent knows that his own signal is not better than the signals others have received. Agents eventually follow others' behavior and disregard their own signals. Then, the optimal behavior of rational agents can prevent complete learning. Smith and Sørensen [2000] show that when signals are of unbounded strength, individuals never fully disregard their own information and complete learning occurs. Monzón and Rapp [2014] present conditions for information aggregation when agents are uncertain both about their own position in the sequence and about the positions of those they observe.

Starting with Dekel and Piccione [2000], a line of research focuses on the outcomes of sequential voting. In Dekel and Piccione [2000], a finite sequence of agents cast votes between two alternatives. Their focus is on the comparison between simultaneous and sequential voting. Dekel and Piccione show that any equilibrium of a *simultaneous* voting game is also an equilibrium when voting is sequential. In Callander [2007], agents vote sequentially and care not only about electing the superior candidate, but also about voting for the winning candidate. Callander shows that a *bandwagon* eventually starts: voters ignore their private information and vote for the leading candidate. Ali and Kartik [2012] present a model motivated by sequential voting, but which encompasses the class of *collective preferences*: an agent's utility increases when others choose an action that matches the unknown state. Ali and Kartik show how herds can arise. My paper differs from this line of research in several dimensions. First, I allow for payoff externalities that can be both positive or negative. My model can accommodate both incentives to conform, and incentives to go against the crowd. Second, agents observe a sample of past behavior instead of the whole history of play. Together with position uncertainty and a positive probability of mistakes, this implies that agents cannot individually determine the aggregate outcome. Third, my focus is not on herds, but rather on whether agents are ex-post satisfied with their action.

Several recent papers have highlighted the importance of payoff externalities in other environments. Eyster, Galeotti, Kartik, and Rabin [2014] present a model of observational learning with congestion. As usual, agents want to match their action to the state of the world. But when previous agents in the sequence choose an action, they make it less attractive for those coming *later*. Eyster et al. study whether learning occurs as a function of congestion costs. Cripps and Thomas [2016] present a model of (possibly informative) queues. Service to those in the queue is provided only in the good state of the world, but at an stochastic rate. Cripps and Thomas study the dynamics of the queue. Arieli [2017] focuses on recurring games: successive generations of agents play the same game. As in my paper, payoffs depend on the unknown state of the world, and also on the actions of others. However, payoff externalities are only local: an agent's utility is affected by the actions of others in the same generation. Arieli studies when complete learning occurs. Besides the points already mentioned, my paper differs from Eyster et al. [2014], Cripps and Thomas [2016] and Arieli [2017] in that an agent's payoff depends on the actions of those before and *also after* him in the sequence. This adds a strategic consideration to the analysis, as agents may affect future decisions.

2. Model

Let $\mathcal{I} = \{1, \ldots, T\}$ be a set of agents, indexed by i. Agents are exogenously placed in a sequence in positions indexed by $t \in \{1 \ldots, T\}$. The random variable Q assigns a position $Q(i)$ to each agent i. Let $q : \{1, \ldots, T\} \to \{1, \ldots, T\}$ be a permutation and $\mathcal{Q}$ be the set of all possible permutations. All permutations are ex-ante equally likely: $\Pr(Q = q) = \frac{1}{T!}$ for all $q \in \mathcal{Q}$. Each individual has *no* ex-ante information about his position in the sequence.[2]

There are two equally likely states of the world $\theta \in \Theta = \{0, 1\}$. Agents must choose

[2] This setup corresponds to the case of symmetric position beliefs as defined in Monzón and Rapp [2014].

between two possible actions $a \in \mathcal{A} = \{0, 1\}$. The timing of the game is as follows. First, nature chooses the state of the world θ and the order of the sequence q. Agents do not observe these directly. Instead, each agent i receives a noisy signal about the state of the world and a sample of past actions. Then he makes a once-and-for-all choice.

payoffs may depend on the actions of others. Let $X \equiv \frac{1}{T}\sum_{j\in\mathcal{I}} a_j$ denote the proportion of agents who choose action 1, with realizations $x \in [0, 1]$. Agent i obtains utility $u(a_i, X, \theta) : \mathcal{A} \times [0, 1] \times \Theta \to \mathbb{R}$, where $u(a_i, X, \theta)$ is a continuous function in X.[3]

2.1 Private Signals

Each agent i receives a private signal $S_{Q(i)}$, with realizations $s \in \mathcal{S}$. Conditional on the true state of the world, signals are i.i.d. across individuals and distributed according to ν_0 if $\theta = 0$ or ν_1 if $\theta = 1$. I assume that ν_0 and ν_1 are mutually absolutely continuous. Then, no perfectly-revealing signals occur with positive probability, and the following likelihood ratio (Radon-Nikodym derivative) exists $l(s) \equiv \frac{d\nu_1}{d\nu_0}(s)$. Let G_θ be the distribution function for this likelihood ratio: $G_\theta(l) \equiv \Pr(l(S) \leq l \mid \theta)$. Since ν_0 and ν_1 are mutually absolutely continuous, the support $\text{supp}(G)$ of G_0 coincides with the support of G_1. I define signal strength as follows.

Definition. Signal Strength. *Signal strength is* ***unbounded*** *if* $0 < G_0(l) < 1$ *for all likelihood ratios* $l \in (0, \infty)$. *Signal strength is* ***bounded*** *if the convex hull of* $\text{supp}(G)$ *is given by* $\text{co}(\text{supp}(G)) = [\underline{l}, \bar{l}]$, *with both* $0 < \underline{l} < 1 < \bar{l} < \infty$.[4]

2.2 The Sample of Past Actions

Agents observe others' actions through a simple sampling rule. Let $h_t = (a_1, a_2, \ldots, a_{t-1})$ denote a possible history of actions up to period $t - 1$. Let H_t be the (random) history at time t, with realizations $h_t \in \mathcal{H}_t$. Agent i in position $q(i) = t$ receives a sample $\xi_t : \mathcal{H}_t \to$

[3]Note that an agent's payoff depends on the actions of both those who came before him and those who come after him in the sequence.

[4]I disregard intermediate cases, since they do not add much to the understanding of observational strategic learning.

Ξ containing the ordered choices of his M predecessors (if available):

$$\xi_t = \begin{cases} \varnothing & \text{if } t = 1 \\ (a_1, \ldots, a_{t-1}) & \text{if } 1 < t \leq M \\ (a_{t-M}, \ldots, a_{t-1}) & \text{if } t > M \end{cases}$$

The first agent observes nobody's action, so he receives an empty sample. Agents in positions $t \in \{2, \ldots, M\}$ observe the actions of all their predecessors. Subsequent agents observe the actions of their M immediate predecessors.

2.3 Strategies, Mistakes and Equilibrium Existence

All information available to an agent is summarized by $\{s, \xi\}$, which is an element of $\mathcal{S} \times \Xi$. I assume that individuals make mistakes with small probability $\varepsilon > 0$, so their strategies are ε–constrained. Formally, agent i's strategy is a function $\sigma_i : \mathcal{S} \times \Xi \rightarrow [\varepsilon, 1 - \varepsilon]$ that specifies a probability $\sigma_i(s, \xi)$ for choosing action 1 given the information available. Σ denotes the set of ε–constrained strategies. Let σ_{-i} be the strategies for all players other than i. Then the profile of play is given by $\sigma = (\sigma_i, \sigma_{-i})$.[5]

Every profile σ induces a probability distribution $\mathbf{P}_\sigma$ over histories H_t, and consequently over proportions X. Profile $\sigma^* = \left(\sigma_i^*, \sigma_{-i}^*\right)$ is a Bayes-Nash equilibrium of the game if

$$E_{\sigma^*}\left[u\left(a_i, X, \theta\right)\right] \geq E_{\left(\sigma_i, \sigma_{-i}^*\right)}\left[u\left(a_i, X, \theta\right)\right] \quad \text{for all } \sigma_i \in \Sigma \text{ and for all } i.$$

A profile of play is symmetric if $\sigma_i = \sigma_j$ for all $i, j \in \mathcal{I}$.

Lemma 1. *For each T there exists a symmetric equilibrium $\sigma^{*,T}$.*

See Appendix A.1 for the proof.

[5]Mistakes are rationally anticipated. This model is equivalent to one in which agents choose from $[0, 1]$, but they know in advance that there is a 2ε chance that their decision will be overruled by a coin flip. An alternative interpretation of this model is as follows. With probability $1 - 2\varepsilon$, an agent chooses rationally from $[0, 1]$. With probability 2ε, the agent is a "behavioral" type. Half of behavioral types always choose action 0, while the others always choose action 1.

2.4 Definition of Strategic Learning

I study the outcomes of large anonymous games, so I let the number of agents grow large and study symmetric equilibria. Agents face a different stage game in each state of the world. Ex-ante, each agent is uncertain not only about the state of the world θ, but also about the realization of the proportion X. An agent receives his private signal and observes the actions of some predecessors. Given this information, he forms beliefs *both* about the underlying state of the world, and about the possible realizations of the proportion X. Then he chooses an action.

I study whether agents can successfully learn both about the state of the world and about the proportion X. In standard observational learning models, complete learning occurs when the fraction of adopters of the superior action approaches one. When payoff externalities exist, I say *strategic learning* occurs whenever agents' actions are *ex-post* optimal given both the state of the world and the realization of the proportion X. I first present two simple examples that illustrate when agents will be ex-post satisfied with their actions. I then introduce the formal definition of strategic learning.

EXAMPLE 1. ANTI–COORDINATION. *Let* $u(1, X, 0) = \frac{1}{5} - X$, $u(1, X, 1) = \frac{4}{5} - X$, *and* $u(0, X, \theta) = 0$.

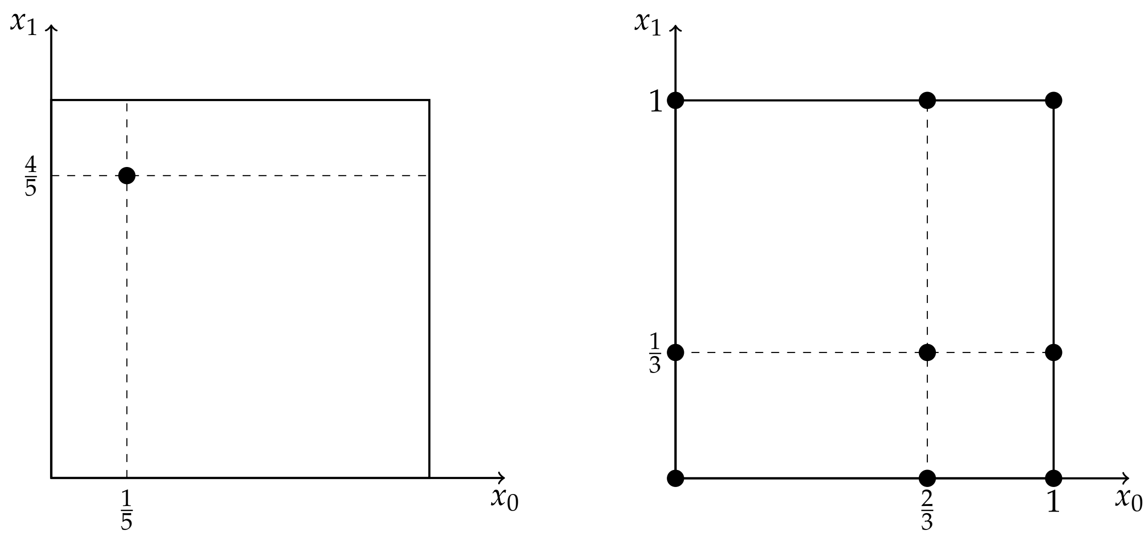
Example 1 presents an environment where choosing action 1 becomes less attractive as more agents also choose it. In state $\theta = 0$ action 1 is preferred as long as $X \leq \frac{1}{5}$, while in state $\theta = 1$, action 1 is preferred whenever $X \leq \frac{4}{5}$. Let x_θ be the realized proportion in state θ, so $x = (x_0, x_1)$ is the vector of realized proportions in each state. When $(x_0, x_1) = (\frac{1}{5}, \frac{4}{5})$ agents are ex-post satisfied with their choices. If instead for example $(x_0, x_1) = (0, \frac{4}{5})$, agents would have preferred choosing action 1 in state $\theta = 0$. In fact, $(\frac{1}{5}, \frac{4}{5})$ is the only vector of realized proportion that makes all agents ex-post satisfied with their actions in both states of the world.

Formally, define the *excess utility* from choosing action 1 in state θ given X as $v_\theta(X) \equiv u(1, X, \theta) - u(0, X, \theta)$. I say that x_θ corresponds to a Nash Equilibrium of the stage game θ (and denote it by $x_\theta \in NE^\theta$) whenever $v_\theta(x_\theta) > 0 \Rightarrow x_\theta = 1$ and $v_\theta(x_\theta) < 0 \Rightarrow x_\theta = 0$. Similarly, $x \in NE$ whenever $x_\theta \in NE^\theta$ for both $\theta \in \{0, 1\}$.

The circle in Figure 1(a) depicts the set NE for Example 1. There is a unique $x_\theta \in NE^\theta$ for each $\theta \in \{0,1\}$, so NE is the singleton $\{(\frac{1}{5}, \frac{4}{5})\}$. Other games can have multiple elements in NE. Consider for example the following simple coordination game.

EXAMPLE 2. COORDINATION. *Let* $u(1, X, 0) = X - \frac{2}{3}$, $u(1, X, 1) = X - \frac{1}{3}$, *and* $u(0, X, \theta) = 0$.

In Example 2, $NE^0 = \{0, \frac{2}{3}, 1\}$ and $NE^1 = \{0, \frac{1}{3}, 1\}$. Then, there are nine elements in NE, depicted in Figure 1(b) with circles.



(a) NE set in Example 1 (Anti–coordination) (b) NE set in Example 2 (Coordination)

Figure 1: NE sets in Examples 1 and 2

It is not obvious a priori whether the realized proportion will be close to elements of NE. The main result in this paper (Proposition 2) shows that this is in fact the case. Intuitively, there is strategic learning when, as the number of agents grows large, the (random) proportion X gets close to NE. Because mistakes occur with positive probability, the proportion X may not get arbitrarily close to elements in NE. This is why I first take the number of agents to infinity and then the probability of mistakes to zero. Let the distance between the realized proportion x and the set NE be defined by $d(x, NE) \equiv \min_{y \in NE} |x - y|$.

DEFINITION. STRATEGIC LEARNING. *There is* strategic learning *when for all* $\delta > 0$ *there* *exists $\tilde{\varepsilon} > 0$, such that*

$$\lim_{T\to\infty} \mathbf{P}_{\sigma^{*,T}} \left(d\left(X, NE\right) < \delta \right) = 1$$

for all sequences of symmetric equilibria $\left\{\sigma^{*,T}\right\}_{T=1}^{\infty}$ in games with probability of mistakes $\varepsilon < \tilde{\varepsilon}$.

3. Results

3.1 Average Action Convergence

The (random) proportion X converges to its expectation in both states of the world. Let the random variable $X_\theta|\sigma$ represent the proportion of agents who choose action one, conditional on the state of the world θ, and given the strategy σ. The vector $X|\sigma = (X_0|\sigma, X_1|\sigma)$ has realizations $x = (x_0, x_1)$ and expectation $E\left[X|\sigma\right] = \left(E_\sigma\left[X_0\right], E_\sigma\left[X_1\right]\right)$. A sequence of symmetric strategy profiles $\left\{\sigma^T\right\}_{T=1}^{\infty}$ induces a sequence of proportions $\left\{X|\sigma^T\right\}_{T=1}^{\infty}$ and a sequence of expected proportions $\left\{E\left[X|\sigma^T\right]\right\}_{T=1}^{\infty}$. As highlighted by the notation, the expected proportion may change with T, and in fact need not converge. I show that in spite of this, $X|\sigma^T$ converges in probability to its expectation.

Proposition 1. Average Action Converges in Probability. *Take any sequence of symmetric strategy profiles $\left\{\sigma^T\right\}_{T=1}^{\infty}$. Then, $X_\theta|\sigma^T - E\left[X_\theta|\sigma^T\right] \xrightarrow{p} 0$. More generally, take any sequence $\left\{\widetilde{\sigma}_i^T\right\}_{T=1}^{\infty}$ of alternative strategies for agent i. Let the profile of play $\widetilde{\sigma}^T = (\widetilde{\sigma}_i^T, \sigma_{-i}^T)$ include i's alternative strategy. Then, $X_\theta|\widetilde{\sigma}^T - E\left[X_\theta|\sigma^T\right] \xrightarrow{p} 0$.*

See Appendix A.2 for the proof.

A symmetric strategy profile σ^T induces a Markov Chain over M-period histories of play. Any agent in positions $t > M$ observes the actions of his M immediate predecessors. As σ^T is symmetric, the likelihood that agent i in position $Q(i) > M$ chooses action 1 given sample ξ is independent of both his identity and his position. Then, σ^T induces a Markov Chain $\{Y_t\}_{M<t\leq T}$ over M-period histories. Moreover, as agents make mistakes with probability $\varepsilon > 0$, the Markov Chain is irreducible and aperiodic. If σ^T was fixed for all T, a standard argument (Ergodic Theorem) would suffice to show Proposition 1. However, there is no guarantee that the equilibrium play is independent of the number of agents. In fact, it is easy to find examples where this is not the case.

A Markov Chain induced by an ε–constrained symmetric strategy profile σ^T converges to its unique stationary distribution geometrically. Fix a strategy profile σ^T and its induced Markov Chain $\{Y_t\}_{t>M}$, but let $t \to \infty$. A coupling argument provides a geometric lower bound on the speed of convergence to the stationary distribution. What is more, for any $\varepsilon > 0$, this lower bound is independent of the particular strategy profile σ^T. As a result, although $\{Y_t\}_{M<t\leq T}$ depends on a particular σ^T, it must approach its expectation as T grows. In fact, $X_\theta|\sigma^T - E\left[X_\theta|\sigma^T\right]$ converges in L^2 norm, so it also converges in probability.

Finally, the long-run behavior of the proportion X does not change when one agent deviates and picks a different strategy. To see this, compare the random proportion $X|\sigma^T$ induced by the symmetric profile to the one $X|\widetilde{\sigma}^T$ induced when one agent deviates. Let i be the agent who deviates and chooses strategy $\widetilde{\sigma}_i$. Agents in positions earlier than $Q(i)$ are not affected by agent i's strategy. Agents in positions right after $Q(i)$ are directly affected. Because of mistakes, the effect that agent i's action has on subsequent actions $t > Q(i)$ vanishes (geometrically) as t increases. So as the total number of agents T increases, the fraction of agents who are directly affected by i's action goes to zero. Then also $X_\theta|\sigma^T - X_\theta|\widetilde{\sigma}^T \xrightarrow{p} 0$.

As I show next, Proposition 1 allows for a simple approximation to the utility agents obtain from playing this game.

3.2 Utility Convergence

Agents' expected utility converges to the *utility of the expected average action*. Agents' expected utility under symmetric profile σ^T is simply

$$u\left(\sigma^T\right) \equiv E_{\sigma^T}\left[u\left(a_i, X, \theta\right)\right] = \frac{1}{2}\sum_{\theta\in\{0,1\}} E_{\sigma^T}\left[X_\theta \cdot u\left(1, X_\theta, \theta\right) + \left(1 - X_\theta\right)\cdot u\left(0, X_\theta, \theta\right)\right].$$

Define the *utility of the expected average action* $\bar{u}^T$ by

$$\bar{u}^T \equiv \frac{1}{2}\sum_{\theta\in\{0,1\}} E_{\sigma^T}\left[X_\theta\right]\cdot u\left(1, E_{\sigma^T}\left[X_\theta\right], \theta\right) + \left(1 - E_{\sigma^T}\left[X_\theta\right]\right)\cdot u\left(0, E_{\sigma^T}\left[X_\theta\right], \theta\right).$$

LEMMA 2. EXPECTED UTILITY CONVERGENCE. *Take any sequence of symmetric strategy profiles* $\{\sigma^T\}_{T=1}^{\infty}$. *Then,* $\lim_{T\to\infty}\left[u\left(\sigma^T\right)-\bar{u}^T\right]=0$.

Proof. By Proposition 1, $X_\theta|\sigma^T - E\left[X_\theta|\sigma^T\right] \xrightarrow{p} 0$. The function $u\left(a_i, X, \theta\right)$ is continuous in X. Then, $X_\theta|\sigma^T \cdot u\left(a_i, X_\theta|\sigma^T, \theta\right) \xrightarrow{p} E_{\sigma^T}\left[X_\theta\right]\cdot u\left(a_i, E_{\sigma^T}\left[X_\theta\right], \theta\right)$ because of the continuous mapping theorem. Moreover, $u\left(a_i, X, \theta\right)$ is bounded, so $X_\theta|\sigma^T \cdot u\left(a_i, X_\theta|\sigma^T, \theta\right)$ is also bounded. Then $\lim_{T\to\infty} E_{\sigma^T}\left[X_\theta \cdot u\left(a, X_\theta, \theta\right)\right] = \lim_{T\to\infty} E_{\sigma^T}\left[X_\theta\right]\cdot u\left(a_i, E_{\sigma^T}\left[X_\theta\right], \theta\right)$ by Portmanteau's Theorem. This leads directly to $\lim_{T\to\infty}\left[u\left(\sigma^T\right)-\bar{u}^T\right]=0$. ■

Proposition 1 also allows for a simple approximation of the expected utility of deviations. Suppose that agent i chooses an alternative strategy $\widetilde{\sigma}_i$ and let $u\left(\widetilde{\sigma}_i^T, \sigma_{-i}^T\right)$ denote the resulting expected utility from this deviation. Define the *approximate utility of the deviation* $\widetilde{u}^T$ as

$$\widetilde{u}^T \equiv \frac{1}{2}\sum_{\theta\in\{0,1\}}\sum_{a\in\mathcal{A}} \mathbf{P}_{\widetilde{\sigma}^T}\left(a_i = a \mid \theta\right)\cdot u\left(a, E_{\sigma^T}\left[X_\theta\right], \theta\right).$$

LEMMA 3. EXPECTED UTILITY OF DEVIATIONS. *Take any sequence of symmetric strategy profiles* $\{\sigma^T\}_{T=1}^{\infty}$ *and a sequence of alternative strategies for agent i:* $\{\widetilde{\sigma}_i^T\}_{T=1}^{\infty}$. *Then,* $\lim_{T\to\infty}\left[u\left(\widetilde{\sigma}_i^T, \sigma_{-i}^T\right)-\widetilde{u}^T\right]=0$.

The proof closely follows that of Lemma 2. See Appendix A.4 for the details.

3.3 The Set of Limit Points

Different profiles of play σ^T induce different distributions over X. Then, the sequence of expected proportions $\{E\left[X|\sigma^T\right]\}_{T=1}^{\infty}$ need not have a limit. Although the proportion X approaches its expectation, this expectation itself may not converge. Then, I focus on the set L of *limit points* for sequences of equilibrium strategies $\{E\left[X|\sigma^T\right]\}_{T=1}^{\infty}$.

DEFINITION. LIMIT POINTS. $x = (x_0, x_1)$ *is a limit point if there exists a sequence of symmetric equilibrium strategy profiles* $\{\sigma^T\}_{T=1}^{\infty}$ *such that for some subsequence* $\{\sigma^{T_\tau}\}_{\tau=1}^{\infty}$, $\lim_{\tau\to\infty} E\left[X|\sigma^{T_\tau}\right] = x$.

The following corollary, which is an immediate consequence of Proposition 1, shows why one should focus on the set L of limit points. As the number of agents grows large, *only* proportions X close to L occur with positive probability

Corollary 1. *Take any sequence of symmetric strategy profiles* $\left\{\sigma^T\right\}_{T=1}^{\infty}$ *and any* $\delta > 0$. *Then* $\lim_{T\to\infty} \mathbf{P}_{\sigma^T}\left(d\left(X, L\right) < \delta\right) = 1$.

See Appendix A.3 for the proof.

The set of limit points L is generated by equilibrium strategies. Optimality considerations allow for a partial characterization of L. Pick a sequence of symmetric equilibria $\left\{\sigma^T\right\}_{T=1}^{\infty}$ and also a sequence of (alternative) ε–constrained strategies for agent i: $\left\{\widetilde{\sigma}_i^T\right\}_{T=1}^{\infty}$. Since σ^T are equilibrium strategies, $u\left(\widetilde{\sigma}_i^T, \sigma_{-i}^T\right) - u\left(\sigma^T\right) \leq 0$ for all σ_{-i}^T and for all T. Computing exactly $u\left(\widetilde{\sigma}_i^T, \sigma_{-i}^T\right)$ and $u\left(\sigma^T\right)$ is not possible in general. It requires specifying payoffs, the signal structure, the number M of agents sampled, and then also computing the equilibrium play. Fortunately, Lemmas 2 and 3 together make it easy to work with alternative strategies. Let the *approximate improvement* Δ^T be given by

$$\Delta^T \equiv \widetilde{u}^T - \bar{u}^T = \frac{1}{2} \sum_{\theta \in \{0,1\}} \left[\mathbf{P}_{\widetilde{\sigma}^T}\left(a_i = 1 \mid \theta\right) - E_{\sigma^T}\left[X_\theta\right]\right] \cdot v_\theta\left(E_{\sigma^T}\left[X_\theta\right]\right).$$

The following corollary provides the foundation to take advantage of the approximate improvement Δ^T.

Corollary 2. *Take any sequence of symmetric* ***equilibrium*** *strategy profiles* $\left\{\sigma^T\right\}_{T=1}^{\infty}$ *and a sequence of* ε*–constrained strategies* $\left\{\widetilde{\sigma}_i^T\right\}_{T=1}^{\infty}$ *for agent* i. *Then* $\limsup_{T\to\infty} \Delta^T \leq 0$.

See Appendix A.5 for the proof.

I present two simple alternative strategies that restrict the possible elements of the set L of limit points. The first one consists in always following a particular action, regardless of the information received. This strategy proves useful when one action dominates the other in the limit. The second strategy consists on copying the action of one of the observed agents, unless the signal received is extremely informative. This strategy resembles the standard improvement principle in observational learning, and is useful when no action strictly dominates the other in the limit.

3.4 Alternative Strategy 1: Always Follow a Given Action

The first alternative strategy is simple: follow a given action, regardless of the information received. Lemma 4 shows how this strategy imposes restrictions on the elements of L.

LEMMA 4. DOMINANCE. *Any limit actions $(x_0, x_1) \in L$ must satisfy:*

$$(x_0 - \varepsilon)\, v_0\,(x_0) + (x_1 - \varepsilon)\, v_1\,(x_1) \geq 0 \tag{1}$$

$$(1 - \varepsilon - x_0) v_0\,(x_0) + (1 - \varepsilon - x_1) v_1\,(x_1) \leq 0 \tag{2}$$

Moreover, let $v_0(x_0)v_1(x_1) \geq 0$. Then, $v_\theta\,(x_\theta) > 0$ implies $(x_0, x_1) = (1 - \varepsilon, 1 - \varepsilon)$ and $v_\theta\,(x_\theta) < 0$ implies $(x_0, x_1) = (\varepsilon, \varepsilon)$.

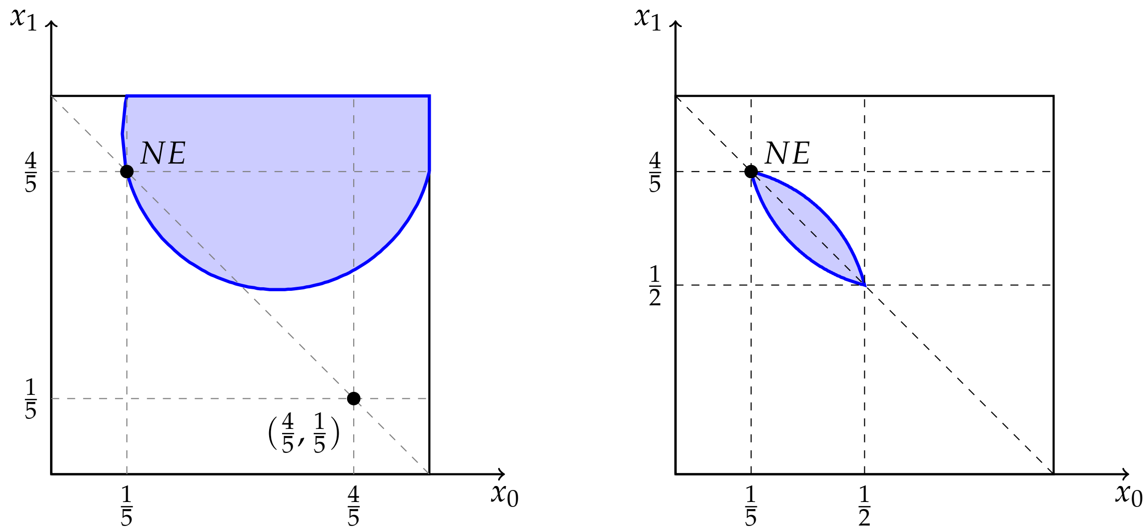
See Appendix A.6 for the proof.

To illustrate how Lemma 4 partially characterizes the long-run outcomes of large games, consider first equation (2). When equation (2) is *not* satisfied, always playing action 1 leads to a utility that is strictly higher than the expected utility of the game. Then, points that do not satisfy equation (2) cannot be limit points. Take again the simple anti–coordination game presented in Example 1. The shaded area in Figure 2(a) shows all points that satisfy equation (2).[6] Take for example $(\frac{4}{5}, \frac{1}{5})$. For a large enough number of players, agents' expected payoffs become arbitrarily close to $\frac{1}{2}[\frac{4}{5}u(1, \frac{4}{5}, 0) + \frac{1}{5}u(1, \frac{1}{5}, 1)] = -\frac{1}{2}(\frac{3}{5})^2$. An agent who *always* chooses action 1 obtains instead payoffs arbitrarily close to $\frac{1}{2}[u(1, \frac{4}{5}, 0) + u(1, \frac{1}{5}, 1)] = 0$. Then, there cannot be a sequence of equilibria that induces $(\frac{4}{5}, \frac{1}{5})$ as limit point.

Equation (1) describes the outcomes not dominated by action 0 instead. In the case of Example 1, equation (1) generates an area symmetric to that presented in Figure 2(a). In fact, it is easy to see that $(\frac{4}{5}, \frac{1}{5})$ is also dominated by always playing action 0. The shaded area in Figure 2(b) represents the possible outcomes that remain after applying Lemma 4 in Example 1.

Outcomes that make agents indifferent between actions in one state but not in the other can only be in L if all agents choose the non dominated action in *both* states, so ei-

[6] The exact shape of the sets depicted in Figures 2 and 3 depend on the value of ε. I present them with $\varepsilon = 0$.



(a) Points *not* dominated by action 1

(b) Possible outcomes after applying Lemma 4

Figure 2: Applying Lemma 4 to Example 1 (Anti–coordination).

ther $x = (1-\varepsilon, 1-\varepsilon)$ or $x = (\varepsilon, \varepsilon)$. Figure 3 illustrates this for the coordination game from Example 2. The shaded area in Figure 3(a) shows the points which are *not* dominated by action 1. The shaded area in Figure 3(b) depicts the outcomes that remain after applying Lemma 4. The non-shaded circles in Figure 3(b) like $(\frac{2}{3}, 1)$ cannot be limit points. These points are not (strictly) dominated by always playing some action. For example, as Figure 3(a) shows, $(\frac{2}{3}, 1)$ is not (strictly) worse than always choosing action 1. However, $(\frac{2}{3}, 1)$ cannot be a limit point because of the last result in Lemma 4.

3.5 Alternative Strategy 2: Improve Upon a Sampled Agent

The second alternative strategy deals with the most interesting case: *non-dominated* actions. Take a limit point $x = (x_0, x_1)$ with $v_0(x_0)v_1(x_1) < 0$. For simplicity, assume first that $v_0(x_0) < 0$ and $v_1(x_1) > 0$, so in the limit, agents want their action to match the state of the world. The question in this case is simple: do agents succeed in matching their actions to the state of the world? In other words, do *non-dominated* actions require $(x_0, x_1) = (0, 1)$? This environment resembles one from observational learning *without payoff externalities*, so the proof here follows arguments similar to those from those envi-

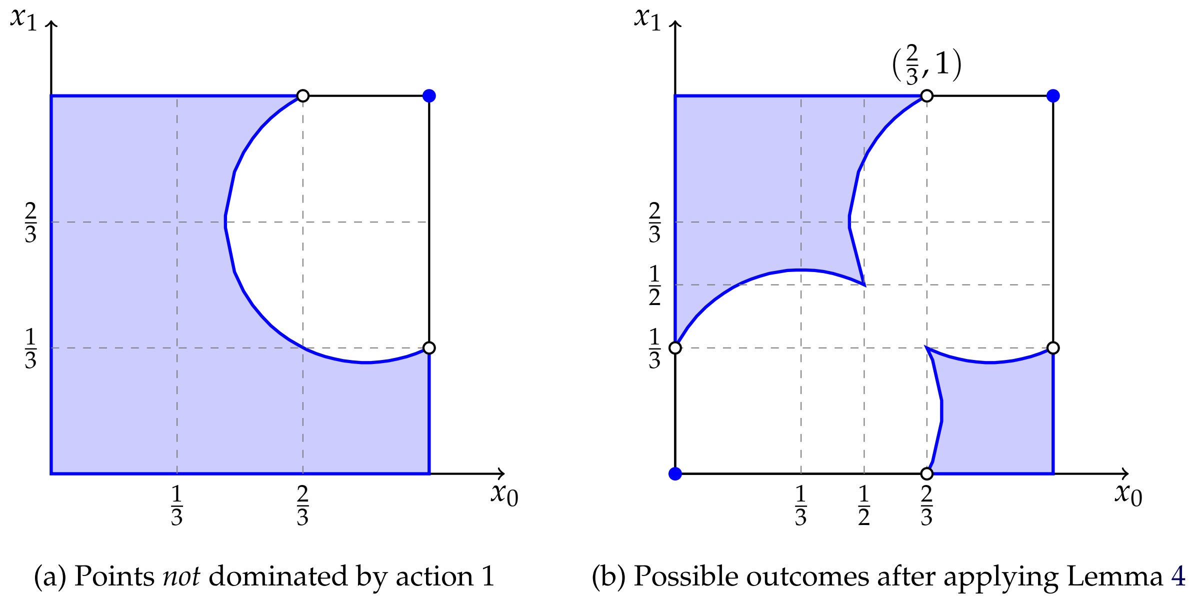


Figure 3: Applying Lemma 4 to Example 2 (Coordination).

ronments.

I introduce an improvement principle to show how observational learning restricts which outcomes can be limit points. Consider a simple strategy. Each individual selects *one* individual at random from his sample. Let $\tilde{\xi} = 1$ if the action of the selected individual is $a = 1$ and $\tilde{\xi} = 0$ otherwise. The simple strategy mandates that the sampled action must be copied, unless a strong enough signal is received. Formally, focus on T big enough so that $v_0(E_{\sigma^T}[X_0]) < 0$ and $v_1(E_{\sigma^T}[X_1]) > 0$. The simple strategy $\tilde{\sigma}^T$ is as follows:

$$\tilde{\sigma}^T\left(\tilde{\xi}, s\right) = \begin{cases} 1 & \text{if } \tilde{\xi} = 1 \text{ and } l(s) \geq \underline{k}^T \equiv \frac{-v_0(E_{\sigma^T}[X_0])}{v_1(E_{\sigma^T}[X_1])} \frac{\mathbf{P}_{\sigma^T}\left(\tilde{\xi}=1|\theta=0\right)}{\mathbf{P}_{\sigma^T}\left(\tilde{\xi}=1|\theta=1\right)} \\ 1 & \text{if } \tilde{\xi} = 0 \text{ and } l(s) \geq \bar{k}^T \equiv \frac{-v_0(E_{\sigma^T}[X_0])}{v_1(E_{\sigma^T}[X_1])} \frac{\mathbf{P}_{\sigma^T}\left(\tilde{\xi}=0|\theta=0\right)}{\mathbf{P}_{\sigma^T}\left(\tilde{\xi}=0|\theta=1\right)} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

This simple strategy improves upon the average utility $\bar{u}^T$ whenever signals are sufficiently informative and mistakes not that common. This is derived from two intuitive reasons. First, as long as signals more informative than the observed action $\tilde{\xi}$ exist, the strategy $\tilde{\sigma}$ is strictly better than just imitating $\tilde{\xi}$. Second, without mistakes, the utility of imitating the observed action $\widetilde{\xi}$ approaches the average utility $\bar{u}^T$ as the number of agents grows large.

Lemma 5. Improvement Principle. *Take any limit point $(x_0, x_1) \in L$ with $v_0(x_0) < 0$ and $v_1(x_1) > 0$. Then,*

$$
\begin{aligned}
&- v_0(x_0)\left[(1-2\varepsilon)x_0\left[G_0\left(\underline{k}\right) - \left(\underline{k}\right)^{-1} G_1\left(\underline{k}\right)\right] - \varepsilon\left(1-2x_0\right)\right] \\
&\quad + v_1(x_1)\left[(1-2\varepsilon)\left(1-x_1\right)\left[[1-G_1(\overline{k})] - \overline{k}[1-G_0(\overline{k})]\right] - \varepsilon\left(2x_1-1\right)\right] \leq 0 \qquad (4)
\end{aligned}
$$

with $\underline{k} = \frac{-v_0(x_0)}{v_1(x_1)}\frac{x_0}{x_1}$ and $\overline{k} \equiv \frac{-v_0(x_0)}{v_1(x_1)}\frac{1-x_0}{1-x_1}$.

See Appendix A.7 for the proof.

When the outcome x does not satisfy Equation (4), agents can profit from following the simple strategy $\widetilde{\sigma}$, so such an outcome cannot be a limit point. The term $G_0\left(\underline{k}\right) - \underline{k}^{-1}G_1\left(\underline{k}\right) \geq 0$ in equation (4) increases in $\underline{k}$ and is strictly positive whenever $\underline{k} > \underline{l}$. Symmetrically, the term $[1-G_1(\overline{k})] - \overline{k}[1-G_0(\overline{k})]$ is decreasing in $\overline{k}$ and strictly positive whenever $\overline{k} < \overline{l}$. Then, as long as $\underline{k} > \underline{l}$ or $\overline{k} < \overline{l}$ there is potential for improvement upon those observed. On the other side, the existence of mistakes may prevent such an improvement.[7]

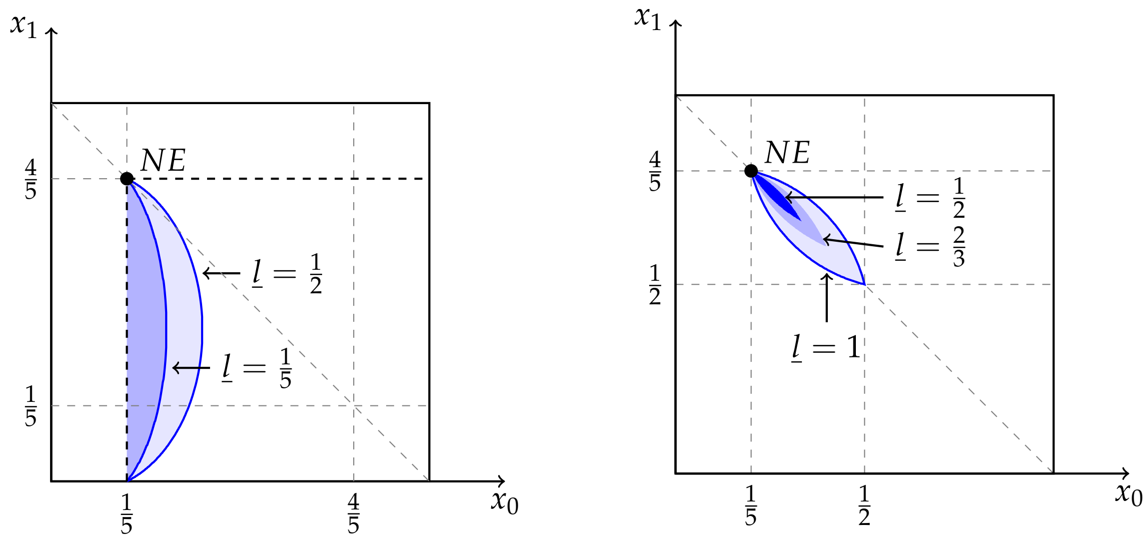
To illustrate how Lemma 5 provides a partial characterization of the outcomes of large games, consider first the anti–coordination game presented in Example 1. Lemma 5 applies when $v_0(x_0) < 0$ and $v_1(x_1) > 0$, which holds whenever $x_0 > \frac{1}{5}$ and $x_1 < \frac{4}{5}$. Take a signal structure with $\overline{l}^{-1} = \underline{l} = \frac{1}{2}$. Points *outside* of the lightly shaded area in Figure 4(a) have $\underline{k} > \underline{l}$. The term $G_0\left(\underline{k}\right) - \underline{k}^{-1}G_1\left(\underline{k}\right)$ is strictly positive there, so for ε small enough, equation (4) cannot hold. Next, take a more informative signal structure: $\overline{l}^{-1} = \underline{l} = \frac{1}{5}$. Points outside of the *dark* shaded area have $\underline{k} > \underline{l}$. As the bounds on the informativeness of the signal become less restrictive, the shaded area becomes smaller.[8]

Symmetrically, whenever $\overline{k} < \overline{l}$, then $[1-G_1(\overline{k})] - \overline{k}[1-G_0(\overline{k})]$ is strictly positive. The area determined by this condition is not depicted in Figure 4(a), but is symmetric to those

[7] Because of mistakes, it can happen that $v_0(x_0) < 0$ and $v_1(x_1) > 0$ but $(x_0, x_1) \neq (\varepsilon, 1-\varepsilon)$. Example 3 in the next section shows how this can happen in the standard observational learning setup.

[8] As before, the exact shape of the sets shown in Figures 4 and 5 depends on the value of ε. I present them with $\varepsilon = 0$.

depicted there. The shaded areas in Figure 4(b) depict outcomes (x_0, x_1) that satisfy *both* conditions: $\underline{k} \leq \underline{l}$ and $\bar{l} \leq \bar{k}$. As $\bar{l}^{-1} = \underline{l}$ gets smaller, the area satisfying $\underline{k} \leq \underline{l} < \bar{l} \leq \bar{k}$ shrinks. Figure 4(b) provides a preview of the main result of this paper. Only outcomes close to NE remain after applying Lemmas 4 and 5.



(a) No improvement from simple strategy

(b) Applying Lemmas 4 and 5 ($\bar{l}^{-1} = \underline{l}$)

Figure 4: Applying Lemma 5 to Example 1 (Anti–coordination).

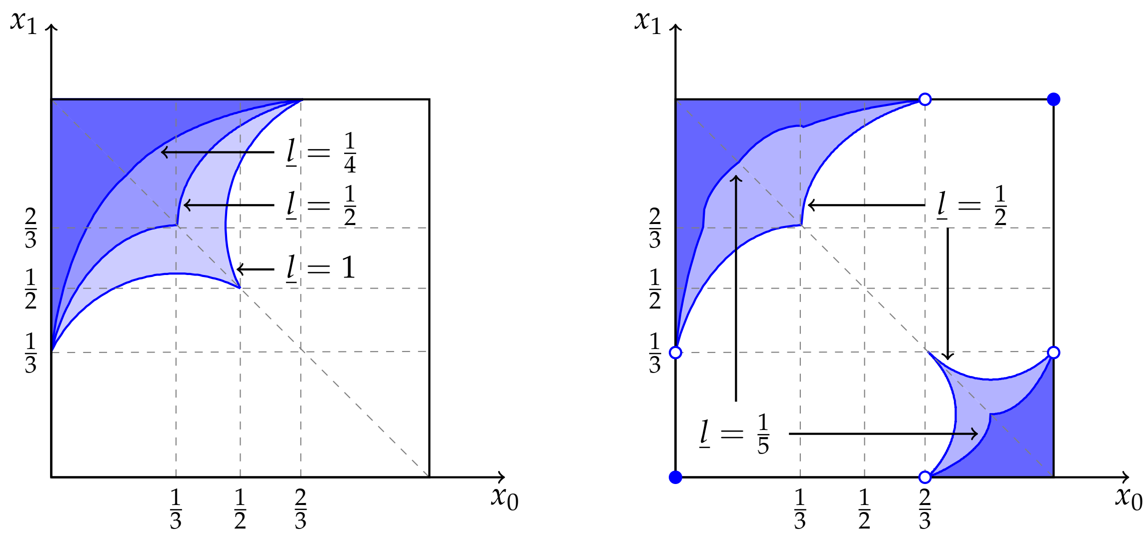
For simplicity, I have only discussed so far the case with $v_0(x_0) < 0$ and $v_1(x_1) > 0$. Lemma 6 presents an improvement principle that applies when $v_0(x_0) > 0$ and $v_1(x_1) < 0$, so agents want their action to match the opposite state of the world. The argument behind Lemma 6 is symmetric to that of Lemma 5. See the Online Appendix for details.

Lemma 6. *Take a limit point* $(x_0, x_1) \in L$ *with* $v_0(x_0) > 0$ *and* $v_1(x_1) < 0$. *Then,*

$$v_0(x_0)\left[(1-2\varepsilon)(1-x_0)\left[G_0(\bar{k}) - (\bar{k})^{-1}G_1(\bar{k})\right] - \varepsilon(2x_0 - 1)\right]$$
$$- v_1(x_1)\left[(1-2\varepsilon)x_1\left[[1 - G_1(\underline{k})] - \underline{k}[1 - G_0(\underline{k})]\right] - \varepsilon(1 - 2x_1)\right] \leq 0.$$

3.6 Strategic Learning

Lemmas 4, 5 and 6 jointly lead to the main result of this paper: there is *strategic learning*. I illustrate this result with the coordination game presented in Example 2. The shaded areas in Figure 5(a) depict the possible outcomes that satisfy equation (4) in Lemma 5 for different values of $\bar{l}^{-1} = \underline{l}$. Lemma 5 applies to outcomes with $v_0(x_0) < 0$ and $v_0(x_0) > 0$, which correspond to $x_0 < \frac{2}{3}$ and $x_1 > \frac{1}{3}$. Lemma 6 applies when $v_0(x_0) < 0$ and $v_0(x_0) > 0$, which correspond to $x_0 > \frac{2}{3}$ and $x_1 < \frac{1}{3}$. Figure 5(b) shows the set of possible limit points that remain after applying Lemmas 4, 5 and 6. The shaded areas shrink when signals become more informative. Points far from the set NE cannot be limit points. Proposition 2 formalizes this intuition.



(a) Lemma 5 ($\bar{l}^{-1} = \underline{l}$)

(b) Lemmas 4, 5 and 6 ($\bar{l}^{-1} = \underline{l}$)

Figure 5: Observational Learning in Games. Example 2 (Coordination).

Proposition 2. *Assume signals are of unbounded strength. Then there is strategic learning.*

See Appendix A.8 for the proof.

3.7 Signals of Bounded Strength

With signals of bounded strength, agents' play need not become arbitrarily close to elements of NE. I show however that there must be some degree of learning through the observations of others. I provide a bound on how far from elements of NE long-run outcomes can be. This result is a direct consequence of Lemmas 4, 5 and 6. Intuitively, whenever an agent's choice is ex-post suboptimal, it is because he was wrong about the state.

To fix ideas, let there be a positive proportion of agents who choose action one in state zero ($x_0 > 0$), but who are ex-post dissatisfied ($v_0(x_0) < 0$). Those agents would have preferred choosing action zero. The loss in the population is approximately $-v_0(x_0)x_0$. Instead, the gain in the population from choosing action one in state one is $v_1(x_1)x_1$. I show that the *ratio* between the loss and the gain must be bounded above by the informativeness of signals. This ratio is given by $\underline{k} = [-v_0(x_0)x_0]/[v_1(x_1)x_1]$. It must happen that $\underline{k} \leq \underline{l}$. Similarly, the ratio between the gain and the loss from choosing *action zero* is given by $\underline{k} = [-v_0(x_0)(1-x_0)]/[v_1(x_1)(1-x_1)]$. And it must happen that $\bar{k} \geq \bar{l}$.

In general, let the set $NE_{(\underline{l},\bar{l})}$ contain all outcomes with ratios bounded by $(\underline{l}, \bar{l})$:

$$x \in NE_{(\underline{l},\bar{l})} \text{ if } \begin{cases} v_0(x_0)v_1(x_1) \geq 0 \Rightarrow x \in NE \\ v_0(x_0) < 0 \text{ and } v_1(x_1) > 0 \Rightarrow \underline{k} \leq \underline{l} < \bar{l} \leq \bar{k} \\ v_0(x_0) > 0 \text{ and } v_1(x_1) < 0 \Rightarrow \bar{k} \leq \underline{l} < \bar{l} \leq \underline{k} \end{cases}$$

The following result shows *bounded strategic learning* must occur. Its definition is analogous to the definition of strategic learning, with NE replaced by $NE_{(\underline{l},\bar{l})}$.

LEMMA 7. BOUNDED STRATEGIC LEARNING. *For all $\delta > 0$ there exists $\tilde{\varepsilon} > 0$, such that*

$$\lim_{T\to\infty} \mathbf{P}_{\sigma^{*,T}} \left(d\left(X, NE_{(\underline{l},\bar{l})}\right) < \delta \right) = 1$$

for all sequences of symmetric equilibria $\{\sigma^{*,T}\}_{T=1}^{\infty}$ in games with probability of mistakes $\varepsilon < \tilde{\varepsilon}$.

The argument behind Lemma 7 is similar to that of Proposition 2. See the Online Appendix for details.

4. Examples and Applications

This paper studies the long-run outcomes of observational learning in games. The examples that follow shed further light in this direction. First, I illustrate the role of mistakes with an example of *pure* observational learning (without payoff externalities). The second example illustrates the key role of the observation of others to attain strategic learning.

Third, I provide an example of a coordination game with multiple equilibria. In one equilibrium agents coordinate on the superior technology, but in a different one agents coordinate on a given technology, regardless of its inherent quality. Finally, I illustrate the long-run outcomes of games with preferences like those from Callander [2007] and Eyster et al. [2014].

4.1 Mistakes in Observational Learning without Payoff Externalities

EXAMPLE 3. STANDARD OBSERVATIONAL LEARNING. *Let* $u(1, X, 1) = u(0, X, 0) = 1$ *and* $u(1, X, 0) = u(0, X, 1) = 0$. *Each agent observes his immediate predecessor:* $M = 1$. *The signal structure is described by* $\nu_1[(0, s)] = s^2$ *and* $\nu_0[(0, s)] = 2s - s^2$ *with* $s \in (0, 1)$.

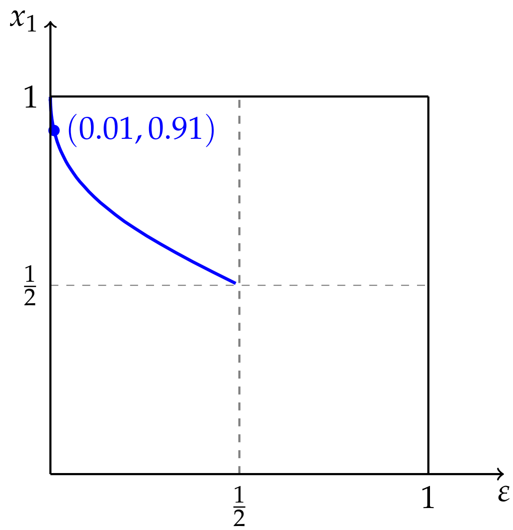
In this symmetric example, the average action X_1 represents the fraction of agents choosing the *right* action. Signals are of unbounded strength and the set $NE = (0, 1)$ is a singleton. Then, Proposition 2 guarantees that X_1 will be δ-close to 1, for low enough ε. This example provides a simple environment to illustrate what happens when ε is positive. What is the link between δ and ε? Is it true that (as the number of agents grows large) X_1 must approach $1 - \varepsilon$? This example shows that this is not the case.

The simple signal and observational structure in Example 3 allows for an analytical solution. As the number of agents grows large, the fraction of adopters of the superior technology approaches $\bar{x}_1 \equiv \frac{1-\varepsilon}{1-2\varepsilon}\left(1 - \sqrt{\frac{\varepsilon}{1-\varepsilon}}\right)$. See the Online Appendix for details. Figure 6(a) shows the long-run fraction of adopters of the superior technology $\bar{x}_1$ as a function of the probability of mistakes ε. For example, when $\varepsilon = 0.01$, $\bar{x}_1 \approx 0.91 < 1 - \varepsilon$.

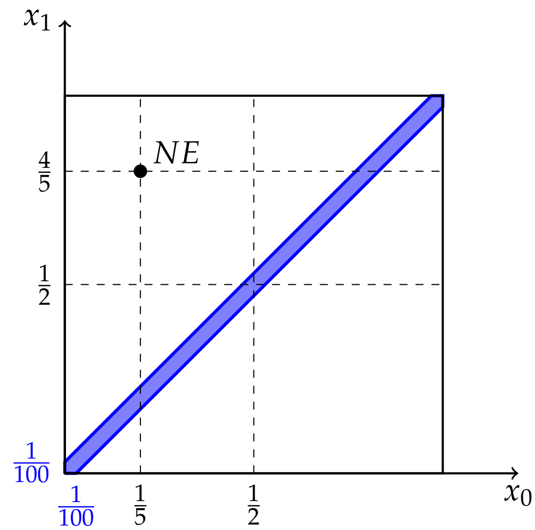
4.2 No Observation of Others' Actions

Consider next an anti-coordination game like the one presented in Example 1, but with agents who do *not* observe others' actions.

EXAMPLE 4. NO OBSERVATION OF OTHERS. *Let* $u(1, X, 0) = \frac{1}{5} - X$, $u(1, X, 1) = \frac{4}{5} - X$, *and* $u(0, X, \theta) = 0$. *Agents do not observe others' actions. The signal structure is as follows. Let* $S = \{0, \frac{1}{2}, 1\}$, *with* $d\nu_1(\frac{1}{2}) = d\nu_0(\frac{1}{2}) = 99/100$, $d\nu_1(1) = d\nu_0(0) = (1 - \gamma)/100$, *and* $d\nu_1(0) = d\nu_0(1) = \gamma$. *Let* $\gamma < 1/2$.



(a) Mistakes and Standard Obs. Learning



(b) No observation of others

Figure 6: Examples 3 and 4

A signal $s = \frac{1}{2}$ is uninformative about the state of the world. Signals $s = 0$ and $s = 1$ are instead informative. As γ gets smaller, signals become closer to being of unbounded strength. So Lemma 7 guarantees that the lower the γ, the closer one gets to strategic learning, *if agents observe the actions of some predecessors*. When there is no observation of others, information cannot get transmitted through actions. It is easy to see that at most $|E[X_1] - E[X_0]| \leq 1/100$. Outcomes outside of the shaded area in Figure 6(b) can never be attained without observing others.

4.3 Application to Common Payoff Functions in the Literature

EXAMPLE 5. CONGESTION. EXAMPLE 1 IN EYSTER ET AL. [2014]. *Payoffs are given by* $u(1, X, \theta) = \theta - kX$, $u(0, X, \theta) = 1 - \theta - k(1 - X)$. *Signals are of unbounded strength.*

An agent obtains an utility of one when he chooses the superior technology. On top of it, others who choose the same action as him exert a congestion effect of amount k.[9] The excess utility function is $v_\theta(X) = 2\theta - 2kX - (1 - k)$. When $k < 1$, $v_0(X) < 0$ and $v_1(X) >$

[9] In Eyster et al. [2014], only predecessors' actions have a negative effect. Instead, in this paper, it is both predecessors and successors. I have adapted the payoff function to account for this.

0 for all X. Then, $NE = \{(0,1)\}$. If instead $k \geq 1$, $NE = \{(\frac{1}{2} - \frac{1}{2k}, \frac{1}{2} + \frac{1}{2k})\}$. Signals are of unbounded strength, so Proposition 2 guarantees that there is strategic learning. The long-run outcome will be the unique element of NE. The analysis is analogous to that for the anti-coordination game presented in Example 1.

EXAMPLE 6. DESIRE TO CONFORM WITH THE MAJORITY. CALLANDER [2007]. *Payoffs are given by* $u(a_i, X, \theta) = \theta f(X) + (1-\theta)(1-f(X)) + k[a_i f(X) + (1-a_i)(1-f(X))]$. *The* continuous *and monotonically increasing function* $f(X)$ *has* $f(0) = 0$ *and* $f(1) = 1$. *Signals are of unbounded strength.*

There is an election with two candidates: zero and one. $f(X)$ denotes the probability that candidate one wins the election given that a fraction X choose him.[10] Each voter obtains a payoff of 1 if the better candidate gets elected. On top of it, he obtains a payoff of k if he votes for the better candidate. The excess utility function is $v_\theta(X) = k(2f(X) - 1)$. An individual cannot affect the result of the election. Then, only the cooperation component remains. The possible long-run outcomes are analogous to those in Example 2.

4.4 Multiple Equilibria in Coordination Games

EXAMPLE 7. COORDINATION. NO SELECTION OF EQUILIBRIA. *Payoffs are as in Example 2:* $u(1, X, 0) = X - \frac{2}{3}$, $u(1, X, 1) = X - \frac{1}{3}$, *and* $u(0, X, \theta) = 0$. *The signal structure is as follows. Let* $S = \{0, \frac{1}{2}, 1\}$, *with* $d\nu_1(\frac{1}{2}) = d\nu_0(\frac{1}{2}) = 99/100$, $d\nu_1(1) = d\nu_0(0) = (1-\gamma)/100$, *and* $d\nu_1(0) = d\nu_0(1) = \gamma$. *Let* $\gamma < 1/2$. *Each agent observes his immediate predecessor:* $M = 1$.

It is easy to show that there is an equilibrium where all agents choose action 1, regardless of what they observe. Under such strategy of play, when the number of agents grows large the proportion X is close to $1 - \varepsilon$ in *both* states of the world. Then, it is always optimal to choose action 1.

Interestingly, there is another equilibrium where agents coordinate on the superior technology. This equilibrium has a simple form. Take a sequence of symmetric strategy

[10] In Callander [2007], $f(X) = 1$ if $X < \frac{1}{2}$, $f(X) = 0$ if $X < \frac{1}{2}$, and $f(X) = \frac{1}{2}$ if $X = \frac{1}{2}$. Instead, in this paper, payoffs are continuous, so $f(X)$ must be continuous.

profiles where $\sigma^T(s, \xi) = \sigma(s, \xi)$ does not change with T and is given by:

$$\sigma(s, \xi) = \begin{cases} s & \text{if } s = \{0, 1\} \\ \xi & \text{if } s = 1/2 \end{cases}$$

Agents follow an informative signal and mimic their predecessor if the signal is uninformative. Under this profile of play, the proportion X is close to γ in state 0 and to $1 - \gamma$ in state 1 (for T large and ε small). This implies that an agent wants his action to match the state of the world. Moreover, the sample is informative about the state of the world. So indeed an agent who receives an uninformative signal copies the action of his predecessor. To sum up, for big enough T, strategy σ is an equilibrium.

5. Discussion

I study the long-run outcomes of observational learning with payoff externalities. In several economic situations, payoffs depend both on an uncertain state of the world and on others' actions. Individuals obtain information about their environment from private signals, and also by observing others. As agents need to learn both about the state of the world and about the play of others, informational externalities are confounded with coordination motives. Agents are uncertain about the true state of nature, so they do not know on which outcome to coordinate on. In addition, even if they knew the state, they would still not observe the aggregate play, so it would not be obvious which action to choose. Finally, a new strategic consideration arises with payoff externalities: agents may change their behavior in order to influence others.

I show that in spite of these confounding factors, there is strategic learning: agents' actions are ex-post optimal given the state of the world and the actions of others. As long as the number of agents grows large, and they sometimes make mistakes, each agent's individual influence on the aggregate outcome becomes negligible. Individuals are aware of this, and so they act as if they could not influence the aggregate play. In large games, the aggregate behavior becomes almost deterministic. I can then translate an environment with payoff externalities into one without them. I use then standard arguments in observational learning to show that information aggregates. Agents are ex-post satisfied with their actions in both states of the world.

References

ALI, S. AND N. KARTIK (2012): "Herding with collective preferences," *Economic Theory*, 51, 601–626.

ARIELI, I. (2017): "Payoff externalities and social learning," *Games and Economic Behavior*, 104, 392 – 410.

BANERJEE, A. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797–817.

BIKHCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1026.

CALLANDER, S. (2007): "Bandwagons and momentum in sequential voting," *The Review of Economic Studies*, 74, 653–684.

CHENG, S.-F., D. M. REEVES, Y. VOROBEYCHIK, AND M. P. WELLMAN (2004): "Notes on equilibria in symmetric games," in *Proceedings of the 6th International Workshop On Game Theoretic And Decision Theoretic Agents GTDT 2004*, 71–78.

CRIPPS, M. W. AND C. D. THOMAS (2016): "Strategic Experimentation in Queues," Working Paper.

DEKEL, E. AND M. PICCIONE (2000): "Sequential Voting Procedures in Symmetric Binary Elections," *Journal of Political Economy*, 108, 34+.

EYSTER, E., A. GALEOTTI, N. KARTIK, AND M. RABIN (2014): "Congested observational learning," *Games and Economic Behavior*, 87, 519–538.

GALLICE, A. AND I. MONZÓN (2016): "Cooperation in Social Dilemmas through Position Uncertainty," Working Paper.

LINDVALL, T. (1992): *Lectures on the Coupling Method*, Wiley Series in Probability and Statistics - Applied Probability and Statistics Section, Wiley.

MONZÓN, I. AND M. RAPP (2014): "Observational Learning with Position Uncertainty," *Journal of Economic Theory*, 154, 375 – 402.

SMITH, L. AND P. SØRENSEN (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68, 371–398.

A. Proofs

A.1 Proof of Lemma 1

The proof of existence of a symmetric equilibrium builds upon Theorem 3 in Cheng, Reeves, Vorobeychik, and Wellman [2004]. Cheng et al. [2004] show that a pure strategy symmetric equilibrium exists in symmetric infinite games with compact, convex strategy sets and continuous and quasiconcave utility functions. I first present Theorem 3 in Cheng et al. [2004] and then show how it applies to the environment in the present paper.

For each player $i \in \mathcal{I}$, let R_i be a set of strategies (with $\rho_i \in R_i$). Agent i's payoffs from profile $(\rho_1, \ldots, \rho_T)$ are denoted by $u_i(\rho_1, \ldots, \rho_T)$. The tuple $\left[\mathcal{I}, \{R_i\}_{i=1}^{T}, \{u_i\}_{i=1}^{T}\right]$ denotes a game.

DEFINITION. SYMMETRIC GAMES (DEFINITION 2 IN CHENG ET AL. [2004]). *A normal-form game is symmetric if the players have identical strategy spaces ($R_i = R$ for all $i \in \mathcal{I}$) and $u_i(\rho_i, \rho_{-i}) = u_j(\rho_j, \rho_{-j})$ for $\rho_i = \rho_j$ and $\rho_{-i} = \rho_{-j}$ for all $i, j \in \mathcal{I}$. Thus we can write $u(\rho_i, \rho_{-i})$ for the utility to any player playing strategy ρ_i in profile ρ. Then, the tuple $[\mathcal{I}, R, u()]$ denotes a symmetric game.*

THEOREM 1. (THEOREM 3 IN CHENG ET AL. [2004]). *A symmetric game $[\mathcal{I}, R, u()]$ with R a nonempty, convex, and compact subset of some Euclidean space and $u(\rho_i, \rho_{-i})$ continuous in (ρ_i, ρ_{-i}) and quasiconcave in ρ_i has a symmetric pure-strategy equilibrium.*

In the current paper, agent i's strategy is a function $\sigma_i : S \times \Xi \to [\varepsilon, 1-\varepsilon]$, with $\sigma_i \in \Sigma$. I collapse the strategy σ_i into the likelihood $\rho_i(\xi, \theta)$ of choosing action 1 given the sample received and the state of the world. Formally, define $\rho_i(\xi, \theta) \equiv \mathbf{P}_{\sigma_i}(a_i = 1 \mid \theta, \xi)$. There is a many to one mapping $\sigma_i \mapsto \rho_i$. It is without loss of generality to work directly with agents choosing ρ_i from the feasible set

$$R_i = \left\{ \rho_i : \rho_i(\xi, \theta) = E\left[\sigma_i\left(S_{Q(i)}, \xi \right) \mid \theta \right] \text{ for some } \sigma_i \in \Sigma_i \right\}.$$

The set of strategies Σ is the same for all agents, so $R_i = R$ for all $i \in \mathcal{I}$. Conveniently, R is a subset of an Euclidean space of dimension $|\Xi| \cdot |\Theta|$. R is non-empty and compact (see Appendix A.2 in Monzón and Rapp [2014] for the proof). Next, take $\rho_i \in R$ and $\rho_i' \in R$, with ρ_i derived from σ_i and ρ_i' from σ_i'. Then

$$\begin{aligned}
\alpha\rho_i(\xi,\theta) + (1-\alpha)\rho_i'(\xi,\theta) &= \alpha E\left[\sigma_i(S_{Q(i)},\xi) \mid \theta\right] + (1-\alpha) E\left[\sigma_i'(S_{Q(i)},\xi) \mid \theta\right] \\
&= E\left[\alpha\sigma_i(S_{Q(i)},\xi) + (1-\alpha)\sigma_i'(S_{Q(i)},\xi) \mid \theta\right]
\end{aligned}$$

As Σ is convex, then $\alpha\rho_i(\xi,\theta) + (1-\alpha)\rho_i'(\xi,\theta) \in R$, so R is convex. Agent i's expected utility as a function of ρ becomes

$$\begin{aligned}
u_i(\rho_i, \rho_{-i}) = \frac{1}{2} \sum_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \sum_{h_t \in \mathcal{H}_t} &\mathbf{P}_{\rho_{-i}}(H_t = h_t \mid \theta) \sum_{\xi \in \Xi} \Pr(\xi \mid h_t) \\
\times \Big[&\rho_i(\xi, \theta) E_{\rho_{-i}}\left[u(1, X, \theta) \mid \theta, h_t, a_t = 1\right] \\
&+ (1 - \rho_i(\xi, \theta)) E_{\rho_{-i}}\left[u(0, X, \theta) \mid \theta, h_t, a_t = 0\right] \Big]
\end{aligned}$$

It is simple to see that $u_i(\rho_i, \rho_{-i})$ is continuous in ρ_i. Others' ρ_{-i} affect $u_i(\rho_i, \rho_{-i})$ through two channels. First, they affect the distribution of H_t. Second, they affect the distribution of X_θ. Utility $u_i(\rho_i, \rho_{-i})$ is continuous in ρ_{-i} through both channels (note that $u(a_i, X, \theta)$ is continuous in X). Therefore, payoffs $u_i(\rho_i, \rho_{-i})$ are continuous in ρ. Finally, note that $u_i(\rho_i, \rho_{-i})$ is linear in $\rho_i(\xi, \theta)$, so it $u(\rho_i, \rho_{-i})$ is quasiconcave in ρ_i. Then, by Theorem 3 in Cheng et al. [2004] there exists $\rho^* \in R$ such that ρ^* is a best response to $\rho_{-i} = (\rho^*, \ldots, \rho^*)$. Thus, if each agent plays a strategy σ^* that maps to ρ^*, all play a best response. As a result, there exists a symmetric equilibrium σ^* of the game. ■

A.2 Proof of Proposition 1

I present first an intermediate Lemma. Let $P = (p_{ij})$ be a transition matrix on a finite state space $\mathcal{Y}$. Assume that the Markov Chain $Y = (Y_n)_{n=0}^{\infty}$ associated with P is aperiodic and irreducible. Let μ denote the unique stationary distribution of Y.

Lemma 8. *Let $\mathcal{Y}^1 \subseteq \mathcal{Y}$ be a non-empty subset of the state space and $\mu^1 \equiv \sum_{y \in \mathcal{Y}^1} \mu_y$. Then, there exists $\rho > 0$ and $K > 0$ such that for any distribution over states in period t:*

$$\left| \Pr\left(Y_{t+n} \in \mathcal{Y}^1\right) - \mu^1 \right| \leq 2(1-\rho)^{(n-K)/K}$$

where K is large enough so that $\rho = \min_{i,j} p_{ij}^{(K)} > 0$.

Proof. The proof is based on a standard coupling argument. It follows closely sections 2.7 and 2.8 of Lindvall [1992]. Let Y'_n be the Markov Chain with transition matrix P but started at the stationary distribution μ. Instead, let Y_n be the Markov Chain with transition matrix P but started at some distribution λ. Let N be the first period in which these two chains meet: $N = \min\left\{k : Y_k = Y'_k\right\}$. Finally let Y''_n be given by:

$$Y''_n = \begin{cases} Y_n & \text{if } n < N \\ Y'_n & \text{if } n \geq N \end{cases}$$

Then

$$\begin{aligned}
\left|\Pr(Y_n = y) - \mu_y\right| &= \left|\Pr(Y_n = y) - \Pr\left(Y'_n = y\right)\right| \\
&= \Big|\Pr(Y_n = y, N \leq n) + \Pr(Y_n = y, N > n) \\
&\qquad - \Pr\left(Y'_n = y, N \leq n\right) - \Pr\left(Y'_n = y, N > n\right)\Big| \\
&= \left|\Pr(Y_n = y, N > n) - \Pr\left(Y'_n = y, N > n\right)\right| \\
\left|\Pr(Y_n = y) - \mu_y\right| &\leq \Pr(Y_n = y, N > n) + \Pr\left(Y'_n = y, N > n\right)
\end{aligned}$$

For the subset $\mathcal{Y}^1 \subseteq \mathcal{Y}$, $\Pr\left(Y_n \in \mathcal{Y}^1\right) = \sum_{y \in \mathcal{Y}^1} \Pr\left(Y_n = y\right)$, so

$$\begin{aligned}
\Pr\left(Y_n \in \mathcal{Y}^1\right) - \mu^1 &= \sum_{x \in \mathcal{Y}^1} \Pr\left(Y_n = y\right) - \mu^1 = \sum_{x \in \mathcal{Y}^1} \left[\Pr\left(Y_n = y\right) - \mu_y\right] \\
\left|\Pr\left(Y_n \in \mathcal{Y}^1\right) - \mu^1\right| &\leq \sum_{x \in \mathcal{Y}^1} \left|\Pr\left(Y_n = y\right) - \mu_y\right| \\
&\leq \sum_{x \in \mathcal{Y}^1} \Pr\left(Y_n = y, N > n\right) + \sum_{x \in \mathcal{Y}^1} \Pr\left(Y'_n = x, N > n\right) \\
&\leq 2 \Pr\left(N > n\right)
\end{aligned}$$

Since $\mathcal{Y}$ is finite, and the Markov Chain Y is irreducible and aperiodic, there exists a finite $K > 0$ large enough so that: $\rho = \min_{i,j} p_{ij}^{(K)} > 0$. Then, for any two distributions μ and λ,

$$\Pr\left(N > n\right) = \Pr\left(Y_i \neq Y'_i \quad \forall\, i \leq n\right) \leq (1 - \rho)^{\lfloor n/K \rfloor},$$

where $\lfloor n/K \rfloor$ is the integer part of n/K. To avoid using $\lfloor n/K \rfloor$, note that $\lfloor n/K \rfloor \geq n/K - 1 = (n - K)/K$. Then, $\left|\Pr\left(Y_n \in \mathcal{Y}^1\right) - \mu^1\right| \leq 2(1-\rho)^{(n-K)/K}$. ■

With Lemma 8 in hand, I turn to the proof of Proposition 1. Let $\{\sigma^\tau\}_{\tau=1}^{\infty}$ be a sequence of symmetric strategy profiles. After the first M periods, all samples are of size M. Let $\mathcal{Y} = \{0,1\}^M$ be the set of all possible histories of length M. Each symmetric strategy profile σ^τ induces a Markov Chain $Y^\tau = (Y_t)_{t \geq M}$ over $\mathcal{Y}$. Since mistakes occur with positive probability, these Markov Chains are irreducible and aperiodic. Then, each Y^τ has a unique stationary distribution, which I denote by μ^τ. After exactly M periods, transition probabilities are bounded below: $\min_{y,y' \in \mathcal{Y} \times \mathcal{Y}} \Pr\left(Y_{n+M} = y' \mid Y_n = y\right) \geq \varepsilon^M$. The lower bound ε^M is independent of the strategy profile σ^τ.

Let $\mathcal{Y}^1$ be all histories where the last agent chose action $a = 1$ and let $\bar{\mu}^\tau \equiv \sum_{y \in \mathcal{Y}^1} \mu_y^\tau$. Then, Lemma 8 guarantees that *for any distribution over states in period t*:

$$\left|\Pr\left(Y_{t+n}^\tau \in \mathcal{Y}^1\right) - \bar{\mu}^\tau\right| \leq 2\left(1 - \varepsilon^M\right)^{(n-M)/M} = 2\left[\left(1 - \varepsilon^M\right)^{\frac{1}{M}}\right]^{n-M} \equiv 2\delta^{(n-M)} \equiv c\delta^n \tag{5}$$

This bound holds for any symmetric strategy profile σ^τ.

In what follows, I fix a state of the world θ, so from now on I drop the subindex θ. Also, I fix a strategy profile σ^τ. I use τ to index strategy profiles and T to index the number of agents. Let $V(\sigma^\tau)$ denote the variance of $X|\sigma^\tau$, for any number of players T: $V(\sigma^\tau) \equiv E_{\sigma^\tau}\left[(X - E_{\sigma^\tau}[X \mid \theta])^2 \mid \theta\right]$. I show that for any $\widetilde{\delta} > 0$ there exists $\widetilde{T} < \infty$ such that: $E_{\sigma^\tau}\left[X^2 \mid \theta\right] - (E_{\sigma^\tau}[X \mid \theta])^2 < \delta$ for all $T > \widetilde{T}$ and for all τ. This implies that

$$\lim_{T\to\infty} E_{\sigma^T}\left[X^2 \mid \theta\right] - (E_{\sigma^T}[X \mid \theta])^2 = 0$$

that is, $X|\sigma^T - E\left[X|\sigma^T\right]$ converges to zero in L^2 norm, which implies convergence in probability.

Fix a strategy profile σ^τ and define $V(\sigma^\tau)$ as follows:

$$\begin{aligned} V(\sigma^\tau) &\equiv E_{\sigma^\tau}\left[\left(\frac{1}{T}\sum_{t=1}^{T} a_t\right)^2\right] - \left(E_{\sigma^\tau}\left[\frac{1}{T}\sum_{t=1}^{T} a_t\right]\right)^2 \\ &= \frac{1}{T^2}\left[\sum_{t=1}^{T}\left(E_{\sigma^\tau}\left[a_t^2\right] - E_{\sigma^\tau}[a_t]^2\right) + 2\sum_{t=1}^{T}\sum_{n=1}^{T-t}\left(E_{\sigma^\tau}[a_t a_{t+n}] - E_{\sigma^\tau}[a_t]\, E_{\sigma^\tau}[a_{t+n}]\right)\right] \end{aligned} \tag{6}$$

It is easy to see that $\sum_{t=1}^{T}\left(E_{\sigma^\tau}\left[a_t^2\right] - E_{\sigma^\tau}[a_t]^2\right) \leq T$. Regarding the remaining terms, note that

$$\begin{aligned} E_{\sigma^\tau}[a_t a_{t+n}] - E_{\sigma^\tau}[a_t]\, E_{\sigma^\tau}[a_{t+n}] &= \mathbf{P}_{\sigma^\tau}(a_t = 1)\, \mathbf{P}_{\sigma^\tau}(a_{t+n} = 1 \mid a_t = 1) \\ &\quad - \mathbf{P}_{\sigma^\tau}(a_t = 1)\, \mathbf{P}_{\sigma^\tau}(a_{t+n} = 1) \\ &= \mathbf{P}_{\sigma^\tau}(a_t = 1)\left[\mathbf{P}_{\sigma^\tau}(a_{t+n} = 1 \mid a_t = 1) - \mathbf{P}_{\sigma^\tau}(a_{t+n} = 1)\right] \\ &\leq \left|\mathbf{P}_{\sigma^\tau}(a_{t+n} = 1 \mid a_t = 1) - \mathbf{P}_{\sigma^\tau}(a_{t+n} = 1)\right| \end{aligned}$$

Given equation (5), $|\mathbf{P}_{\sigma^\tau}(a_{t+n} = 1 \mid a_t = 1) - \bar{\mu}| < c\delta^n$ and $|\mathbf{P}_{\sigma^\tau}(a_{t+n} = 1) - \bar{\mu}| < c\delta^{(t+n)}$ for any σ^τ. Then,

$$\left|\mathbf{P}_{\sigma^\tau}(a_{t+n} = 1 \mid a_t = 1) - \mathbf{P}_{\sigma^\tau}(a_{t+n} = 1)\right| < c\delta^n + c\delta^{t+n} \leq 2c\delta^n$$

So the second term in equation (6) becomes:

$$
\begin{aligned}
2\sum_{t=1}^{T}\sum_{n=1}^{T-t}\left(E_{\sigma^\tau}\left[a_t a_{t+n}\right]-E_{\sigma^\tau}\left[a_t\right]E_{\sigma^\tau}\left[a_{t+n}\right]\right) &\leq 2\sum_{t=1}^{T}\sum_{n=1}^{T-t}2c\delta^n = 4c\sum_{t=1}^{T}\sum_{n=1}^{T-t}\delta^n \\
&\leq 4c\sum_{t=1}^{T}\frac{\delta\left(1-\delta^{T-t}\right)}{1-\delta} \leq 4c\sum_{t=1}^{T}\frac{\delta}{1-\delta} \\
&= 4c\frac{\delta}{1-\delta}T
\end{aligned}
$$

Then, for all σ^τ

$$
V(\sigma^\tau) \leq \frac{1}{T}\left(1+4c\frac{\delta}{1-\delta}\right)
$$

where $\left(1+4c\frac{\delta}{1-\delta}\right)$ is independent of σ.

Then, pick any $b>0$. There exists $\widetilde{T}$ such that for all $T>\widetilde{T}$, and for all σ^τ, $V(\sigma^\tau)<b$. So in particular, for all $b>0$ there exists $\widetilde{T}$ such that for all $T>\widetilde{T}$, $V(\sigma^T)<b$. That is, $V(\sigma^T)\to 0$. ■

The proof of the second part of Proposition 1 is as follows. Let agent i be in position $t=Q(i)$. Define two Markov Chains, both with the same transition matrix P. These chains start right after agent i plays. Their only difference is the starting distribution over states. First, $(Y_n)_{n\geq t+1}$ has agent i following strategy σ_i. Second, $\left(\widetilde{Y}_n\right)_{n\geq t+1}$ has agent i following strategy $\widetilde{\sigma}_i$. As before, let N be the first period in which these two chains meet. By equation (5), $\Pr(N>n)\leq c\delta^n$. Note that for any $N=n$,

$$
\begin{aligned}
X|\sigma^T - X|\widetilde{\sigma}^T = \frac{1}{T}\Bigg[&\sum_{t=1}^{Q(i)-1}\left(a_t|\sigma^T - a_t|\widetilde{\sigma}^T\right) + \sum_{t=Q(i)}^{Q(i)+n-1}\left(a_t|\sigma^T - a_t|\widetilde{\sigma}^T\right) \\
&+ \sum_{t=Q(i)+n}^{T}\left(a_t|\sigma^T - a_t|\widetilde{\sigma}^T\right)\Bigg]
\end{aligned}
$$

But $a_t|\sigma^T = a_t|\widetilde{\sigma}^T$ for $t\in\{1,Q(i)-1\}$ and for $t\in\{Q(i)+n,T\}$. Then,

$$
\left|X|\sigma^T - X|\widetilde{\sigma}^T\right| = \left|\frac{1}{T}\sum_{t=Q(i)}^{Q(i)+n-1}\left(a_t|\sigma^T - a_t|\widetilde{\sigma}^T\right)\right| \leq \frac{n}{T}
$$

To sum up, for any strategy profile σ^T,

$$\Pr\left(\left|X|\sigma^T - X|\widetilde{\sigma}^T\right| \geq \frac{n}{T}\right) \leq c\delta^n$$

Then for all $b > 0$, there exists n such that $b \geq c\delta^n$. Fix b and n. There is always a T, so that $n/T < b$. Then,

$$\Pr\left(\left|X|\sigma^T - X|\widetilde{\sigma}^T\right| \geq b\right) \leq \Pr\left(\left|X|\sigma^T - X|\widetilde{\sigma}^T\right| \geq \frac{n}{T}\right) \leq c\delta^n \leq b$$

Finally, note that both $X|\sigma^T - X|\widetilde{\sigma}^T \xrightarrow{p} 0$ and $X|\sigma^T - E\left[X|\sigma^T\right] \xrightarrow{p} 0$. Then also $X|\widetilde{\sigma}^T - E\left[X|\sigma^T\right] \xrightarrow{p} 0$. ■

A.3 Proof of Corollary 1

The distance $d(X, L)$ can be bounded above as follows:

$$\begin{aligned} d(X,L) &= \min_{y\in L}|X-y| \leq \min_{y\in L}\left[|X - E_{\sigma^T}[X]| + |E_{\sigma^T}[X] - y|\right] \\ &\leq |X - E_{\sigma^T}[X]| + \min_{y\in L}|E_{\sigma^T}[X] - y| = |X - E_{\sigma^T}[X]| + d\left(E_{\sigma^T}[X], L\right) \end{aligned}$$

The set L includes all limit points for convergent subsequences of $\{E_{\sigma^T}[X]\}_{T=1}^{\infty}$. Then $\lim_{T\to\infty} d\left(E_{\sigma^T}[X], L\right) = 0$. For some $\widetilde{T}$ large enough, $d\left(E_{\sigma^T}[X], L\right) < \delta/2$ for all $T > \widetilde{T}$. Then $\mathbf{P}_{\sigma^T}\left(d(X,L) < \delta\right) \geq \mathbf{P}_{\sigma^T}\left(|X - E_{\sigma^T}[X]| < \delta/2\right)$. Finally, Proposition 1 guarantees that $\lim_{T\to\infty}\mathbf{P}_{\sigma^T}\left(|X - E_{\sigma^T}[X]| < \delta/2\right) = 1$. ■

A.4 Proof of Lemma 3

Agent i's expected utility $u\left(\widetilde{\sigma}_i^T, \sigma_{-i}^T\right)$ is given by:

$$\begin{aligned} u\left(\widetilde{\sigma}_i^T, \sigma_{-i}^T\right) &= E_{\widetilde{\sigma}^T}\left[u\left(a_i, X, \theta\right)\right] = \frac{1}{2}\sum_{\theta\in\{0,1\}}\sum_{a\in\mathcal{A}} E_{\widetilde{\sigma}^T}\left[u\left(a, X, \theta\right)\mathbb{1}\left\{a_i = a\right\} \mid \theta\right] \\ &= \frac{1}{2}\sum_{\theta\in\{0,1\}}\sum_{a\in\mathcal{A}} E_{\widetilde{\sigma}^T}\left[u\left(a, X_\theta, \theta\right) \mid a_i = a\right]\mathbf{P}_{\widetilde{\sigma}^T}\left(a_i = a \mid \theta\right). \end{aligned}$$

Then,

$$u\left(\widetilde{\sigma}_i^T,\sigma_{-i}^T\right)-\widetilde{u}^T=\frac{1}{2}\sum_{\theta\in\{0,1\}}\sum_{a\in\mathcal{A}}\mathbf{P}_{\widetilde{\sigma}^T}\left(a_i=a\mid\theta\right)$$
$$\times\left[E_{\widetilde{\sigma}^T}\left[u\left(a,X_\theta,\theta\right)\mid a_i=a\right]-u\left(a,E_{\sigma^T}\left[X_\theta\right],\theta\right)\right]$$

If $\lim_{T\to\infty}\mathbf{P}_{\widetilde{\sigma}^T}\left(a_i=a\mid\theta\right)=0$, then trivially

$$\lim_{T\to\infty}\mathbf{P}_{\widetilde{\sigma}^T}\left(a_i=a\mid\theta\right)\left[E_{\widetilde{\sigma}^T}\left[u\left(a,X_\theta,\theta\right)\mid a_i=a\right]-u\left(a,E_{\sigma^T}\left[X_\theta\right],\theta\right)\right]=0.$$

Assume instead that there exists $\delta>0$ such that $\mathbf{P}_{\widetilde{\sigma}^T}\left(a_i=a\mid\theta\right)\geq\delta$ infinitely often. By Proposition 1, for any $\delta>0$, $\lim_{T\to\infty}\mathbf{P}_{\widetilde{\sigma}^T}\left(\left|X_\theta-E_{\sigma^T}\left[X_\theta\right]\right|\geq\delta\right)=0$. Then, it is also true that for any $\delta>0$, $\lim_{T\to\infty}\mathbf{P}_{\widetilde{\sigma}^T}\left(\left|X_\theta-E_{\sigma^T}\left[X_\theta\right]\right|\geq\delta\mid a_i=a\right)=0$.[11] So by Portmanteau's Theorem, $\lim_{T\to\infty}E_{\widetilde{\sigma}^T}\left[u\left(a,X_\theta,\theta\right)\mid a_i=a\right]=\lim_{T\to\infty}u\left(a,E_{\sigma^T}\left[X_\theta\right],\theta\right)$. This leads directly to $\lim_{T\to\infty}\left[u\left(\widetilde{\sigma}_i^T,\sigma_{-i}^T\right)-\widetilde{u}^T\right]=0$. ■

A.5 Proof of Corollary 2

$$\begin{aligned}\limsup_{T\to\infty}\Delta^T&=\limsup_{T\to\infty}\left[\widetilde{u}^T-u\left(\widetilde{\sigma}_i^T,\sigma_{-i}^T\right)+u\left(\widetilde{\sigma}_i^T,\sigma_{-i}^T\right)-u\left(\sigma^T\right)+u\left(\sigma^T\right)-\bar{u}^T\right]\\&\leq\limsup_{T\to\infty}\left[\widetilde{u}^T-u\left(\widetilde{\sigma}_i^T,\sigma_{-i}^T\right)\right]+\limsup_{T\to\infty}\left[u\left(\widetilde{\sigma}_i^T,\sigma_{-i}^T\right)-u\left(\sigma^T\right)\right]\\&\quad+\limsup_{T\to\infty}\left[u\left(\sigma^T\right)-\bar{u}^T\right]\end{aligned}$$

Lemmas 2 and 3 imply that $\lim_{T\to\infty}\left[u\left(\sigma^T\right)-\bar{u}^T\right]$ and $\lim_{T\to\infty}\left[\widetilde{u}^T-u\left(\widetilde{\sigma}_i^T,\sigma_{-i}^T\right)\right]=0$, respectively. Next, σ^T are equilibrium strategies, so $u\left(\widetilde{\sigma}_i^T,\sigma_{-i}^T\right)-u\left(\sigma^T\right)\leq0$ for all σ_{-i}^T and for all T. These two facts together imply that

$$\limsup_{T\to\infty}\Delta^T\leq\limsup_{T\to\infty}\left[u\left(\widetilde{\sigma}_i^T,\sigma_{-i}^T\right)-u\left(\sigma^T\right)\right]\leq0.\ ■$$

[11] To see this note that: $\mathbf{P}_{\widetilde{\sigma}}\left(\left|X_\theta-E_\sigma\left[X_\theta\right]\right|>\delta\right)=\sum_{a\in\mathcal{A}}\mathbf{P}_{\widetilde{\sigma}}\left(\left|X_\theta-E_\sigma\left[X_\theta\right]\right|>\delta\mid\theta,a_i=a\right)\mathbf{P}_{\widetilde{\sigma}}\left(a_i=a\mid\theta\right)$. By Proposition 1, $\lim_{T\to\infty}\mathbf{P}_{\widetilde{\sigma}^T}\left(\left|X_\theta-E_{\sigma^T}\left[X_\theta\right]\right|>\delta\right)=0$. Then, if $\mathbf{P}_{\widetilde{\sigma}^T}\left(a_i=a\mid\theta\right)\geq\delta$ infinitely often, it must be the case that $\lim_{T\to\infty}\mathbf{P}_{\widetilde{\sigma}^T}\left(\left|X_\theta-E_{\sigma^T}\left[X_\theta\right]\right|>\delta\mid a_i=a\right)=0$.

A.6 Proof of Lemma 4

Lemma 4 deals with the case in which an action is dominant (either weakly or strictly) in the limit. Consider two alternative strategies, $\widetilde{\sigma}^0$: "always play action 0", and $\widetilde{\sigma}^1$: "always play action 1". Define accordingly $\Delta^{0,T} \equiv \frac{1}{2}\sum_{\theta\in\{0,1\}}\left(\varepsilon - E_{\sigma^T}\left[X_\theta\right]\right)v_\theta\left(E_{\sigma^T}\left[X_\theta\right]\right)$ and $\Delta^{1,T} \equiv \frac{1}{2}\sum_{\theta\in\{0,1\}}\left(1-\varepsilon - E_{\sigma^T}\left[X_\theta\right]\right)v_\theta\left(E_{\sigma^T}\left[X_\theta\right]\right)$. Then, by Corollary 2,

$$\limsup_{T\to\infty}\Delta^{0,T} = \frac{1}{2}\sum_{\theta\in\{0,1\}}\left(\varepsilon - x_\theta\right)v_\theta\left(x_\theta\right) \leq 0 \quad \text{and}$$
$$\limsup_{T\to\infty}\Delta^{1,T} = \frac{1}{2}\sum_{\theta\in\{0,1\}}\left(1-\varepsilon - x_\theta\right)v_\theta\left(x_\theta\right) \leq 0. \blacksquare$$

Next, assume $v_0(x_0)v_1(x_1) \geq 0$. Then, if $v_\theta(x_\theta) < 0$, equation (1) requires $x_\theta = \varepsilon$. If, on the other side, $v_\theta(x_\theta) > 0$, equation (2) requires $x_\theta = 1-\varepsilon$.

The rest of the proof is a direct result of the following Lemma:

Lemma 9. *If $x_\theta = \varepsilon$ for some $\theta \in \{0,1\}$, then $x = (\varepsilon, \varepsilon)$. Similarly, if $x_\theta = 1-\varepsilon$ for some $\theta \in \{0,1\}$, then $x = (1-\varepsilon, 1-\varepsilon)$.*

Proof. Assume that $x_1 = 1-\varepsilon$, but $x_0 \neq 1-\varepsilon$. The proof is analog for all other cases. The expected proportion $E_{\sigma^T}\left[X_\theta\right]$ can be expressed as follows:

$$E_{\sigma^T}\left[X_\theta\right] = E_{\sigma^T}\left[\frac{1}{T}\sum_{t=1}^{T}a_t \mid \theta\right] = \frac{1}{T}\sum_{t=1}^{T}E_{\sigma^T}\left[a_t \mid \theta\right] = \frac{1}{T}\sum_{t=1}^{T}\mathbf{P}_{\sigma^T}\left(a_t = 1 \mid \theta\right)$$
$$= \mathbf{P}_{\sigma^T}\left(a_i = 1 \mid \theta\right) = \sum_{\xi\in\Xi}\mathbf{P}_{\sigma^T}\left(\xi \mid \theta\right)\int_{s\in S}\sigma^T(s,\xi)dv_\theta(s)$$

Let $\Xi^M \subset \Xi$ be the set of all samples with exactly M actions. All agents in positions $M < t \leq T$ receive samples $\xi_t \in \Xi^M$. Since mistakes occur with positive probability $\varepsilon > 0$, all samples $\xi \in \Xi^M$ occur with positive probability: $\mathbf{P}_{\sigma^T}\left(\xi \mid \theta\right) \geq \varepsilon^M$ for any strategy profile σ^T. Then, $\lim_{T\to\infty}\int_{s\in S}\sigma^T(s,\xi)dv_1(s) = 1-\varepsilon$ for all $\xi \in \Xi^M$. Since $\sigma^T(s,\xi) \leq 1-\varepsilon$, then, for any $\tilde{c} > 0$

$$\lim_{T\to\infty}\int_{s\in S}\mathbb{1}\left\{\sigma^T(s,\xi) \geq 1-\varepsilon-\tilde{c}\right\}dv_1(s) = 1.$$

I show next that the previous equation must also hold for measure ν_0. That is, for all $\tilde{c} > 0$

$$\lim_{T\to\infty} \int_{s\in S} \mathbb{1}\left\{\sigma^T(s,\xi) \geq 1-\varepsilon-\tilde{c}\right\} d\nu_0(s) = 1. \tag{7}$$

which implies that $\lim_{T\to\infty} \int_{s\in S} \sigma^T(s,\xi) d\nu_0(s) = 1-\varepsilon$ for all $\xi \in \Xi^M$, and so $x_0 = 1-\varepsilon$.

To see why equation (7) must hold for measure ν_0, consider the sequence of sets $\{S^t\}_{t=1}^{\infty}$ with $S^t = \mathbb{1}\{\sigma^T(s,\xi) < 1-\varepsilon-\tilde{c}\}$. We know that $\lim_{T\to\infty} \int_{s\in S^t} d\nu_1(s) = 0$. Assume that for some $c > 0$, $\int_{s\in S^t} d\nu_0(s) \geq c > 0$ for all t. Pick $l \in (0,\infty)$ such that[12]

$$0 < \int_{\{s:l(s)\leq l\}} d\nu_0(s) \leq c.$$

Then,

$$\begin{aligned}
\int_{\{s:l(s)\leq l\}} d\nu_0(s) &\leq c \leq \int_{s\in S^t} d\nu_0(s) \\
\int_{\{s:l(s)\leq l, s\notin S^t\}} d\nu_0(s) &\leq \int_{\{s:l(s)>l, s\in S^t\}} d\nu_0(s) \\
\int_{\{s:l(s)\leq l, s\notin S^t\}} l(s)^{-1} d\nu_1(s) &\leq \int_{\{s:l(s)>l, s\in S^t\}} l(s)^{-1} d\nu_1(s) \\
l^{-1} \int_{\{s:l(s)\leq l, s\notin S^t\}} d\nu_1(s) &\leq l^{-1} \int_{\{s:l(s)>l, s\in S^t\}} d\nu_1(s) \\
G_1(l) = \int_{\{s:l(s)\leq l\}} d\nu_1(s) &\leq \int_{\{s\in S^t\}} d\nu_1(s)
\end{aligned}$$

Because of absolute continuity, since $G_0(l) > 0$, then $G_1(l) > 0$. So for all elements of $\{S^t\}_{t=1}^{\infty}$, $\int_{s\in S^t} d\nu_1(s) \geq G_1(l) > 0$. Then, $\int_{s\in S^t} d\nu_1(s)$ cannot converge to zero. ∎

A.7 Proof of Lemma 5

Let $\pi_\theta^T \equiv \mathbf{P}_{\sigma^T}\left(\widetilde{\xi} = 1 \mid \theta\right)$. I show first the following intermediate lemma.

Lemma 10. *Let $T \geq 2M$. For any sequence of strategy profiles $\{\sigma^T\}_{T=1}^{\infty}$, and for $\theta \in \{0,1\}$,* $\lim_{T\to\infty} \pi_\theta^T - E_{\sigma^T}[X_\theta] = 0$.

[12] It may happen that the lowest possible interval $\{s : l(s) \leq l\}$ with positive mass starts with a mass point (say at $\tilde{l}$). If so, its mass may be $\int_{\{s:l(s)\leq\tilde{l}\}} d\nu_0(s) > c$. In such a case, consider $\alpha \in (0,1)$ with $\alpha \int_{\{s:l(s)\leq\tilde{l}\}} d\nu_0(s) = c$. The same argument holds.

Proof. Fix the state of the world θ.

$$
\begin{aligned}
\pi_\theta^T &= \frac{1}{T}\sum_{t=1}^{T}\mathbf{P}_{\sigma^T}\left(\widetilde{\xi}_t = 1\right) \\
&= \frac{1}{T}\left[\sum_{t=2}^{M}\frac{1}{t-1}\sum_{\tau=1}^{t-1}\mathbf{P}_{\sigma^T}\left(a_\tau = 1\right) + \sum_{t=M+1}^{T}\frac{1}{M}\sum_{\tau=t-M}^{t-1}\mathbf{P}_{\sigma^T}\left(a_\tau = 1\right)\right] \\
&= \frac{1}{T}\left[\sum_{\tau=1}^{T-1}\mathbf{P}_{\sigma^T}\left(a_\tau = 1\right)\sum_{t=\tau+1}^{\min\{\tau+M,T\}}\left[\min\left\{t-1, M\right\}\right]^{-1}\right] \\
&= \frac{1}{T}\left[\sum_{\tau=1}^{M-1}\mathbf{P}_{\sigma^T}\left(a_\tau = 1\right)\sum_{t=\tau+1}^{\tau+M}(t-1)^{-1} + \sum_{\tau=M}^{T-M}\mathbf{P}_{\sigma^T}\left(a_\tau = 1\right)\sum_{t=\tau+1}^{\tau+M}M^{-1}\right. \\
&\qquad \left. + \sum_{\tau=T-M+1}^{T-1}\mathbf{P}_{\sigma^T}\left(a_\tau = 1\right)\sum_{t=\tau+1}^{T}M^{-1}\right] \\
&= \frac{1}{T}\sum_{\tau=1}^{T}\mathbf{P}_{\sigma^T}\left(a_\tau = 1\right) \\
&\qquad + \frac{1}{T}\left[\sum_{\tau=1}^{M-1}\mathbf{P}_{\sigma^T}\left(a_\tau = 1\right)\left(\sum_{t=\tau}^{\tau+M-1}t^{-1} - 1\right) - \sum_{\tau=T-M+1}^{T}\mathbf{P}_{\sigma^T}\left(a_\tau = 1\right)\left(1 - \frac{T-\tau}{M}\right)\right]
\end{aligned}
$$

So

$$
\begin{aligned}
\pi_\theta^T - E_{\sigma^T}\left[X_\theta\right] = \frac{1}{T}&\left[\sum_{\tau=1}^{M-1}\mathbf{P}_{\sigma^T}\left(a_\tau = 1\right)\left(\sum_{t=\tau}^{\tau+M-1}t^{-1} - 1\right)\right. \\
&\left. - \sum_{\tau=T-M+1}^{T}\mathbf{P}_{\sigma^T}\left(a_\tau = 1\right)\left(1 - \frac{T-\tau}{M}\right)\right]
\end{aligned}
$$

Then, it follows directly that $\lim_{T\to\infty}\pi_\theta^T - E_{\sigma^T}\left[X_\theta\right] = 0$. ■

With Lemma 10 in hand, I turn to the proof of Lemma 5. Given the simple strategy, the approximate improvement is given by:

$$
\begin{aligned}
\Delta^T &= \frac{1}{2}\sum_{\theta\in\{0,1\}} v_\theta\left(E_{\sigma^T}\left[X_\theta\right]\right)\left[\varepsilon + (1-2\varepsilon)\left[\pi_\theta^T[1 - G_\theta(\underline{k}^T)] + (1-\pi_\theta^T)[1 - G_\theta(\bar{k}^T)]\right]\right. \\
&\qquad\qquad \left. - E_{\sigma^T}\left[X_\theta\right]\right] \\
&= \frac{1}{2}\sum_{\theta\in\{0,1\}} v_\theta\left(E_{\sigma^T}\left[X_\theta\right]\right)\left[\pi_\theta^T - E_{\sigma^T}\left[X_\theta\right] + (1-2\pi_\theta^T)\varepsilon\right]
\end{aligned}
$$

$$+\frac{1-2\varepsilon}{2}\sum_{\theta\in\{0,1\}} v_\theta\left(E_{\sigma^T}\left[X_\theta\right]\right)\left[-\pi_\theta^T G_\theta\left(\underline{k}^T\right)+\left(1-\pi_\theta^T\right)\left[1-G_\theta\left(\overline{k}^T\right)\right]\right]$$

$$=\frac{1}{2}\sum_{\theta\in\{0,1\}} v_\theta\left(E_{\sigma^T}\left[X_\theta\right]\right)\left[\pi_\theta^T-E_{\sigma^T}\left[X_\theta\right]+(1-2\pi_\theta^T)\varepsilon\right]+\frac{1-2\varepsilon}{2}\Bigg[$$

$$\left(-v_0\left(E_{\sigma^T}\left[X_0\right]\right)\right)\pi_0^T\left[G_0\left(\underline{k}^T\right)-\frac{v_1\left(E_{\sigma^T}\left[X_1\right]\right)}{-v_0\left(E_{\sigma^T}\left[X_0\right]\right)}\frac{\pi_1^T}{\pi_0^T}G_1\left(\underline{k}^T\right)\right]$$

$$+v_1\left(E_{\sigma^T}\left[X_1\right]\right)\left(1-\pi_1^T\right)\left[\left[1-G_1\left(\overline{k}^T\right)\right]-\frac{-v_0\left(E_{\sigma^T}\left[X_0\right]\right)}{v_1\left(E_{\sigma^T}\left[X_1\right]\right)}\frac{1-\pi_0^T}{1-\pi_1^T}\left[1-G_0\left(\overline{k}^T\right)\right]\right]\Bigg]$$

$$=\frac{1}{2}\sum_{\theta\in\{0,1\}} v_\theta\left(E_{\sigma^T}\left[X_\theta\right]\right)\left[\pi_\theta^T-E_{\sigma^T}\left[X_\theta\right]\right]+\frac{1-2\varepsilon}{2}\Bigg[$$

$$\left(-v_0\left(E_{\sigma^T}\left[X_0\right]\right)\right)\left[\pi_0^T\left[G_0\left(\underline{k}^T\right)-\left(\underline{k}^T\right)^{-1}G_1\left(\underline{k}^T\right)\right]-\frac{\varepsilon}{1-2\varepsilon}\left(1-2\pi_0^T\right)\right]$$

$$+v_1\left(E_{\sigma^T}\left[X_1\right]\right)\left[\left(1-\pi_1^T\right)\left[\left[1-G_1\left(\overline{k}^T\right)\right]-\overline{k}^T\left[1-G_0\left(\overline{k}^T\right)\right]\right]-\frac{\varepsilon\left(2\pi_1^T-1\right)}{1-2\varepsilon}\right]\Bigg]$$

Let $\underline{k}\equiv\frac{-v_0(x_0)}{v_1(x_1)}\frac{x_0}{x_1}$ and $\overline{k}\equiv\frac{-v_0(x_0)}{v_1(x_1)}\frac{1-x_0}{1-x_1}$. Note that $\lim_{T\to\infty}\underline{k}^T=\underline{k}$ and $\lim_{T\to\infty}\overline{k}^T=\overline{k}$. However, $G_\theta(l)$ may be discontinuous if there are mass points. In spite of this,

$$\lim_{T\to\infty} G_0\left(\underline{k}^T\right)-\left(\underline{k}^T\right)^{-1}G_1\left(\underline{k}^T\right)=G_0\left(\underline{k}\right)-\left(\underline{k}\right)^{-1}G_1\left(\underline{k}\right). \tag{8}$$

To see this, first let $\lim_{l\searrow\underline{k}} G_\theta(l)$ denote the limit when l approaches $\underline{k}$ from the right. Since $G_\theta(l)$ is always right-continuous, then equation (8) holds. Next, let $\lim_{l\nearrow\underline{k}} G_\theta(l)$ denote the limit when l approaches $\underline{k}$ from the left. If $G_\theta(l)$ is left-continuous at $\underline{k}$, then again equation (8) holds. Recall that $l(s)=\frac{d\nu_1}{d\nu_0}(s)$. For $l\in(0,\infty)$, if $G_\theta(l)$ is *not* left-continuous at $\underline{k}$, then $\int_{l(s)=\underline{k}} d\nu_\theta(s)>0$ for both $\theta\in\{0,1\}$. Then,

$$\begin{aligned}
\lim_{l\nearrow\underline{k}} G_0(l)-l^{-1}G_1(l) &= \int_{l(s)<\underline{k}} d\nu_0(s)-\underline{k}^{-1}\int_{l(s)<\underline{k}} d\nu_1(s)\\
&= \int_{l(s)\le\underline{k}} d\nu_0(s)-\int_{l(s)=\underline{k}} d\nu_0(s)\\
&\quad -\underline{k}^{-1}\left(\int_{l(s)\le\underline{k}} d\nu_1(s)-\int_{l(s)=\underline{k}} d\nu_1(s)\right)\\
&= G_0\left(\underline{k}\right)-\left(\underline{k}\right)^{-1}G_1\left(\underline{k}\right)-\int_{l(s)=\underline{k}} d\nu_0(s)+\underline{k}^{-1}\int_{l(s)=\underline{k}} d\nu_1(s)
\end{aligned}$$

$$
\begin{aligned}
&= G_0\left(\underline{k}\right) - \left(\underline{k}\right)^{-1} G_1\left(\underline{k}\right) - \int_{l(s)=\underline{k}} d\nu_0(s) + \underline{k}^{-1} \int_{l(s)=\underline{k}} \frac{d\nu_1(s)}{d\nu_0(s)} d\nu_0(s) \\
&= G_0\left(\underline{k}\right) - \left(\underline{k}\right)^{-1} G_1\left(\underline{k}\right) - \int_{l(s)=\underline{k}} d\nu_0(s) + \underline{k}^{-1} \int_{l(s)=\underline{k}} \underline{k} d\nu_0(s) \\
&= G_0\left(\underline{k}\right) - \left(\underline{k}\right)^{-1} G_1\left(\underline{k}\right)
\end{aligned}
$$

The same argument guarantees that

$$
\lim_{T\to\infty} \left[1 - G_1\left(\bar{k}^T\right)\right] - \bar{k}^T \left[1 - G_0\left(\bar{k}^T\right)\right] = \left[1 - G_1\left(\bar{k}\right)\right] - \bar{k}\left[1 - G_0\left(\bar{k}\right)\right]. \tag{9}
$$

Given equations (8) and (9),

$$
\begin{aligned}
\lim_{T\to\infty} \Delta^T = \frac{-v_0(x_0)}{2} & \left[(1-2\varepsilon)x_0 \left[G_0\left(\underline{k}\right) - \left(\underline{k}\right)^{-1} G_1\left(\underline{k}\right)\right] - \varepsilon\left(1 - 2x_0\right)\right] \\
+ \frac{v_1(x_1)}{2} & \left[(1-2\varepsilon)\left(1 - x_1\right)\left[\left[1 - G_1\left(\bar{k}\right)\right] - \bar{k}\left[1 - G_0\left(\bar{k}\right)\right]\right] - \varepsilon\left(2x_1 - 1\right)\right]
\end{aligned}
$$

So Corollary 2 leads directly to equation (4). ■

A.8 Proof of Proposition 2

I present first the following proposition.

Proposition 3. (Proposition 11 in Monzón and Rapp [2014]). *For all $l \in (\underline{l}, \bar{l})$, $G_\theta(l)$ satisfies:*

$$
l > \frac{G_1(l)}{G_0(l)} \quad \text{and} \quad l < \frac{1 - G_1(l)}{1 - G_0(l)}
$$

Moreover, if $k' \geq k$ then,

$$
\begin{aligned}
\left[1 - G_1(k)\right] - k\left[1 - G_0(k)\right] &\geq \left[1 - G_1\left(k'\right)\right] - k'\left[1 - G_0\left(k'\right)\right] \\
G_0\left(k'\right) - G_1\left(k'\right)\left(k'\right)^{-1} &\geq G_0(k) - G_1(k)(k)^{-1}
\end{aligned}
$$

See Monzón and Rapp [2014] for the proof.

Let $NE_\delta = \left\{x \in [0,1]^2 : d\left(x, NE\right) \leq \delta\right\}$ be the set of all points which are δ–close to elements of NE and let L^ε denote the set of limit points in a game with mistake probability

$\varepsilon > 0$. I show first the following Lemma.

LEMMA 11. LIMIT SET APPROACHES NE. *For any* $\delta > 0$, $\exists\, \tilde{\varepsilon} > 0 : L^{\varepsilon} \subseteq NE_{\delta}\ \forall \varepsilon < \tilde{\varepsilon}$.

Proof. By contradiction. Assume that there exists 1) a sequence of mistake probabilities $\{\varepsilon^n\}_{n=1}^{\infty}$ with $\lim_{n\to\infty} \varepsilon^n = 0$, and 2) an associated sequence $\{x^n\}_{n=1}^{\infty}$ with $x^n \in L^{\varepsilon^n}$ for all n, but 3) $x^n \notin NE_{\delta}$ for all n. Since $x^n \in [0,1]^2$ for all n, this sequence has a convergent subsequence $\{x^{n_m}\}_{m=1}^{\infty}$ with $\lim_{m\to\infty} x^{n_m} = \bar{x}$. If $v_0(\bar{x}_0) = v_1(\bar{x}_1) = 0$, then $\bar{x} \in NE$, so for m large enough, $x^{n_m} \in NE_{\delta}$. Then, it must be the case that $v_{\theta}(\bar{x}_{\theta}) \neq 0$ for some θ.

Assume that $v_1(\bar{x}_1) > 0$. I show next that this requires $\bar{x}_1 = 1$. Pick $\tilde{m}$ large enough so that $v_1(x_1^{n_m}) > 0$ for all $m > \tilde{m}$. For all m with $v_0(x_0^{n_m}) \geq 0$, Lemma 4 implies that $x^{n_m} = (1-\varepsilon^{n_m}, 1-\varepsilon^{n_m})$. So if $v_0(x_0^{n_m}) \geq 0$ infinitely often, then $\bar{x}_1 = 1$. Similarly, for all m with $v_0(x_0^{n_m}) < 0$, by Lemma 5 equation (4) must hold:

$$\begin{aligned} &\frac{-v_0(x_0^{n_m})}{2}\Big[\overbrace{(1-2\varepsilon^{n_m})}^{\to 1}\overbrace{x_0^{n_m}\left[G_0\left(\underline{k}^{n_m}\right) - \left(\underline{k}^{n_m}\right)^{-1} G_1\left(\underline{k}^{n_m}\right)\right]}^{\geq 0} - \overbrace{\varepsilon\,(1-2x_0)}^{\to 0}\Big] \\ &\quad + \frac{v_1(x_1^{n_m})}{2}\Bigg[\underbrace{(1-2\varepsilon^{n_m})}_{\to 1}\underbrace{(1-x_1^{n_m})\left[\left[1-G_1\left(\overline{k}^{n_m}\right)\right] - \overline{k}^{n_m}\left[1-G_0\left(\overline{k}^{n_m}\right)\right]\right]}_{\geq 0} \\ &\quad - \underbrace{\varepsilon^{n_m}\left(2x_1^{n_m}-1\right)}_{\to 0}\Bigg] \leq 0 \end{aligned} \tag{10}$$

Proposition 3 guarantees both that $\left[\left[1-G_1\left(\overline{k}^{n_m}\right)\right] - \overline{k}^{n_m}\left[1-G_0\left(\overline{k}^{n_m}\right)\right]\right] \geq 0$ and that $\left[G_0\left(\underline{k}^{n_m}\right) - \left(\underline{k}^{n_m}\right)^{-1} G_1\left(\underline{k}^{n_m}\right)\right] \geq 0$. Then, as equation (10) shows, when $\varepsilon^{n_m} \to 0$ only non-negative terms may remain. Assume next that $\bar{x}_1 < 1$. Then $\lim_{m\to\infty} v_1(x_1^{n_m})(1 - x_1^{n_m}) = v_1(\bar{x}_1)(1-\bar{x}_1) > 0$. As $\overline{k} = -[v_0(x_0)(1-x_0)]/[v_1(x_1)(1-x_1)]$, this implies that $\lim_{m\to\infty} \overline{k}^{n_m} < \infty$. Since signals are of unbounded strength, then

$$\lim_{m\to\infty}\left[\left[1-G_1\left(\overline{k}^{n_m}\right)\right] - \overline{k}^{n_m}\left[1-G_0\left(\overline{k}^{n_m}\right)\right]\right] > 0.$$

To summarize, whenever $\bar{x}_1 < 1$, equation (10) is not satisfied for small enough ε^{n_m}. This proves that $\bar{x}_1 = 1$.

Analogous arguments (using also Lemma 6) imply that if $v_{\theta}(\bar{x}_{\theta}) > 0$, then $\bar{x}_{\theta} = 1$ and that if $v_\theta(\bar{x}_\theta) < 0$, then $\bar{x}_\theta = 0$. So $\bar{x} \in NE$, and thus I have reached a contradiction. ■

With Lemma 11 the proof of Proposition 2 is straightforward. Fix $\delta/2 > 0$ and let $\tilde{\varepsilon}$ be as given by Lemma 11. Write:

$$
\begin{aligned}
d\left(X, NE\right) = \min_{y \in NE} |X - y| = \min_{y \in NE} |X - l + l - y| &\leq |X - l| + \min_{y \in NE} |l - y| \quad \text{for any } l \\
&\leq d\left(X, L^{\varepsilon}\right) + \min_{y \in NE} |l - y| \quad \text{ for } l \in \arg\min_{l \in L^{\varepsilon}} |X - l| \\
&\leq d\left(X, L^{\varepsilon}\right) + \delta/2 \quad \forall \varepsilon < \tilde{\varepsilon}, \text{ by Lemma 11.}
\end{aligned}
$$

Then, for any σ, $\mathbf{P}_\sigma\left(d\left(X, NE\right) < \delta\right) \geq \mathbf{P}_\sigma\left(d\left(X, L^{\varepsilon}\right) < \delta/2\right)$. By Corollary 1, for all $\delta/2 > 0$, and all sequences of symmetric equilibria:

$$
\lim_{T \to \infty} \mathbf{P}_{\sigma^{T,*}}\left(d\left(X, NE\right) < \delta\right) \geq \lim_{T \to \infty} \mathbf{P}_{\sigma^{T,*}}\left(d\left(X, L^{\varepsilon}\right) < \delta/2\right) = 1 \; ■
$$