The Social Learning Barrier

Florian Brandl†
University of Bonn

We consider long-lived agents who interact repeatedly in a social network. In each period, each agent learns about an unknown state by observing a private signal and her neighbors' actions in the previous period before taking an action herself. Our main result shows that the learning rate of the slowest learning agent is bounded from above independently of the number of agents, the network structure, and the agents' strategies. Applying this result to equilibrium learning with rational agents shows that the learning rate of all agents in any equilibrium is bounded under general conditions. This extends recent findings on equilibrium learning and demonstrates that the limitation stems from an inherent tradeoff between optimal action choices and information revelation rather than strategic considerations.

1. Introduction

How fast do individuals learn from repeatedly observing each other's actions in social networks? The amount of private information in large networks is vast, so efficient information aggregation would lead to rapid learning. However, we show that information aggregation fails under general conditions: the rate of learning of the slowest learning agent is bounded from above by a constant that does not depend on the size and structure of the network and the agents' behavior. This has various direct consequences for learning in equilibrium by rational and forward-looking agents and applications to domains such as product choice, voting, technology adoption, and opinion formation.

In our model, long-lived agents interact with each other over an infinite number of periods in a network. The state of the world is fixed but unknown. In each period, each agent receives a private signal about the state and observes the actions of her neighbors in the previous period before choosing an action herself. The signals are independent and identically distributed across periods conditional on the state. An agent's flow utility in a period depends on her action and the state, but not the other agents' actions, and is unobserved. We assume all agents share the

†florian.brandl@uni-bonn.de

same generic flow utility function and quantify the learning rate by how fast the probability of agents choosing suboptimal actions vanishes.

Our main result (Theorem 1) shows that information aggregation fails under general conditions: for any number of agents, any network structure, and any strategies for the agents, some agent learns no faster than a fixed upper bound. This bound only depends on the marginal distributions of the agents' private signals and not on correlations between them. When combined with an imitation argument of Huang, Strack, and Tamuz (2024), this strengthens recent results on the learning rate in equilibrium with rational agents who are either myopic or geometrically discount future payoffs of Harel, Mossel, Strack, and Tamuz (2021) and Huang et al. (2024).[1] Hence, our main insight is that bounds on the learning rates arise not from equilibrium strategy restrictions, but from an inherent trade-off: agents choosing the payoff-maximizing action vs. using their actions to inform others.

In networks with many agents, the total amount of private information is vast. If information sharing were fully efficient, all agents would learn very fast. However, our result shows that learning remains bounded, independent of network size, implying that as n grows almost all private information is in fact lost. To see what causes the breakdown of information aggregation, take the perspective of a social planner who can design the agents' strategies to maximize the rate at which the probability of suboptimal action choices vanishes. The strategies must trade off two competing objectives: on the one hand, with high probability, all agents must choose the same action in most periods; on the other hand, an agent's action choice must contain information about her private signals. The second objective requires that an agent's action depends on her private signals sufficiently often, which conflicts with the first objective. In more detail, in order for an agent to make mistakes with rapidly vanishing probability in state A, she needs to choose the corresponding optimal action a if all of her neighbors do so, even if her private signals indicate another state. But if all agents are reluctant to switch away from a in that situation, then, with positive probability, in state B all agents will eventually keep choosing a forever. There a is suboptimal, so learning breaks down entirely. This failure mode is reminiscent of the information cascades in the herding model of Bikhchandani, Hirshleifer, and Welch (1992) and the "rational groupthink" event of Harel et al. (2021), which drives the bound on learning with myopic agents.

The obtained bound on the learning rate is tight: we show that a social planner can design strategies for which each agent's learning rate gets arbitrarily close to the upper bound if there are sufficiently many agents and the network is strongly connected (Theorem 2).[2] Here, we assume that the signals are conditionally independent and identically distributed across agents and not only across periods. This shows that the above tradeoff is not absolute: agents can match the correct action more frequently than an isolated agent and have their actions be

[1] See Huang et al. (2024, Lemma 2) for the imitation argument with geometrically discounting agents. Similar imitation principles are common in the literature (see, e.g., Smith and Sorensen, 2000; Gale and Kariv, 2003; Golub and Sadler, 2017).

[2] A network is strongly connected if there is an observational path from any agent to any other agent.

informative about their private signals at the same time. For complete networks, the strategies we use are simple. Each agent follows her past private signals if those are highly indicative of some state, which ensures that these actions are very likely correct. Whenever an agent's private signals do not decisively favor one of the states, she follows the most popular action of the previous period. We use results from large deviations theory to show that most agents' signals strongly indicate the true state in most periods so that the most popular action in any period is very likely correct. In either case, an agent's action is very likely correct.

We then turn to learning in equilibrium by rational agents who geometrically discount future payoffs. The imitation argument of Huang et al. (2024) shows that all agents learn at the same rate in any equilibrium and for any number of agents in a strongly connected network. Hence, any bound on the learning rate of the slowest learning agent gives the same bound for every agent in that case. This allows us to recover the results of Harel et al. (2021) and Huang et al. (2024)—showing that all agents learn at a bounded rate in any equilibrium independently of the number of agents in any strongly connected network—under more general conditions. Most notably, we do not assume that the signals are conditionally independent across agents. Since our bound on the learning rate of the slowest learning agent applies to any strategies, it holds for equilibrium learning independently of agents' evaluation of future payoffs, for misspecified but otherwise rational agents, and for agents who use non-Bayesian heuristics. Moreover, whenever an imitation argument is available, the bound extends to all agents (not just the slowest one).

Finally, we demonstrate that the failure of information aggregation persists in a richer information environment by considering a model variation where agents observe their neighbors' signals in addition to their actions. In complete networks, all information is then public and learning is very fast in a large network with rational agents. However, a common feature of social networks is that each agent only has a small number of neighbors. As a corollary of our main result, we show that the size of the largest neighborhood determines the equilibrium learning rate. In more detail, we assume that the signals are conditionally independent across agents and periods and agents are rational and geometrically discount future payoffs. Then, for any number of agents and any strongly connected network, no agent's learning rate in any equilibrium exceeds an upper bound that only depends on the size of the largest neighborhood in the network and the marginal distributions of private signals.

The rest of the paper is structured as follows. Section 2 discusses related work and Section 3 introduces the model. In Section 4, we recall known results on learning by an isolated agent and with publicly observable signals. Section 5 states the main result, explains the ideas underlying its proof, and shows that the established bound on the learning rate is tight. Section 6 applies this result to learning in equilibrium with geometrically discounting agents and considers the model variation where agents observe their neighbors' signals and actions. Section 7 concludes with a discussion of modeling assumptions and future directions. All proofs are in the Appendix.

2. Related Work

Most of the literature has focused on equilibrium learning, non-Bayesian agents, non-recurring private signals, or short-lived agents.

Studying models with multiple periods and long-lived rational agents is challenging. Agents may choose suboptimal actions today to induce other agents to reveal information tomorrow, which requires analyzing higher-order beliefs. In recent work, Huang et al. (2024) show that in essentially the same model as in the present paper, the rate of learning of all agents in any equilibrium is bounded independently of the number of agents and the structure of the network. This result follows from the following elegant argument: if the information contained in the agents' actions is too precise, agents will ignore their private signals, so that actions will cease to reveal information; thus, actions can only contain a bounded amount of information, which implies that learning is bounded. Harel et al. (2021) obtained a similar conclusion in a restricted setting with myopic agents and networks in which each agent observes all other agents' actions. They derive their result using large deviations theory, and their work is methodologically closer to ours. Both papers rely on agents being fully rational, playing equilibrium strategies, and geometrically discounting future payoffs—assumptions that our results show are not necessary.

Because of the difficulties arising from Bayesian learning with repeated interactions, the literature has focused on learning heuristics and non-Bayesian agents. The literature following DeGroot (1974) assumes that agents observe each other's beliefs and form tomorrow's belief via a simple updating heuristic such as linear aggregation (see, e.g., Golub and Jackson, 2010). Another approach is to relax the assumptions that agents are fully Bayesian. For example, Bala and Goyal (1998) assume agents respond rationally to private signals and others' random payoffs but ignore the informational content of others' actions. Meanwhile, the agents in Molavi, Tahbaz-Salehi, and Jadbabaie (2018) use a heuristic: they combine past beliefs and then update that aggregate belief rationally based on their private signals.

Another strand of the literature starting with Geanakoplos and Polemarchakis (1982), Bacharach (1985), and Parikh and Krasucki (1990) considers models in which rational and long-lived agents receive a private signal once before the first period and repeatedly observe the actions of other agents, and studies whether agents converge on the same action. Gale and Kariv (2003) allow for social networks in which agents observe their neighbors' actions and show that eventually, all agents converge on the same action. In the same model, Mossel, Sly, and Tamuz (2014, 2015) study the probability that agents converge on the correct action as the number of agents goes to infinity and show that this depends on the network structure. Vives (1993) considers a continuum of agents with a continuous action space and shows that information aggregation can still be slow if observations of actions are noisy as in the case of observing market prices. In contrast to our model, agents do not receive private signals in later periods.

In the classical herding model (Bikhchandani, Hirshleifer, and Welch, 1992; Banerjee, 1992; Smith and Sorensen, 2000), a single short-lived agent arrives in each period and observes a private signal as well as her predecessors' actions. Learning can fail in this setting since rational agents may ignore their private signals and follow their predecessors' actions, leading to herding on the wrong action. The analysis of this model is substantially different from the present model. Since each agent acts only once, informational feedback loops need not be considered. Arieli, Babichenko, Müller, Pourbabee, and Tamuz (2025) consider a variation of the herding model, in which agents are condescending by underestimating the quality of the others' private information. This misspecification decreases the probability of herding on the wrong action and improves learning compared to correct specification if condescension is mild. Harel et al. (2021) conjecture that the same misspecification improves learning in the present model as well. The strategies we use to show that learning in networks can be faster than in autarky can be seen as a mixture of extreme condescension and extreme anti-condescension since the agents completely ignore others' actions most of the time and otherwise ignore their private signals. In a variant of the herding model, the state changes stochastically over time (Moscarini, Ottaviani, and Smith, 1998; Lévy, Pęski, and Vieille, 2024; Huang, 2024). We maintain the assumption that the state is persistent throughout. A recent survey of Bikhchandani, Hirshleifer, Tamuz, and Welch (2021) summarizes the work on models with short-lived agents.

The bandit literature considers models in which rational agents learn in repeated interactions from observing each other's actions and payoffs (Bolton and Harris, 1999; Keller, Rady, and Cripps, 2005; Keller and Rady, 2010; Heidhues, Rady, and Strack, 2015). The main differences to our model are that in the bandit problem, the agents have an experimentation motive and all information is public. This induces a free-rider problem that has no analog in our model.

3. The Model

Let $N = \{1, \ldots, n\}$ be the set of agents and let $T = \{1, 2, \ldots\}$ be the set of periods. Each agent has the same possibly infinite set of actions A and chooses an action in each period. If x is a vector indexed by N and $i \in N$, then x^i denotes its i-th coordinate, and if x is indexed by T and $t \in T$, then x_t is its period-t coordinate, $x_{\leq t}$ is its restriction to the periods $\{1, \ldots, t\}$, and $x_{<t}$ is its restriction to the periods $\{1, \ldots, t-1\}$. The set of states of the world is Ω, which is assumed to be finite. The true state $\omega \in \Omega$ is a random variable with full support distribution $\pi_0 \in \Delta(\Omega)$. We assume that ω and all other random variables defined below live on a probability space with probability measure P. For a state θ, we write $\mathrm{E}_\theta(\cdot) = \mathrm{E}(\cdot \mid \theta)$ and $P_\theta(\cdot) = P(\cdot \mid \theta)$ for the corresponding conditional expectation and conditional probability.

3.1. Agents' Payoffs

All agents have the same utility function $u \colon A \times \Omega \to \mathbb{R}$ that depends on their own action and the state, and $u(a, \omega)$ is an agent's flow utility for choosing the action a in any period. An agent's utility is independent of other agents' actions so the interactions between the agents are purely informational. We assume that the optimal action a_θ for every state $\theta \in \Omega$ exists and is unique.

$$\{a_\theta\} = \arg\max_{a \in A} u(a, \theta)$$

We also assume that $a_\theta \neq a_{\theta'}$ for any two distinct states $\theta, \theta' \in \Omega$.[3] We say that a_ω is the correct action and any other action is a mistake. The requirement that no action is optimal in two different states avoids trivial cases, and the uniqueness of the optimal action for each state prevents an agent from communicating additional information through the choice of the optimal action without making a mistake. This tradeoff between choosing the correct action and communicating information about private signals is the main tension in our model. We explain this in detail in Remark 1. Since we quantify learning by the probability of choosing the correct action, the sole role of utility functions is to identify the correct action. More fine-grained characteristics of the utility functions are only relevant for analyzing equilibria with geometrically discounting agents (cf. Section 6).

3.2. Agents' Information

The prior distribution π_0 is commonly known. In each period $t \in T$, each agent i privately observes a signal $\mathfrak{s}_t^i$ from a set of signals S, which is assumed to be a standard Borel space. Conditional on each state θ, $\mathfrak{s}_t^i$ has distribution $\mu_\theta^i \in \Delta(S)$ and $\mathfrak{s}_t = (\mathfrak{s}_t^i)_{i \in N}$ has distribution $\mu_\theta \in \Delta(S^N)$, and $(\mathfrak{s}_t)_{t \in T}$ are independent conditional on ω. That is, signals are conditionally independent across periods but not necessarily across agents. We assume that $\mu_\theta, \mu_{\theta'}$ are mutually absolutely continuous for any two states θ, θ', so that no signal excludes any state with certainty and signals are informative about any pair of states. Observing the signal realization $s \in S$ changes the log-likelihood ratio of the observing agent i between the states θ and θ' by

$$\ell^i_{\theta,\theta'}(s) = \log \frac{d\mu^i_\theta}{d\mu^i_{\theta'}}(s)$$

We assume that $\ell^i_{\theta,\theta'}(s)$ is bounded and not identically 0 for any two distinct states, so that each agent's signals have bounded but non-trivial informativeness about any pair of distinct states. Let $\ell^i_{\theta,\theta'} = \ell^i_{\theta,\theta'}(\mathfrak{s}^i_1)$ for any pair of states θ, θ' and any agent i. The private signals of agent i up to any period t induce the private log-likelihood ratio

$$L^i_{\theta,\theta',t} = \log \frac{\pi_0(\theta)}{\pi_0(\theta')} + \sum_{r \leq t} \ell^i_{\theta,\theta'}(\mathfrak{s}^i_r)$$

Each agent i observes the actions of her neighbors $N^i \subset N$, and we assume $i \in N^i$ so that each agent observes her own action. The directed graph induced by these neighborhoods is called the network, and we assume that it is common knowledge among the agents. A network

[3] The assumption that all agents have the same utility function is purely for notational convenience. All proofs remain valid with the obvious adjustments provided each agent's utility function satisfies the preceding genericity assumptions and the agents know each other's utility functions (or at least each other's optimal actions in each state).

is strongly connected if there is an observational path from any agent to any other agent, and complete if each agent's neighborhood is N.

Agents do not observe each other's signals. Moreover, agents know their utility function (and thus everyone's utility function) but do not observe the flow utility $u(a, \omega)$ of any agent, including themselves. The latter assumption shuts down any experimentation motives and is common for models of learning without experimentation. Our formulation includes a model in which agents receive noisy signals about their flow utility today through tomorrow's signal.[4]

Thus, the information available to agent i in period t before choosing an action consists of the actions of all of i's neighbors in all previous periods and i's signals in all periods up to and including t. We say that $A^{|N^i|\times(t-1)}$ is the set of public histories of agent i, S^t is the set of private histories for each agent, and $\mathcal{I}^i_{\leq t} = S^t \times A^{|N^i|\times(t-1)}$ is the collection of information sets of agent i before choosing an action in period t.

3.3. Agents' Strategies

A pure strategy for agent i is a sequence $\sigma^i = (\sigma^i_t)_{t\in T}$ of measurable functions $\sigma^i_t \colon \mathcal{I}^i_{\leq t} \to A$ from i's information sets in period t to the set of actions, and a pure strategy profile $\sigma = (\sigma^1, \ldots, \sigma^n)$ consists of a strategy for each agent. A pure strategy profile σ induces a random sequence of action profiles: for each i, $a^i_1 = \sigma^i_1(\mathfrak{s}^i_1)$, and for each $t > 1$, $a^i_t = \sigma^i_t(\mathfrak{s}^i_{\leq t}; (a^j_{<t})_{j\in N^i})$, where i's public history $(a^j_{<t})_{j\in N^i}$ is given by the actions of her neighbors in the periods preceding t. For $i \in N$, $t \in T$, and $H_{\leq t} \in A^{N\times t}$, we denote by $\mathcal{S}^i(H_{\leq t}) = \{s^i_{\leq t} \in S^t \colon \forall r \leq t, \sigma^i_r(s^i_{\leq r}; (H^j_{<r})_{j\in N^i}) = H^i_r\}$ the set of those trajectories of i's signals consistent with $H_{\leq t}$, and write $\mathcal{S}(H_{\leq t}) = \prod_{i\in N} \mathcal{S}^i(H_{\leq t})$ for the trajectories of signal profiles consistent with $H_{\leq t}$. Thus, play follows $H_{\leq t}$ if and only if each agent i receives signals in $\mathcal{S}^i(H_{\leq t})$.

We say that agent i makes a mistake in period t if $a^i_t \neq a_\omega$, and that i learns at rate r if

$$r = \liminf_{t\to\infty} -\frac{1}{t} \log P\left(a^i_t \neq a_\omega\right)$$

If the limit exists, the probability of a mistake in period t is $e^{-rt+o(t)}$.[5] This definition of the learning rate is common in the literature (see, e.g., Vives, 1993; Hann-Caruthers et al., 2018; Molavi et al., 2018; Rosenberg and Vieille, 2019; Harel et al., 2021; Huang et al., 2024).[6] Any

[4]Formally, consider the case that agent i's flow utility for action a in period t is $\tilde{u}(a, \mathfrak{s}^i_{t+1})$ for an action and signal-dependent utility function $\tilde{u} \colon A \times S \to \mathbb{R}$. Agents thus observe their flow utility in period t through their signal in the next period. If we define $u(a, \theta) = \mathrm{E}_\theta\left(\tilde{u}(a, \mathfrak{s}^i_1)\right)$, then both models are equivalent in terms of expected payoffs at the time of choosing an action. This connection has also been noted by Rosenberg, Solan, and Vieille (2009), Harel et al. (2021), and Huang et al. (2024).

[5]We use the asymptotic notation $o(t)$ for a function that grows slower than t as t goes to infinity. That is, $f(t) \in o(t)$ if $\lim_{t\to\infty} \frac{f(t)}{t} = 0$.

[6]An alternative, more qualitative asymptotic definition of the learning rate considers the rate at which the expected difference between the utility of the correct action and an agent's action goes to 0, i.e.,

$$\liminf_{t\to\infty} -\frac{1}{t} \log \mathrm{E}\left(u(a_\omega, \omega) - u(a^i_t, \omega)\right)$$

For finite action sets, both definitions coincide. For infinite action sets and assuming utilities are bounded, the learning rate is weakly higher for the quantitative definition and may be arbitrarily high even when the bound on the rate of learning for pure strategies entails the same bound for mixed strategies for the same instance with a larger signal space and an additional signal component that is uninformative about the state.[7] Hence, restricting to pure strategies comes at no loss in generality, and we do so throughout.

3.4. Leading Example

The following simple instance of the model already presents most of the arising complexities and can serve as a leading example. There are two states and two actions, say, $\Omega = \{f, g\}$ and $A = \{a_f, a_g\}$. Each agent has utility 1 for matching the state and 0 for failing to match the state. The network is complete. The signals are binary ($S = \{s_f, s_g\}$) and conditionally independent and identically distributed across agents and periods, and each agent's signal in each period matches the state with probability p for some $p \in (\frac{1}{2}, 1)$ (i.e., $\mu^i_f(s_f) = p$ and $\mu^i_g(s_g) = p$). We illustrate our results using this example at the end of Section 5.

4. Autarky and Public Signals

We revisit two settings as benchmarks: a single agent learning in autarky and several agents observing each other's signals.

If there is only a single agent and this agent i chooses actions optimally based on her private signals, it follows from classical large deviations results for random walks that for some $r^i_{\text{aut}} > 0$ determined by the distribution of signals,

$$P\left(a^i_t \neq a_\omega\right) = e^{-r^i_{\text{aut}} t + o(t)}$$

Hence, the limit $\lim_{t\to\infty} -\frac{1}{t} \log P\left(a^i_t \neq a_\omega\right)$ exists and is equal to r^i_{aut}, which we call i's autarky learning rate. In particular, the probability of a mistake goes to zero as time goes to infinity. For a proof of this result, see Dembo and Zeitouni (2009, Theorem 2.2.30) or Harel et al. (2021, Fact 1) in the present context for the case of two states. We provide more details in Appendix B.

Now consider any number of agents with public signals: each agent observes all other agents' signals. If the agents' signals are conditionally independent and identically distributed across

qualitative learning rate is 0. Hence, the qualitative definition of the learning rate is most reasonable for finitely many actions.

[7] More precisely, replace the signal space S by $\tilde{S} = S \times [0,1]$, and for each state θ, let $\tilde{\mu}_\theta$ be the product distribution on $\tilde{S}^N$ with marginal μ_θ with respect to S^N and the uniform distribution on $[0,1]^N$ as its marginal with respect to $[0,1]^N$. We may choose the signals' second coordinates so that they are independent of the state and any other signals. Then, the informativeness of each agent's signals remains unchanged and the signal profile distributions remain mutually absolutely continuous and conditionally independent across periods. The second coordinate of a signal in $\tilde{S}$ can be used to map any mixed strategy σ for the instance with signals in S to a pure strategy $\tilde{\sigma}$ for signals in $\tilde{S}$ that is behaviorally equivalent, i.e., conditional on each state, the induced distributions of sequences of action profiles (i.e., distributions on $A^{n\times T}$) are the same for σ and $\tilde{\sigma}$. Likewise, for each pure strategy profile with signal space $\tilde{S}$, there is a behaviorally equivalent mixed strategy profile with signal space S.

agents and periods, then the signals of n agents in a single period are as informative as those of a single agent over n periods. Hence n agents observing each other's signals and choosing actions optimally learn n times as fast as a single agent in autarky and the rate of learning is nr_{aut}, where $r_{\text{aut}} = r^i_{\text{aut}}$ does not depend on i.

$$P\left(a^i_t \neq a_\omega\right) = e^{-nr_{\text{aut}}t+o(t)}$$

In particular, the rate of learning grows linearly in the number of agents and can thus become arbitrarily large provided there are sufficiently many agents. By contrast, when the signals are not conditionally independent across agents, even two agents can learn much faster than a single agent with the same distribution of private signals. For our leading example (cf. Section 3.4), if the probability that both agents simultaneously receive the wrong signal is sufficiently low but positive, then an observer of the joint signals can achieve an arbitrarily high learning rate.

5. Coordinated Learning

We study how fast agents can learn from observing their private signals and their neighbors' actions. First, we take the perspective of a social planner who can design the agents' strategies and aims to maximize the learning rate of the slowest learning agent. In particular, we ask whether one can design strategies such that, in a sufficiently large and strongly-connected network, every agent's learning rate can exceed any fixed bound.

Our first result shows that there is an upper bound on the learning rate of the slowest learning agent that is independent of the number of agents, the network structure, and the strategy profile imposed by the social planner. Hence, information aggregation breaks down not because of equilibrium constraints on strategies, but because of a fundamental tradeoff between choosing the correct action and using actions to communicate information. More precisely, each agent's strategy needs to trade off choosing the action that is most likely to be correct in the current period and using actions to inform other agents about the agent's private signals to reduce others' probability of mistakes in future periods. The fact that learning is bounded shows that there is no way to achieve both objectives simultaneously.

Theorem 1 (Learning is bounded)**.** *For any number of agents n, any network, and any strategies $\sigma^1, \ldots, \sigma^n$, some agent learns at rate at most $r_{\text{bdd}} = \min_{\theta\neq\theta'} \max_{i\in N} \mathrm{E}_\theta \left(\ell^i_{\theta,\theta'}\right)$.*

In other words, for any strategy profile, there is some agent i such that for each $\epsilon > 0$, we have $P\left(a^i_t \neq a_\omega\right) \geq e^{-(r_{\text{bdd}}+\epsilon)t}$ for infinitely many periods t. So fast learning by some agents always comes at the cost of other agents' learning. It is clear from the expression for r_{bdd} that it only depends on the marginal signal distributions μ^i_θ and is determined by the two states that are hardest to distinguish. Theorem 1 only asserts that some agent's learning rate meets the prescribed bound, and it is easy to construct strategies for which a large fraction of agents learn very fast if there are many agents.[8]

[8] Consider the instance in Section 3.4 with the following strategies: in each period, each odd-numbered agent

While it is a priori harder to prove a result for all strategy profiles rather than for, say, equilibria with rational agents, the greater generality makes clear that the argument cannot rely on analyzing belief dynamics or incentives. Moreover, it is clearly without loss to consider only complete networks, which is not obvious for equilibrium learning. The proof of Theorem 1 greatly simplifies when additionally assuming that the signals are conditionally independent across agents and loosening the bound on the learning rate. We record this corollary to Theorem 1 below and give an independent short proof in the Appendix that illustrates the main tension in the model.

Theorem 1' (Learning is bounded, weak form)**.** Assume the signals are conditionally independent across agents and periods. For any number of agents n, any network, and any strategies $\sigma^1, \ldots, \sigma^n$, some agent learns at rate at most $\tilde{r}_{\mathrm{bdd}} = 2 \min_{\theta,\theta'} \max_{i \in N} \sup_{s \in S} |\ell^i_{\theta,\theta'}(s)|$.

We sketch the proof of Theorem 1' for the case of two states f and g below. The intuition is as follows. If all agents learn faster than the prescribed rate, each agent chooses a_g except for a set of signal trajectories with very small probability in state g in any sufficiently late period. But then in state f, these sets still have moderately small probability and so with positive probability, all agents choose a_g in each sufficiently late period, contradicting that agents learn at all.

Proof sketch of Theorem 1'. In more detail, assume for contradiction that all agents learn at a rate of at least $\tilde{r}_{\mathrm{bdd}} + 2\epsilon$ for some positive ϵ. A preliminary lemma turns the limit defining the learning rate into a statement about each sufficiently late period at a small cost in the learning rate. More precisely, Lemma 4 exhibits a history $H_{\leq t_0}$ with positive probability in state g (and thus also in state f) such that

$$P_g\left(a^i_t \neq a_g \mid H_{\leq t_0}\right) \leq e^{-(\tilde{r}_{\mathrm{bdd}}+\epsilon)t}$$

for each agent i and each period $t > t_0$, and t_0 may be chosen arbitrarily large. Denoting by H the history that follows $H_{\leq t_0}$ up to t_0 and for which each agent chooses a_g in each period after t_0, it is easy to see that $P_g\left(H \mid H_{\leq t_0}\right) \geq \frac{1}{2}$ if t_0 is large enough. We say that agent i defects in period $t > t_0$ if i is the first to deviate from H, i.e., if play follows $H_{<t}$ in the first $t-1$ periods and $a^i_t \neq a_g$. Consider now the set D^i_t of those trajectories of i's signals for which i defects in period t. We aim to bound the probability of each D^i_t conditional on f and $H_{\leq t_0}$.

If i receives signals in D^i_t and no other agent defects before t, then $a^i_t \neq a_g$, and so

$$P_g\left(D^i_t, \forall j \neq i, H^j_{<t} \mid H_{\leq t_0}\right) \leq P_g\left(a^i_t \neq a_g \mid H_{\leq t_0}\right) \leq e^{-(\tilde{r}_{\mathrm{bdd}}+\epsilon)t}$$

On the other hand, since signals are conditionally independent across agents and periods, the joint probability of these two events equals the product of their probabilities. But

$$P_g\left(\forall j \neq i, H^j_{<t} \mid H_{\leq t_0}\right) \leq P_g\left(H \mid H_{\leq t_0}\right) \geq \frac{1}{2}$$

chooses a_θ if her private signal is s_θ, and each even-numbered agent chooses optimally based on the actions of the odd-numbered agents. Then, the odd-numbered agents do not learn at all, and each even-numbered agent learns at rate $\frac{n}{2} r_{\mathrm{aut}}$, which grows linearly with the number of agents.

and so $P_g\left(D_t^i \mid H_{\le t_0}\right) \le 2e^{-(\tilde{r}_{\text{bdd}}+\epsilon)t}$. Since signals are conditionally independent across agents, we have $P_g\left(D_t^i \mid H_{\le t_0}\right) = P_g\left(D_t^i \mid \prod_{j\in N}\mathcal{S}^j(H_{\le t_0})\right) = P_g\left(D_t^i \mid \mathcal{S}^i(H_{\le t_0})\right)$ and similarly for f instead of g, where $\mathcal{S}^j(H_{\le t_0})$ denotes those trajectories of j's signals that are consistent with $H_{\le t_0}$. Hence, from the definition of $\tilde{r}_{\text{bdd}}$ and the conditional independence of signals across periods, we now see that

$$P_f\left(D_t^i \mid H_{\le t_0}\right) = P_f\left(D_t^i \mid \mathcal{S}^i(H_{\le t_0})\right) \le e^{\tilde{r}_{\text{bdd}}t} P_g\left(D_t^i \mid \mathcal{S}^i(H_{\le t_0})\right) = e^{\tilde{r}_{\text{bdd}}t} P_g\left(D_t^i \mid H_{\le t_0}\right) \le 2e^{-\epsilon t}$$

since the probability of D_t^i increases by a factor of at most $e^{\frac{1}{2}\tilde{r}_{\text{bdd}}t}$ and the probability of $\mathcal{S}^i(H_{\le t_0})$ decreases by a factor of at most $e^{\frac{1}{2}\tilde{r}_{\text{bdd}}t_0}$ when conditioning on f rather than g. If t_0 is large enough, the probability that any agent defects following $H_{\le t_0}$ is thus at most, say, $\frac{1}{2}$ even in state f. But then, with positive probability, all agents choose the incorrect action a_g in all periods after t_0 in state f, which contradicts that their learning rate is positive. □

The second result shows that the bound in Theorem 1 is tight under additional assumptions. Thus, observational learning in networks can improve the learning rate compared to a single agent learning in autarky for suitably designed strategies.

Theorem 2 (Coordination improves learning)**.** *Assume the signals are conditionally independent and identically distributed across agents and periods. For any $\epsilon > 0$, there is n_0 such that for all $n \ge n_0$ and any strongly connected network, there exist strategies $\sigma^1, \ldots, \sigma^n$ such that each agent learns at rate at least $r_{\text{bdd}} - \epsilon$.*

In other words, for any positive ϵ, there exist strategies so that $P(a_t^i \neq a_\omega) \le e^{-(r_{\text{bdd}}-\epsilon)t+o(t)}$ for each agent i if the number of agents is large enough and the network is strongly connected. This shows that the bound in Theorem 1 is tight. The assumption that the signals are conditionally independent across agents is needed. Indeed, if the agents' signals were perfectly correlated conditional on the state, no agent could exceed the autarky learning rate since even observing other agents' signals would not give additional information.

For complete networks, the strategies we construct are easy to describe: each agent follows her private signals as long as those are sufficiently decisive, and otherwise she follows the action taken by most agents in the previous period. To make this more precise, consider again the case of two states f and g. In state $\theta \in \{f, g\}$, agent i's log-likelihood ratio for θ over θ' in period t is expected to be roughly $\mathrm{E}_\theta\left(\ell_{\theta,\theta'}\right)t$, and as long as it is larger than $(\mathrm{E}_\theta\left(\ell_{\theta,\theta'}\right) - \epsilon)t$, agent i chooses a_θ independently of the other agents' actions. Otherwise, she chooses the action that most agents chose in period $t-1$.

These strategies lead to faster learning than learning in autarky. First, agents decide based on their private signals only if those clearly favor one state, and so in that case, mistakes are less likely than when always relying on one's private signals. Second, most agents decide based on their private signals for late periods since then each agent's likelihood ratio is close to its expectation with high probability, and each of these actions is correct with high probability independently of the others. Hence, the most popular action in any period is very likely to

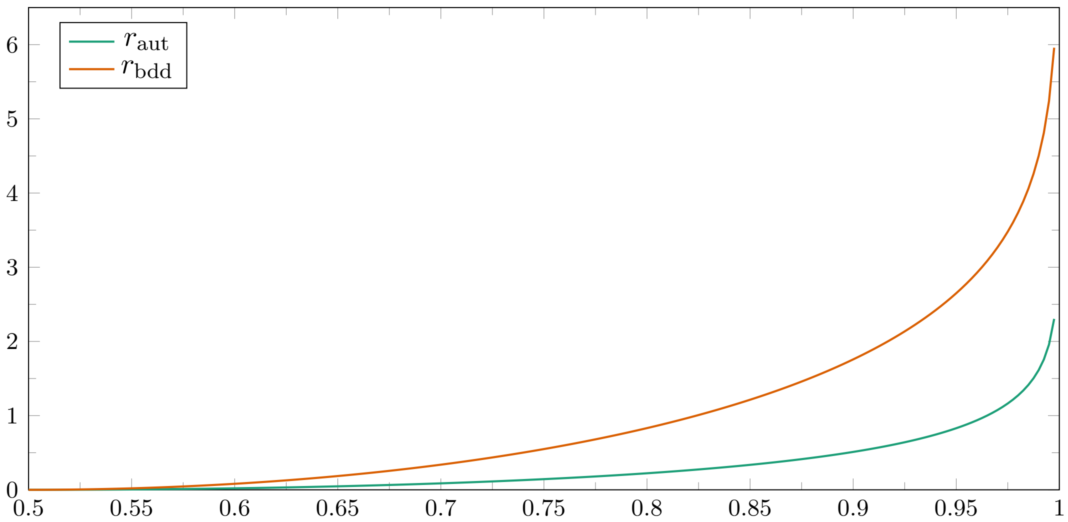

Figure 1: Bounds on the learning rates for the example in Section 3.4. The horizontal axis corresponds to the parameter p determining the informativeness of the signal distributions. Higher values of p correspond to more informative signals. The vertical axis indicates the rate of learning: r_{aut} is the learning rate of a single agent learning in autarky obtained from Proposition 1 in Appendix B; r_{bdd} is the upper bound on the learning rate of the slowest learning agent in an arbitrary network and with arbitrary strategies obtained from Theorem 1; and for any strongly connected network with a large number of agents with the strategies described after Theorem 2 (and in more detail in its proof), each agent learns at a rate close to r_{bdd}.

be correct if there are many agents. So either case improves upon learning in autarky. Note, however, that each agent could unilaterally achieve a higher rate than r_{bdd} by following the most popular action in each period. In particular, these strategies are not in equilibrium for rational and geometrically discounting agents.

We illustrate the bounds from Theorem 1 and Theorem 2 for the example in Section 3.4. Recall that there are two states f and g and two signals, and each agent's signal in each period matches the state with probability $p \in (\frac{1}{2}, 1)$. A calculation shows that

$$\mathrm{E}_g(\ell_{g,f}) = \mathrm{E}_f(\ell_{f,g}) = (2p-1)\log\frac{p}{1-p}$$

Hence, $r_{\text{bdd}} = (2p-1)\log\frac{p}{1-p}$. The expression for r_{aut} involves a minimization problem and cannot be stated in closed form. For $p = 0.75$, we have numerically that $r_{\text{aut}} \approx 0.144$ and $r_{\text{bdd}} \approx 0.549$. Hence, for these signal distributions, observing conditionally independent signals of four agents allows for faster learning than the slowest-learning agent in any network for any number of agents with any strategies. Figure 1 illustrates the bounds on the learning rates for other values of p.

6. Equilibrium Learning

We turn to learning in equilibrium with geometrically discounting agents. More precisely, suppose that all agents share a common discount rate $\delta \in [0, 1)$.[9] The expected utility of agent i for a strategy profile σ is

$$u^i(\sigma) = \sum_{t \in T} \delta^{t-1} \mathrm{E}\left(u(a_t^i, \omega)\right)$$

where a_t^i is i's action in period t for the strategy profile σ. The case $\delta = 0$ corresponds to myopic agents. A strategy profile is a Nash equilibrium if no agent can increase her expected utility by unilaterally changing her strategy. Since mixed strategy profiles map one-to-one to behaviorally equivalent pure strategy profiles for larger signal spaces (cf. Footnote 7), it suffices to establish a bound on the learning rate for pure strategy equilibria.

Huang et al. (2024) show that in any strongly connected network, all agents learn at the same rate in any equilibrium. This allows us to leverage any bound on the slowest-learning agent's rate to conclude the same bound for every agent.[10]

Lemma 1 (Huang et al., 2024, Lemma 2)**.** *For any number of agents, any strongly connected network, and any discount factor $\delta \in [0, 1)$, all agents learn at the same rate in any equilibrium.*

This lemma and the bound on the learning rate of the slowest learning agent from Theorem 1 show that no agent can learn at a rate faster than r_{bdd} in any equilibrium for any number of agents if the network is strongly connected.

Corollary 1. *For any number of agents, any strongly connected network, and any discount factor $\delta \in [0, 1)$, each agent learns at rate at most r_{bdd} in any equilibrium.*

This extends Theorem 1 of Huang et al. (2024) to the case when the signals are not conditionally independent across agents and achieves a better bound on the learning rate.[11] Harel et al. (2021) obtain a better bound than Corollary 1 for two states and complete networks when the signals are conditionally independent and identically distributed across agents and periods and the agents are myopic ($\delta = 0$). The main feature of these results is that each agent's learning rate is bounded from above independently of the number of agents, showing that all but a vanishing fraction of private information is lost in large networks. We are however unable to give a nontrivial lower bound on the equilibrium learning rate analogous to Theorem 2 (cf. Remark 3).

[9] The assumption that the discount factor is the same for all agents is purely for notational simplicity. All results remain valid with heterogeneous discount factors.

[10] The proof of Lemma 2 of Huang et al. (2024) does not make use of the fact that signals are conditionally independent across agents and thus applies in the current setting.

[11] The bound on the learning rate obtained by Huang et al. (2024) is $2 \max_{\theta \neq \theta'} \max_{i \in N} \sup_{s \in S} |\ell^i_{\theta,\theta'}(s)| \geq \tilde{r}_{\text{bdd}} > 2r_{\text{bdd}}$. However, their model differs slightly from ours. They allow the signal space and the distribution of signals to depend on an agent's identity and the period as long as the log-likelihood ratios of signals are bounded uniformly over states, agents, and periods. On the other hand, they assume that the action space and the signal space are finite and that signals are conditionally independent across agents and periods.

For the rest of this section, we depart from the model and assume that agents observe their neighbors' actions and signals, and ask if this improves equilibrium learning. Clearly, a non-trivial bound on the learning rate cannot be independent of the network structure. Indeed, if a rational agent observes everyone's signals, she learns at rate nr_{aut}, which is not bounded independently of n. The problem becomes interesting when the neighborhoods have bounded size. Denote by $\Delta = \max_{i\in N} |N^i|$ the size of the largest neighborhood. We show that even if agents observe their neighbors' actions and signals, the learning rate of each agent is at most Δr_{bdd} in any equilibrium and any strongly connected network for any number of agents. This result assumes that signals are conditionally independent and identically distributed across agents and periods.

Corollary 2. *Assume the signals are conditionally independent across agents and periods. For any number of agents, any strongly connected network, and any discount factor $\delta \in [0,1)$, each agent learns at rate at most Δr_{bdd} in any equilibrium.*

First, it is clear from Lemma 1 that all agents learn at the same rate in any equilibrium. It thus suffices to show that some agent learns at most at the claimed rate. To this end, we embed the setup of Corollary 2 in the model where each agent only observes her neighbors' actions but not their signals. For each $i \in N$ and $t \in T$, let $\tilde{\mathfrak{s}}_t^i = (\mathfrak{s}_t^j)_{j\in N^i}$ be the vector with the signals of all of i's neighbors in period t. The distribution of $\tilde{\mathfrak{s}}_t^i$ in state $\theta \in \Omega$ is the $|N^i|$-fold product $\tilde{\mu}_\theta^i = (\mu_\theta^i)^{\otimes |N^i|}$ of μ_θ^i. Thus, defining $\tilde{\ell}^i_{\theta,\theta'} = \log \frac{d\tilde{\mu}^i_\theta}{d\tilde{\mu}^i_{\theta'}}(\tilde{\mathfrak{s}}_1^i)$ for two states $\theta, \theta' \in \Omega$, we have $\tilde{\ell}^i_{\theta,\theta'} = \sum_{j\in N^i} \ell^j_{\theta,\theta'}$ since the signals $(\mathfrak{s}_1^j)_{j\in N^i}$ are conditionally independent. Thus,

$$\min_{\theta,\theta'} \max_{i\in N} \mathrm{E}_\theta\left(\tilde{\ell}^i_{\theta,\theta'}\right) = \min_{\theta,\theta'} \max_{i\in N} \sum_{j\in N^i} \mathrm{E}_\theta\left(\ell^j_{\theta,\theta'}\right) \le \max_{i\in N} |N^i| \min_{\theta,\theta'} \max_{i\in N} \mathrm{E}_\theta\left(\ell^i_{\theta,\theta'}\right) = \Delta r_{\text{bdd}}$$

Applying Theorem 1 to the modified signals $(\tilde{\mathfrak{s}}_t^i)_{i\in N, t\in T}$ gives that some agent learns at rate at most Δr_{bdd}, concluding the proof. Note that modified signals are not conditionally independent across agents so that the full generality of Theorem 1 was needed.

7. Discussion

We conclude with several remarks about model variations and open problems.

Remark 1 (Genericity of the utility function). We have assumed that the agents' utility functions are suitably generic, i.e., that there is a unique optimal action in each state and no action is optimal in two different states. The second assumption is necessary to make the problem interesting: if the same action is optimal in all states, there is no need for information and all agents can choose optimally from the first period onward. The first assumption forces a tradeoff between correct action choices and actions as signaling devices. By contrast, if there are two optimal actions in each state and signals are binary (as in Section 3.4), then each agent can choose an action optimally based on her available information and simultaneously communicate her private signal in each period. Thus, actions are sufficient for identifying private signals and learning is as fast as if agents observed their neighbors' private signals, so that Theorem 1 fails. Note that the above strategies are even in equilibrium if the network is complete.[12]

Remark 2 (Arbitrarily correlated signals). We have assumed that the distributions of signal profiles are mutually absolutely continuous across states and that the signals are conditionally independent across periods. Theorem 1 breaks down emphatically without these assumptions.

Appropriate correlation across agents or periods allows even a small number of agents or a single agent in a small number of periods to learn the state with certainty. For example, consider the instance in Section 3.4 with signal precision $p = \frac{2}{3}$. If there are three agents and the distribution of signal profiles is such that exactly two agents receive a signal matching the state conditional on either state and each agent chooses the action matching her signal in the first period, then all agents know the state after period 1 and can choose optimally in all future periods. For correlation across periods, consider a single agent who receives exactly two signals matching the state in the first three periods. This reveals the state after at most three periods.

Remark 3 (Lower bounds for equilibrium learning). Theorem 2 shows that non-trivial information aggregation is possible if the agents follow prescribed strategies. However, it remains an open problem if there is any equilibrium for any network for which every agent learns faster than a single agent in autarky. Answering this question would likely require either new conceptual insights into equilibrium learning or constructing an equilibrium in which learning exceeds the autarky benchmark.

A related question is whether mediation of the information exchanged between the agents can improve equilibrium learning. The mediator observes all agents' actions but not their private signals and sends a private message to each agent. A special case is designing the network structure. For example, it could be beneficial for equilibrium learning if the mediator rewards an agent for deviating from a consensus action by giving her access to more agents' actions in future periods. This incentivizes agents to act based on their private signals rather than herd, potentially improving equilibrium learning

Remark 4 (Random networks). In an extension of our model, the network is drawn randomly (and independently from the state and the signals) according to some distribution in each period. The upper bound on the learning rate from Theorem 1 clearly remains valid in this setting since it holds even for complete networks. Theorem 2 also survives so long as all agents know each period's network and all networks are strongly connected. We sketch the necessary changes to the strategies in Footnote 21 in the proof of Theorem 2.

[12] This equilibrium is reminiscent of a construction by Heidhues et al. (2015), who consider bandit problems where agents observe each other's actions, but not the payoffs. They allow agents to communicate via cheap talk messages and show that cheap talk equilibria can replicate any equilibrium with publicly observable payoffs. In our model, a multiplicity of optimal actions enables similar cheap talk communication and can restore the public information case in equilibrium.

Acknowledgements

The author thanks Ilan Kremer for comments on the paper and acknowledges support by the DFG under the Excellence Strategy EXC-2047.

APPENDIX

A. Preliminaries

In this appendix, we recall a classic result from the theory of large deviations of random walks and start by setting up the notation.

Let ℓ be a bounded and non-degenerate random variable. The cumulant generating function of ℓ is

$$\lambda(z) = \log \mathrm{E}\left(e^{z\ell}\right)$$

Note that $\lambda(z)$ is finite for each $z \in \mathbb{R}$ since ℓ is bounded. The Fenchel-Legendre transform of λ is

$$\lambda^*(\eta) = \sup_{z \in \mathbb{R}} \eta z - \lambda(z)$$

We collect some properties of λ and λ^*.

Lemma 2 (Dembo and Zeitouni, 2009, Lemma 2.2.5)**.** *Let $I^* = \{\eta \in \mathbb{R} \colon \exists z \in \mathbb{R}, \lambda'(z) = \eta\}$. Then,*[13]

(i) λ is strictly convex, and λ^* is non-negative and convex.

(ii) For all $\eta \geq \mathrm{E}(\ell)$,

$$\lambda^*(\eta) = \sup_{z \geq 0} \eta z - \lambda(z)$$

and for all $\eta \leq \mathrm{E}(\ell)$,

$$\lambda^*(\eta) = \sup_{z \leq 0} \eta z - \lambda(z)$$

In particular, λ^* is non-decreasing on $[\mathrm{E}(\ell), \infty)$ and strictly increasing on $I^* \cap [\mathrm{E}(\ell), \infty)$, and it is non-increasing on $(-\infty, \mathrm{E}(\ell)]$ and strictly decreasing on $I^* \cap (-\infty, \mathrm{E}(\ell)]$. Moreover, $\lambda^*(\mathrm{E}(\ell)) = 0$.

(iii) λ is differentiable with

$$\lambda'(z) = \frac{\mathrm{E}\left(\ell e^{z\ell}\right)}{\mathrm{E}\left(e^{z\ell}\right)}$$

and if $\lambda'(z) = \eta$, then $\lambda^*(\eta) = \eta z - \lambda(z)$.

It follows from (iii) that $\lambda'(0) = \mathrm{E}(\ell)$ and that $\lambda'(z) \to \sup \ell$ as $z \to \infty$ and $\lambda'(z) \to \inf \ell$ as $z \to -\infty$. Hence, $I^* = (\inf \ell, \sup \ell)$, and so I^* contains an open neighborhood of $\mathrm{E}(\ell)$ (where we use that ℓ is non-degenerate).

[13] Since we assume that ℓ is bounded and non-degenerate, Lemma 2 avoids some case distinctions compared to Lemma 2.2.5 of Dembo and Zeitouni (2009) and allows strengthening convexity of λ to strict convexity.

Let $\ell_1, \ell_2, \ldots$ be independent random variables with the same distribution as ℓ, and for $t \in T$, let $L_t = \sum_{r \leq t} \ell_r$. The following result shows that the probability that L_t is less than ηt (plus a lower-order term) decreases exponentially in t at rate $\lambda^*(\eta)$ if η is smaller than $\mathrm{E}(\ell)$, and similarly if η is larger than $\mathrm{E}(\ell)$.

Theorem 3 (Cramér, 1938)**.** *Let $\eta \in \mathbb{R}$. If $\inf_{z\in\mathbb{R}} \lambda'(z) < \eta \leq \mathrm{E}(\ell)$, then*

$$P(L_t \leq \eta t + o(t)) = e^{-\lambda^*(\eta)t + o(t)}$$

and if $\mathrm{E}(\ell) \leq \eta < \sup_{z\in\mathbb{R}} \lambda'(z)$, then

$$P(L_t \geq \eta t + o(t)) = e^{-\lambda^*(\eta)t + o(t)}$$

In the stated version, Theorem 3 follows from Theorem 2.2.3 of Dembo and Zeitouni (2009) by recalling that λ^* is non-increasing on $(-\infty, \mathrm{E}(\ell)]$ and non-decreasing on $[\mathrm{E}(\ell), \infty)$, or by applying Theorem 6 of Harel et al. (2021) to ℓ and $-\ell$.

For $\theta, \theta' \in \Omega$ and $i \in N$, let $\lambda^i_{\theta,\theta'}$ be the cumulant generating function of $\ell^i_{\theta,\theta'}$, and denote by $(\lambda^i_{\theta,\theta'})^*$ its Fenchel-Legendre transform. It is not hard to show that $\lambda^i_{\theta,\theta'}(z) = \lambda^i_{\theta',\theta}(-(z+1))$ and $(\lambda^i_{\theta,\theta'})^*(\eta) = (\lambda^i_{\theta',\theta})^*(-\eta) - \eta$.[14] Hence, by Lemma 2(ii),

$$(\lambda^i_{\theta,\theta'})^*(-\mathrm{E}_{\theta'}(\ell^i_{\theta',\theta})) = (\lambda^i_{\theta',\theta})^*(\mathrm{E}_{\theta'}(\ell^i_{\theta',\theta})) + \mathrm{E}_{\theta'}(\ell^i_{\theta',\theta}) = \mathrm{E}_{\theta'}(\ell^i_{\theta',\theta}) \tag{1}$$

Since $\lambda^i_{\theta,\theta'}(0) = 0$, $\lambda^i_{\theta,\theta'}(-1) = \lambda^i_{\theta',\theta}(0) = 0$, and since $\lambda^i_{\theta,\theta'}$ is strictly convex, $\lambda^i_{\theta,\theta'}$ attains its minimum on $(-1, 0)$. Then, using again that $\lambda^i_{\theta,\theta'}$ is strictly convex,

$$(\lambda^i_{\theta,\theta'})^*(0) = \sup_{z\in\mathbb{R}} -\lambda^i_{\theta,\theta'}(z) = -\min_{z\in(0,1)} \lambda^i_{\theta,\theta'}(z) < (\lambda^i_{\theta,\theta'})'(0) = \mathrm{E}_\theta(\ell^i_{\theta,\theta'}) \tag{2}$$

B. Single-Agent Learning

We recall some results for a single agent learning in autarky and extend those to more than two states.

When there are only two states, the rate of learning a single agent can achieve in autarky is well-known. It essentially follows from Theorem 3 and appears as Fact 1 of Harel et al. (2021).[15]

[14] These relations appear in Lemma 6 of Harel et al. (2021). Since they use different sign conventions to define $\lambda^i_{\theta,\theta'}$ and $(\lambda^i_{\theta,\theta'})^*$, we reproduce the argument here. First,

$$\lambda^i_{\theta,\theta'}(z) = \log \int_S e^{z \log \frac{d\mu^i_\theta}{d\mu^i_{\theta'}}(s)} d\mu^i_\theta(s) = \log \int_S \left(\frac{d\mu^i_\theta}{d\mu^i_{\theta'}}(s)\right)^z d\mu^i_\theta(s) = \log \int_S \left(\frac{d\mu^i_{\theta'}}{d\mu^i_\theta}(s)\right)^{-(z+1)} d\mu^i_{\theta'}(s) = \lambda^i_{\theta',\theta}(-(z+1))$$

Thus,

$$(\lambda^i_{\theta,\theta'})^*(\eta) = \sup_{z\in\mathbb{R}} \eta z - \lambda^i_{\theta,\theta'}(z) = \sup_{z\in\mathbb{R}} \eta z - \lambda^i_{\theta',\theta}(-(z+1)) = \sup_{z\in\mathbb{R}} (-\eta) z - \lambda^i_{\theta',\theta}(z) - \eta = (\lambda^i_{\theta',\theta})^*(-\eta) - \eta$$

[15] Note that $0 \in (I^i_{g,f})^*$ by the remarks after Lemma 2 and the fact that $\sup \ell^i_{g,f} > 0$ and $\inf \ell^i_{g,f} < 0$.

Proposition 1 (Harel et al., 2021, Fact 1). *Let $\Omega = \{f, g\}$. The rate of learning of agent i in autarky is $(\lambda^i_{g,f})^*(0)$. More precisely, the probability that agent i makes a mistake in period t when choosing actions optimally based on her private signals is*

$$P\left(a^1_t \neq a_\omega\right) = e^{-(\lambda^i_{g,f})^*(0)t+o(t)}$$

For more than two states, the optimal rate of learning is determined by the two states that are the hardest to distinguish. More precisely, the optimal rate of learning equals the minimum of the optimal learning rates when restricting to any pair of states. As for the case of two states, the "maximum likelihood strategy" achieves the highest learning rate: in any period, choose the action that is optimal in a most probable state. This result can be obtained from Theorem 2.2.30 of Dembo and Zeitouni (2009). We state it here along with a proof that is specific to our setting.

Corollary 3. *The probability that agent i learning in autarky and choosing actions optimally makes a mistake in period t is*

$$P\left(a^i_t \neq a_\omega\right) = e^{-r^i_{\text{aut}}t+o(t)}$$

where

$$r^i_{\text{aut}} = \min_{\theta\neq\theta'}(\lambda^i_{\theta,\theta'})^*(0)$$

Proof. Let

$$r = \sup_{\sigma^i} \liminf_{t\to\infty} -\frac{1}{t}\log P\left(a^i_t \neq a_\omega\right)$$

where $a^i_t = \sigma^i_t(\mathfrak{s}^i_1, \dots, \mathfrak{s}^i_t)$ and the supremum is taken over all strategies $\sigma^i = (\sigma^i_1, \sigma^i_2, \dots)$ of agent i in autarky. Hence, r is the optimal rate of learning, and we have to show that $r = r^i_{\text{aut}}$.

Step 1 ($r \leq r^i_{\text{aut}}$). Assume for contradiction that $r > r^i_{\text{aut}}$, and let σ^i be a strategy such that

$$\liminf_{t\to\infty} -\frac{1}{t}\log P\left(a^i_t \neq a_\omega\right) > r^i_{\text{aut}}$$

Then, there are $f, g \in \Omega$ such that $\liminf_{t\to\infty} -\frac{1}{t}\log P\left(a^i_t \neq a_\omega\right) > (\lambda^i_{g,f})^*(0)$. Note that σ^i is also a strategy for the problem after restricting to the subset $\{f, g\}$ of states, and as such achieves a rate of learning of

$$\liminf_{t\to\infty} -\frac{1}{t}\log P\left(a^i_t \neq a_\omega \mid \omega \in \{f,g\}\right) \geq \liminf_{t\to\infty} -\frac{1}{t}\log P\left(a^i_t \neq a_\omega\right) > (\lambda^i_{g,f})^*(0)$$

where the first inequality follows from the fact that $P\left(a^i_t \neq a_\omega \mid \omega \in \{f,g\}\right)P\left(\omega \in \{f,g\}\right) \leq P\left(a^i_t \neq a_\omega\right)$ and $P\left(\omega \in \{f,g\}\right) > 0$. But this contradicts Proposition 1.

Step 2 ($r \geq r^i_{\text{aut}}$). It suffices to find a strategy that achieves a learning rate of at least r^i_{aut}. Let $\sigma^i = (\sigma^i_1, \sigma^i_2, \dots)$ be a strategy with

$$a^i_t = \sigma^i_t(\mathfrak{s}^i_1, \dots, \mathfrak{s}^i_t) \in \{a_\theta \in A \colon \theta \in \Omega, L^i_{\theta,\theta',t} \geq 0 \;\forall \theta' \in \Omega\}$$

for all $t \in T$. Thus, σ^i is a "maximum-likelihood strategy", i.e., it chooses the optimal action for a most probable state.[16] Then, for all $\theta \in \Omega$,

$$P_\theta\left(a_t^i \neq a_\theta\right) \leq \sum_{\theta' \in \Omega} P_\theta\left(L_{\theta,\theta',t}^i \leq 0\right) = \sum_{\theta' \in \Omega} e^{-(\lambda_{\theta,\theta'}^i)^*(0)t+o(t)} \leq e^{-r_{\text{aut}}^i + o(t)}$$

where the first inequality follows from the definition of σ^i, the equality follows from Theorem 3, and the second inequality uses the definition of r_{aut} (and may require adjusting the lower-order term $o(t)$). Hence,

$$P\left(a_t^i \neq a_\omega\right) = \sum_{\theta \in \Omega} P_\theta\left(a_t^i \neq a_\theta\right) P\left(\omega = \theta\right) \leq e^{-r_{\text{aut}}^i + o(t)}$$

as required.

□

C. Omitted Proofs From Section 5

In this section, we prove the upper and lower bound on the optimal learning rate claimed in Theorem 1 and Theorem 2.

We start with three auxiliary lemmas. The first heuristically states the following. Fix a (one-dimensional) random walk with i.i.d. increments, an upper (lower) slope larger (smaller) than the expectation of each increment, and a probability threshold smaller than 1. Then, for each period t_0, the probability that the random walk remains inside an affine wedge with the given upper and lower slopes and sufficiently large intercepts in all periods after t_0 conditional on being well inside the wedge at t_0 exceeds the threshold.

Lemma 3. *Let $\ell_1, \ell_2, \dots$ be bounded, non-degenerate i.i.d. random variables and let $a^+, a^- \in \mathbb{R}$ with $a^+ > \mathrm{E}(\ell_1)$ and $a^- < \mathrm{E}(\ell_1)$. For $t \in T$, let $L_t = \sum_{r \leq t} \ell_r$, and let $L = (L_1, L_2, \dots)$. Then, for each $\delta > 0$, there is $K > 0$ such that for each $b \in \mathbb{R}$, each $t_0 \in T$, and almost every (w.r.t. the distribution of $(L_1, \dots, L_{t_0})$) $x_{\leq t_0} \in \mathbb{R}^{t_0}$ with $b - \frac{K}{2} \leq x_{t_0} \leq b + \frac{K}{2}$,*

$$P\left(L \in \mathcal{W} \mid L_{\leq t_0} = x_{\leq t_0}\right) \geq 1 - \delta$$

where $\mathcal{W} = \{x \in \mathbb{R}^T : \forall t \geq t_0, b - K + a^-(t - t_0) \leq x_t \leq b + K + a^+(t - t_0)\}$.

Proof. Denote by λ the cumulant generating function of ℓ_1 and denote by λ^* its Fenchel-Legendre transform. First, observe that the statement becomes stronger if a^+ and a^- are closer to $\mathrm{E}(\ell_1)$. Hence, we may and will assume that $\mathrm{E}(\ell_1) < a^+ < \sup_{z \in \mathbb{R}} \lambda'(z)$ and $\inf_{z \in \mathbb{R}} \lambda'(z) < a^- < \mathrm{E}(\ell_1)$, so that Theorem 3 applies.

Fix $\delta > 0$. Let $b \in \mathbb{R}$, $t_0 \in T$, and $x_{\leq t_0} \in \mathbb{R}^{t_0}$ with $b - \frac{K}{2} \leq x_{t_0} \leq b + \frac{K}{2}$. Let $\tilde{a}^+ = \frac{1}{2}(\mathrm{E}(\ell_1) + a^+)$ and $\tilde{a}^- = \frac{1}{2}(\mathrm{E}(\ell_1) + a^-)$. By Lemma 2(ii), $\lambda^*(\tilde{a}^+) > 0$ and $\lambda^*(\tilde{a}^-) > 0$. Thus, for all $t \geq t_0$, we

[16] Note that σ^i is well-defined since $L_{\theta,\theta',t}^i + L_{\theta',\theta'',t}^i = L_{\theta,\theta'',t}^i$ ensuring that the right-hand side in the definition of σ^i is always non-empty.

have by Theorem 3 that

$$P\left(L_t - L_{t_0} \geq \tilde{a}^+(t-t_0) + o(t-t_0)\right) = e^{-\lambda^*(\tilde{a}^+)(t-t_0)+o(t-t_0)}, \text{ and}$$
$$P\left(L_t - L_{t_0} \leq \tilde{a}^-(t-t_0) + o(t-t_0)\right) = e^{-\lambda^*(\tilde{a}^-)(t-t_0)+o(t-t_0)}$$

Thus, there is $t_1 \geq t_0$ such that

$$P\left(\forall t \geq t_1, a^-(t-t_0) \leq L_t - L_{t_0} \leq a^+(t-t)\right) \geq 1-\delta$$

Let $K = 2(t_1 - t_0)\sup|\ell_1|$. Then, it follows that

$$P\left(\forall t \geq t_0, -\frac{K}{2} + a^-(t-t_0) \leq L_t - L_{t_0} \leq \frac{K}{2} + a^+(t-t_0)\right) \geq 1-\delta$$

But $\ell_1, \ell_2, \ldots$ are independent and $b - \frac{K}{2} \leq x_{t_0} \leq b + \frac{K}{2}$ and so

$$P\left(L \in \mathcal{W} \mid L_{\leq t_0} = x_{\leq t_0}\right) \geq P\left(\forall t \geq t_0, -\frac{K}{2} + a^-(t-t_0) \leq L_t - L_{t_0} \leq \frac{K}{2} + a^+(t-t_0)\right) \geq 1-\delta$$

which finishes the proof. □

The second lemma states the following. Assume the agents follow strategies for which each of them learns at a strictly positive rate r. Then, there is a history $H_{\leq t_0}$ up to t_0 such that conditional on $H_{\leq t_0}$, each agent's probability of a mistake decreases exponentially at a rate close to r from t_0 onward. Heuristically, this turns the limit defining the learning rate into a statement about each sufficiently late period at a small cost in the learning rate.

Lemma 4. *Fix the number of agents n, a state $g \in \Omega$, and a learning rate $r > 0$. Let $\sigma^1, \ldots, \sigma^n$ be strategies such that for each $i \in N$,*

$$\liminf_{t\to\infty} -\frac{1}{t}\log P\left(a_t^i \neq a_g\right) \geq r$$

Then, for each $\epsilon > 0$, there are $t_0 \in T$ and $H_{\leq t_0} \in A^{N\times t_0}$ such that $P_g\left(H_{\leq t_0}\right) > 0$ and for each $i \in N$ and $t > t_0$,

$$P_g\left(a_t^i \neq a_g \mid H_{\leq t_0}\right) \leq e^{-(r-\epsilon)t}$$

Moreover, t_0 may be chosen arbitrarily large.

Proof. By assumption, for each $i \in N$ and each $t \in T$,

$$P_g\left(a_t^i \neq a_g\right) \leq e^{-rt+o(t)}$$

First, for $\tilde{t}_0 \in T$ large enough,

$$\sum_{i\in N, t>\tilde{t}_0} P_g\left(a_t^i \neq a_g\right) \leq \frac{1}{2}$$

and

$$\sum_{i\in N, t>\tilde{t}_0} P_g\left(a_t^i \neq a_g\right) = \sum_{H_{\leq \tilde{t}_0}\in A^{N\times \tilde{t}_0}, P_g\left(H_{\leq \tilde{t}_0}\right)>0} P_g\left(H_{\leq \tilde{t}_0}\right) \sum_{i\in N, t>\tilde{t}_0} P_g\left(a_t^i \neq a_g \mid H_{\leq \tilde{t}_0}\right)$$

and, thus, there is $H^*_{\leq \tilde{t}_0} \in A^{N \times \tilde{t}_0}$ such that $P_g\left(H^*_{\leq \tilde{t}_0}\right) > 0$ and

$$\sum_{i \in N, t > \tilde{t}_0} P_g\left(a^i_t \neq a_g \mid H^*_{\leq \tilde{t}_0}\right) \leq \frac{1}{2} \tag{3}$$

Let $H \in A^{N \times T}$ such that $H_{\leq \tilde{t}_0} = H^*_{\leq \tilde{t}_0}$ and for each $i \in N$ and each $t > \tilde{t}_0$, $H^i_t = a_g$. Note that, by (3),

$$P_g\left(H \mid H_{\leq \tilde{t}_0}\right) \geq \frac{1}{2} \tag{4}$$

By the assumption of the lemma, for each $i \in N$, each $t_0 \geq \tilde{t}_0$ large enough, and each $t > t_0$,

$$P_g\left(a^i_t \neq a_g \mid H_{\leq \tilde{t}_0}\right) \leq \frac{1}{2} e^{-(r-\epsilon)t}$$

and, thus, by (4),

$$P_g\left(a^i_t \neq a_g \mid H_{\leq t_0}\right) \leq 2P_g\left(a^i_t \neq a_g \mid H_{\leq t_0}\right) P_g\left(H_{\leq t_0} \mid H_{\tilde{t}_0}\right) \leq 2P_g\left(a^i_t \neq a_g \mid H_{\leq \tilde{t}_0}\right) \leq e^{-(r-\epsilon)t}$$

which proves the claim. □

It is well-known that if μ_t, ν_t, $t \in \{1, 2\}$, are probability measures such that ν_t is absolutely continuous with respect to μ_t, then $\nu_1 \otimes \nu_2$ is absolutely continuous with respect to $\mu_1 \otimes \mu_2$, where $\nu_1 \otimes \nu_2$ is the product measure on the corresponding product space.[17] Thus, since the distributions of signal profiles in different states are mutually absolutely continuous and signals are conditionally independent across periods, the same is true for the distributions of signal profile trajectories up to any period t_0. This is the content of the third lemma. We omit the easy proof.

Lemma 5. *For each $t \in T$ and each $\theta \in \Omega$, denote by $\mu^{\otimes t}_\theta$ the t-fold product measure of μ_θ on $S^{N \times t}$. Then, for each $t \in T$ and each $\theta, \theta' \in \Omega$, $\mu^{\otimes t}_\theta$ and $\mu^{\otimes t}_{\theta'}$ are mutually absolutely continuous.*

Now we are ready for the proof of Theorem 1.

Theorem 1 (Learning is bounded)**.** *For any number of agents n, any network, and any strategies $\sigma^1, \ldots, \sigma^n$, some agent learns at rate at most $r_{\text{bdd}} = \min_{\theta \neq \theta'} \max_{i \in N} \mathrm{E}_\theta\left(\ell^i_{\theta,\theta'}\right)$.*

Proof. Let $f, g \in \Omega$ such that $r_{\text{bdd}} = \max_{i \in N} \mathrm{E}_f\left(\ell^i_{f,g}\right)$. Fix strategies $\sigma^1, \ldots, \sigma^n$. It is clearly without loss of generality to assume that $N^i = N$ for each $i \in N$. Assume for contradiction that there is $\epsilon > 0$ such that for each $i \in N$ and each $t \in T$,

$$P\left(a^i_t \neq a_\omega\right) \leq e^{-(r_{\text{bdd}}+6\epsilon)t + o(t)} \tag{5}$$

We write $r = r_{\text{bdd}}$ in the rest of the proof for convenience. For each $\theta \in \Omega$, we denote by $P_\theta(\cdot \mid \mathfrak{s} = \cdot)$ a regular conditional probability, and write $P_\theta(\cdot \mid \mathfrak{s} = \cdot) = P_\theta(\cdot \mid \cdot)$ again for convenience.[18]

As promised by Lemma 4, there are $t_0 \in T$ and $H \in A^{N \times T}$ such that

[17] A proof of this fact can be found here: `https://math.stackexchange.com/q/1042323`.

[18] That is, (i) for each $s \in S^T$, $P_\theta(\cdot \mid \mathfrak{s} = s)$ is a probability measure on the underlying probability space, (ii) for each event E, $P_\theta(E \mid \mathfrak{s} = \cdot)$ is measurable, and (iii) for each event E and each measurable set $\mathcal{S} \subset S^{N \times T}$, $P_\theta\left(E \cap \mathfrak{s}^{-1}(\mathcal{S})\right) = \int_{\mathcal{S}} P_\theta(E \mid \mathfrak{s} = s) dP \circ \mathfrak{s}^{-1}(s)$.

(i) $P_g\left(H_{\leq t_0}\right) > 0$,

(ii) for each $i \in N$ and each $t > t_0$, $P_g\left(a_t^i \neq a_g \mid H_{\leq t_0}\right) \leq e^{-(r+5\epsilon)t}$,

(iii) $\sum_{t>t_0} e^{-\epsilon t} < \frac{1}{32n^2}$ and $\epsilon t_0 \geq K$, and

(iv) for each $i \in N$ and each $t > t_0$, $H_t^i = a_g$,

where K is the constant obtained from applying Lemma 3 to $\ell_1, \ell_2, \ldots$ each with distribution $P_f\left(\ell_{f,g}^i \in \cdot\right)$, $a^+ = \mathrm{E}_f\left(\ell_{f,g}^i\right) + \epsilon$, $a^- = \mathrm{E}_f\left(\ell_{f,g}^i\right) - \epsilon$, and $\delta = \frac{1}{2n}$ for each $i \in N$ separately and taking the maximum.[19]

We proceed in multiple steps.

Step 1. For each $i \in N$, each $t \in T$, and each $s_{\leq t}^i \in S^t$, denote by $L_{g,f}^i(s_{\leq t}^i)$ agent i's log-likelihood ratio for g over f after observing the private signals $s_{\leq t}^i$:

$$L_{g,f}^i(s_{\leq t}^i) = \log \frac{\pi_0(g)}{\pi_0(f)} + \sum_{t' \leq t} \ell_{g,f}^i(s_{t'}^i)$$

First, we show that there is a set of signal profile trajectories $\mathcal{S}_{\leq t_0} \subset S^{N \times t_0}$ up to t_0 and $m^1, \ldots, m^n \in \mathbb{R}$ such that (i) $\mathcal{S}_{\leq t_0} \subset \mathcal{S}(H_{\leq t_0})$ (i.e., each trajectory of profiles in $\mathcal{S}_{\leq t_0}$ induces $H_{\leq t_0}$), (ii) $P_g\left(\mathcal{S}_{\leq t_0}\right) > 0$, (iii) for each $i \in N$, each $t > t_0$, and almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$, $P_g\left(a_t^i \neq a_g \mid s_{\leq t_0}\right) \leq e^{-(r+4\epsilon)t}$, and (iv) for each $i \in N$ and each $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$, $(m^i - \frac{\epsilon}{2})t_0 \leq L_{g,f}^i(s_{\leq t_0}^i) \leq (m^i + \frac{\epsilon}{2})t_0$ (i.e., all trajectories in $\mathcal{S}_{\leq t_0}$ induce roughly the same log-likelihood ratios).

For each $i \in N$ and each $t > t_0$, let $\mathcal{S}_{\leq t_0}(i,t) \subset \mathcal{S}(H_{\leq t_0})$ be the set of signal profile trajectories up to t_0 that induce $H_{\leq t_0}$ and after which the probability that agent i does not choose a_g in period t and state g is at least $e^{-(r+4\epsilon)t}$. Formally, for each $i \in N$ and each $t > t_0$, let $\mathcal{S}_{\leq t_0}(i,t) \subset \mathcal{S}(H_{\leq t_0})$ such that for almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}(i,t)$,

$$P_g\left(a_t^i \neq a_g \mid s_{\leq t_0}\right) \geq e^{-(r+4\epsilon)t}$$

and for almost every $s_{\leq t_0} \notin \mathcal{S}_{\leq t_0}(i,t)$, the reverse inequality holds. Then, for each $i \in N$ and each $t > t_0$,

$$\begin{aligned} e^{-(r+4\epsilon)t} P_g\left(\mathcal{S}_{\leq t_0}(i,t) \mid H_{\leq t_0}\right) &\leq \int_{\mathcal{S}_{\leq t_0}(i,t)} P_g\left(a_t^i \neq a_g \mid s_{\leq t_0}\right) dP_g\left(s_{\leq t_0} \mid H_{\leq t_0}\right) \\ &\leq P_g\left(a_t^i \neq a_g \mid H_{\leq t_0}\right) \\ &\leq e^{-(r+5\epsilon)t} \end{aligned}$$

and thus, $P_g\left(\mathcal{S}_{\leq t_0}(i,t) \mid H_{\leq t_0}\right) \leq e^{-\epsilon t}$. Then, let

$$\hat{\mathcal{S}}_{\leq t_0} = \mathcal{S}(H_{\leq t_0}) \setminus \bigcup_{i \in N, t > t_0} \mathcal{S}_{\leq t_0}(i,t)$$

[19] Here, $P_f\left(\ell_{f,g}^i \in \cdot\right)$ denotes the distribution of $\ell_{f,g}^i$ conditional on state f.

for which we have

$$P_g\left(\hat{\mathcal{S}}_{\leq t_0} \mid H_{\leq t_0}\right) \geq 1 - \sum_{i\in N, t>t_0} P_g\left(\mathcal{S}_{\leq t_0}(i,t) \mid H_{\leq t_0}\right) \geq 1 - n\sum_{t>t_0} e^{-\epsilon t} > 0$$

It thus follows from $P_g\left(H_{\leq t_0}\right) > 0$ that $P_g\left(\hat{\mathcal{S}}_{\leq t_0}\right) > 0$, and that for each $i \in N$, each $t > t_0$, and almost every $s_{\leq t_0} \in \hat{\mathcal{S}}_{\leq t_0}$,

$$P_g\left(a_t^i \neq a_g \mid s_{\leq t_0}\right) \leq e^{-(r+4\epsilon)t}$$

by construction of $\hat{\mathcal{S}}_{\leq t_0}$. Since $\mathbb{R}^n$ can be covered by countably many cubes with side length ϵ, there are $\mathcal{S}_{\leq t_0} \subset \hat{\mathcal{S}}_{\leq t_0}$ and $m^1, \ldots, m^n \in \mathbb{R}$ such that $P_g\left(\mathcal{S}_{\leq t_0}\right) > 0$ and (iv) holds. Moreover, $\mathcal{S}_{\leq t_0}$ satisfies (i) and (iii) since those are preserved under passing to subsets. This finishes the construction.

Step 2. Second, let

$$\begin{aligned}\mathcal{W} =\{&s \in S^{N\times T} \colon \forall i \in N, \forall t \geq t_0,\\ &(m^i - \epsilon)t_0 - (\mathrm{E}_f\left(\ell^i_{f,g}\right) + \epsilon)(t - t_0) \leq L^i_{g,f}(s^i_{\leq t}) \leq (m^i + \epsilon)t_0 - (\mathrm{E}_f\left(\ell^i_{f,g}\right) - \epsilon)(t - t_0)\}\end{aligned}$$

be those trajectories of profiles such that the log-likelihood ratio for g over f of each agent i based only on her private signals remains in a wedge with slopes $-\mathrm{E}_f\left(\ell^i_{f,g}\right) + \epsilon$ and $-\mathrm{E}_f\left(\ell^i_{f,g}\right) - \epsilon$ for every period from t_0 onward. By Property (iv) in Step 1, $\mathcal{S}_{\leq t_0} \subset \mathcal{W}_{\leq t_0}$. Moreover, by Lemma 3 and (iii) in Step 1, for almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$, $P_f\left(\mathcal{W} \mid s_{\leq t_0}\right) \geq 1 - \frac{n}{2n} = \frac{1}{2}$. Note that for each $i \in N$, each $t \geq t_0$, and each $\mathcal{V}^i_{\leq t} \subset \mathcal{W}^i_{\leq t}$,

$$e^{-(m^i+\epsilon)t_0+(\mathrm{E}_f\left(\ell^i_{f,g}\right)-\epsilon)(t-t_0)} P_g\left(\mathcal{V}^i_{\leq t}\right) \leq P_f\left(\mathcal{V}^i_{\leq t}\right) \leq e^{-(m^i-\epsilon)t_0+(\mathrm{E}_f\left(\ell^i_{f,g}\right)+\epsilon)(t-t_0)} P_g\left(\mathcal{V}^i_{\leq t}\right) \tag{6}$$

Step 3. Third, we show that for each $i \in N$, each $t \geq t_0$, and almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$,

$$P_g\left(a_t^i \neq a_g \mid s_{\leq t_0}, H_{<t}\right) \leq 2e^{-(r+4\epsilon)t} \tag{7}$$

Fix $i \in N$ and $t > t_0$. By Step 2, Item (iii) in Step 1, and the choice of t_0, for almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$,

$$\begin{aligned}P_g\left(H_{<t} \mid s_{\leq t_0}\right) &\geq 1 - P_g\left(\neg H_{<t} \mid s_{\leq t_0}\right)\\ &\geq 1 - \sum_{j\in N, t'>t_0} P_g\left(a^j_{t'} \neq a_g \mid s_{\leq t_0}\right)\\ &\geq 1 - n\sum_{t'>t_0} e^{-(r+4\epsilon)t'}\\ &\geq \frac{1}{2}\end{aligned}$$

and, moreover,

$$P_g\left(a_t^i \neq a_g \mid s_{\leq t_0}, H_{<t}\right) P_g\left(H_{<t} \mid s_{\leq t_0}\right) \leq P_g\left(a_t^i \neq a_g \mid s_{\leq t_0}\right)$$

Then (7) follows using again Item (iii) in Step 1.

We continue by setting up for the rest of the proof. For each $i \in N$ and each $t > t_0$, we say that agent i defects in period t if play follows $H_{<t}$ up to period $t-1$ and i does not choose a_g in period t, and we define

$$D_t^i = \left\{ s^i \in S^T : \sigma_t^i(s_{\leq t}^i; H_{<t}) \neq a_g \right\}$$

as those infinite trajectories s^i for i such that i defects at t for $s_{\leq t}^i$. Note that, by (7), for almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$,

$$P_g\left(D_t^i \mid s_{\leq t_0}, H_{<t}\right) = P_g\left(a_t^i \neq a_g \mid s_{\leq t_0}, H_{<t}\right) \leq 2e^{-(r+4\epsilon)t} \tag{8}$$

The rest of the proof proceeds as follows. To simplify the outline, we pretend that $t_0 = 0$ (and thus $\mathcal{S}_{\leq t_0} = \{\emptyset\}$ contains only the empty signal profile trajectory) and that $\mathcal{W} = S^{N\times T}$. The goal is to show that $P_f(H) > 0$, which would imply that each agent's learning rate is 0 and thus contradict the assumption.

Here is an initial attempt that fails but is a useful starting point. It would suffice to show that for each i and each $t > t_0$, $P_f\left(D_t^i\right) \leq e^{-\epsilon t}$, which would follow from $P_g\left(D_t^i\right) \leq e^{-(r+3\epsilon)t}$ by the definition of r. This is close to (8), except that (8) involves conditioning on $H_{<t}$. Unfortunately, $P_g\left(D_t^i\right) \leq e^{-(r+3\epsilon)t}$ does not follow from (8) since $P_g\left(D_t^i \mid H_{<t}\right)$ can be arbitrarily small compared to $P_g\left(D_t^i\right)$. This dead end inspires a more nuanced approach.

We split up agent i's infinite signal trajectories. For each t, E_t^i consists of those trajectories s^i for which there is some j such that $P_g\left(D_t^j \mid s^i, H_{<t}\right)$ is not too small and t is the first period for which there is such j, and F^i consists of those trajectories s^i for which $P_g\left(D_t^j \mid s^i, H_{<t}\right)$ is small for all j and t. For a trajectory $s^i \in E_t^i \cup F^i$, $P_g\left(H_{<t} \mid s^i\right)$ is large since no agent is likely to defect in any period before t conditional on s^i (see (10) below). Thus, one can bound the probability of E_t^i in terms of its probability conditional on $H_{<t}$, which in turn can be bounded in terms of the probability that some agent defects in period t conditional on a trajectory in E_t^i and $H_{<t}$ by the definition of E_t^i (see (11) below). Similarly, for each t, one can bound the probability of those trajectories in F^i for which agent i defects in period t in terms of the same probability conditional on $H_{<t}$ (see (12) below). Combining both estimates and using that defections are very unlikely in state g by assumption, gives that the intersection of D_t^i with each of the sets $E_{t'}^i$, $t' \in T$, and F^i has very low probability in state g and thus still moderately low probability in state f. But then, it is unlikely that any agent defects in state f, and all agents indefinitely play a_g with positive probability.

Step 4. Fourth, since signals are conditionally independent across periods, for each $i, j \in N$ and $t > t_0$,

$$P_g\left(a_t^j \neq a_g \mid \mathfrak{s}_{\leq t_0}, \mathfrak{s}^i, H_{<t}\right) = P_g\left(a_t^j \neq a_g \mid \mathfrak{s}_{\leq t_0}, \mathfrak{s}_{\leq t}^i, H_{<t}\right) \tag{9}$$

almost surely. For each $i \in N$ and each $t > t_0$, define inductively

$$E_t^i(s_{\leq t_0}) = \left\{ s^i \in \{s_{\leq t_0}^i\} \times S^{\{t_0+1, t_0+2, \dots\}} : \exists j \in N, P_g\left(a_t^j \neq a_g \mid s_{\leq t_0}, s^i, H_{<t}\right) \geq e^{-\epsilon t} \right\} \setminus \bigcup_{t > t' > t_0} E_{t'}^i(s_{\leq t_0})$$

and define

$$F^i(s_{\le t_0}) = \left\{ s^i \in \{s^i_{\le t_0}\} \times S^{\{t_0+1,t_0+2,\dots\}} \colon \forall j \in N, \forall t > t_0, P_g\left(a^j_t \neq a_g \mid s_{\le t_0}, s^i, H_{<t}\right) < e^{-\epsilon t} \right\}$$

and note that $\{E^i_t(s^i_{\le t_0})\colon t > t_0\} \cup \{F^i(s_{\le t_0})\}$ is a partition of $\{s^i_{\le t_0}\} \times S^{\{t_0+1,t_0+2,\dots\}}$. Intuitively, $E^i_t(s_{\le t_0})$ is the set of infinite trajectories of i's signals with prefix $s^i_{\le t_0}$ such that conditional on $s_{\le t_0}$, s^i, and $H_{<t}$, some agent defects with appreciable probability in period t and no agent does so in any period between t_0 and t, and $F^i(s_{\le t_0})$ contains those infinite trajectories for which no agent defects with appreciable probability in any period after t_0. It may help to think of the sets $E^i_t(s_{\le t_0})$ as "empty" in the sense that some other agent exhausts an appreciable fraction of them through defections, and to think of the sets $F^i(s_{\le t_0})$ as "full" in the sense that they are not significantly reduced by a defection of any agent.

We establish bounds on the probabilities of these sets. First, for each $i \in N$, each $t > t_0$, almost every $s_{\le t_0} \in \mathcal{S}_{\le t_0}$, and almost every $s^i \in E^i_t(s_{\le t_0}) \cup F^i(s_{\le t_0})$,

$$\begin{aligned}
P_g\left(H_{<t} \mid s_{\le t_0}, s^i\right) &= \prod_{t > t' > t_0} P_g\left(H_{<t'+1} \mid s_{\le t_0}, s^i, H_{<t'}\right) \\
&= \prod_{t > t' > t_0} \left(1 - P_g\left(\exists j \in N, a^j_{t'} \neq a_g \mid s_{\le t_0}, s^i, H_{<t'}\right)\right) \\
&\ge 1 - \sum_{j \in N, t > t' > t_0} P_g\left(a^j_{t'} \neq a_g \mid s_{\le t_0}, s^i, H_{<t'}\right) \qquad (10) \\
&\ge 1 - n \sum_{t > t' > t_0} e^{-\epsilon t'} \\
&\ge \frac{1}{2}
\end{aligned}$$

where the fourth step uses that $s^i \notin E^i_{t'}(s_{\le t_0})$ for $t' < t$. Thus, for each $i \in N$, each $t > t_0$, and almost every $s_{\le t_0}$,

$$\begin{aligned}
P_g\left(E^i_t(s_{\le t_0}) \mid s_{\le t_0}\right) &= \int_{E^i_t(s_{\le t_0})} dP_g\left(s^i \mid s_{\le t_0}\right) \\
&\le 2 \int_{E^i_t(s_{\le t_0})} P_g\left(H_{<t} \mid s_{\le t_0}, s^i\right) dP_g\left(s^i \mid s_{\le t_0}\right) \\
&= 2P_g\left(E^i_t(s_{\le t_0}) \cap H_{<t} \mid s_{\le t_0}\right) \\
&\le 2P_g\left(E^i_t(s_{\le t_0}) \mid s_{\le t_0}, H_{<t}\right) \\
&= 2 \int_{E^i_t(s_{\le t_0})} dP_g\left(s^i \mid s_{\le t_0}, H_{<t}\right) \qquad (11) \\
&\le 2e^{\epsilon t} \int_{E^i_t(s_{\le t_0})} \sum_{j \in N} P_g\left(a^j_t \neq a_g \mid s_{\le t_0}, s^i, H_{<t}\right) dP_g\left(s^i \mid s_{\le t_0}, H_{<t}\right) \\
&\le 2e^{\epsilon t} \sum_{j \in N} P_g\left(a^j_t \neq a_g \mid s_{\le t_0}, H_{<t}\right) \\
&\le 4ne^{\epsilon t} e^{-(r+4\epsilon)t} \\
&= 4ne^{-(r+3\epsilon)t}
\end{aligned}$$

where the second step uses (10), the sixth step uses the definition of $E^i_t(s_{\le t_0})$, and the second to last step uses (7). In words, the left hand side is the probability that, conditional on g and $s_{\le t_0}$, agent i observes an infinite trajetory conditional on which some agent defects with appreciable probability in period t and no agent defects with appreciable probability in any period before t. Second, for each $i \in N$, each $t > t_0$, and almost every $s_{\le t_0} \in \mathcal{S}_{\le t_0}$,

$$\begin{aligned} P_g\left(\left(D^i_t \cap F^i(s_{\le t_0})\right)_{\le t} \mid s_{\le t_0}\right) &\le 2P_g\left(\left(D^i_t \cap F^i(s_{\le t_0})\right)_{\le t} \mid s_{\le t_0}, H_{<t}\right) \\ &\le 2P_g\left(a^i_t \neq a_g \mid s_{\le t_0}, H_{<t}\right) \\ &\le 4e^{-(r+4\epsilon)t} \end{aligned} \tag{12}$$

where the first step follows from repeating the first four steps in (11) and the third step follows from (7). Here, the left hand side is the probability that conditional on g and $s_{\le t_0}$, agent i observes a trajetory up to t for which she defects in period t and conditional on which no agent defects with appreciable probability in any period from t_0+1 to t.

Step 5. We now show that, conditional on state f and almost every signal profile trajectory $s_{\le t_0} \in \mathcal{S}_{\le t_0}$, with positive probability, no agents defects in any period.

First, for each $i \in N$, each $t > t_0$, and almost every $s_{\le t_0} \in \mathcal{S}_{\le t_0}$,

$$\begin{aligned} P_f\left(E^i_t(s_{\le t_0}) \mid s_{\le t_0}, \mathcal{W}\right) &\le 2P_f\left(E^i_t(s_{\le t_0}) \cap \mathcal{W} \mid s_{\le t_0}\right) \\ &\le 2P_f\left((E^i_t(s_{\le t_0}) \cap \mathcal{W}^i)_{\le t} \mid s_{\le t_0}\right) \\ &\le 2e^{(m^i+\epsilon)t_0}e^{-(m^i-\epsilon)t_0+(\mathrm{E}_f(\ell^i_{f,g})+\epsilon)(t-t_0)}P_g\left((E^i_t(s_{\le t_0}) \cap \mathcal{W}^i)_{\le t} \mid s_{\le t_0}\right) \\ &\le 2e^{2\epsilon t_0+(\mathrm{E}_f(\ell^i_{f,g})+\epsilon)(t-t_0)}P_g\left((E^i_t(s_{\le t_0}) \cap \mathcal{W}^i)_{\le t} \mid s_{\le t_0}\right) \\ &\le 2e^{(r+2\epsilon)t}P_g\left((E^i_t(s_{\le t_0}) \cap \mathcal{W}^i)_{\le t} \mid s_{\le t_0}\right) \\ &\le 8ne^{-\epsilon t} \end{aligned}$$

where the first step uses that $P_f\left(\mathcal{W} \mid s_{\le t_0}\right) \ge \frac{1}{2}$ by Step 2, the third step uses (6), and the last step uses (11) and the fact that, by (9), whether an infinite trajectory of i's signals is in $E^i_t(s_{\le t_0})$ only depends on its prefix up to and including t. Thus,

$$P_f\left(\cup_{i \in N, t > t_0} E^i_t(s_{\le t_0}) \mid s_{\le t_0}, \mathcal{W}\right) \le 8n^2 \sum_{t > t_0} e^{-\epsilon t} \le \frac{1}{4} \tag{13}$$

Second, similar to above, using (12) instead of (11), for each $i \in N$, each $t > t_0$, and almost every $s_{\le t_0} \in \mathcal{S}_{\le t_0}$,

$$\begin{aligned} P_f\left(D^i_t \cap F^i(s_{\le t_0}) \mid s_{\le t_0}, \mathcal{W}\right) &\le 2P_f\left(D^i_t \cap F^i(s_{\le t_0}) \cap \mathcal{W} \mid s_{\le t_0}\right) \\ &\le 2P_f\left(\left(D^i_t \cap F^i(s_{\le t_0})_{\le t} \cap \mathcal{W}^i\right)_{\le t} \mid s_{\le t_0}\right) \\ &\le 2e^{(r+2\epsilon)t}P_g\left(\left(D^i_t \cap F^i(s_{\le t_0}) \cap \mathcal{W}^i\right)_{\le t} \mid s_{\le t_0}\right) \\ &\le 8e^{-2\epsilon t} \end{aligned}$$

and thus,

$$P_f\left(\cup_{i \in N, t > t_0} D^i_t \cap F^i(s_{\le t_0}) \mid s_{\le t_0}, \mathcal{W}\right) \le 4n \sum_{t > t_0} e^{-2\epsilon t} \le \frac{1}{4} \tag{14}$$

Then, for almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$,

$$\begin{aligned}
1 - P_f\left(H \mid s_{\leq t_0}, \mathcal{W}\right) &\leq P_f\left(\cup_{i\in N, t>t_0} D_t^i \mid s_{\leq t_0}, \mathcal{W}\right) \\
&= P_f\left(\cup_{i\in N, t, t'>t_0} D_t^i \cap E_{t'}^i(s_{\leq t_0}) \mid s_{\leq t_0}, \mathcal{W}\right) \\
&+ P_f\left(\cup_{i\in N, t>t_0} D_t^i \cap F^i(s_{\leq t_0}) \mid s_{\leq t_0}, \mathcal{W}\right) \\
&\leq P_f\left(\cup_{i\in N, t>t_0} E_t^i(s_{\leq t_0}) \mid s_{\leq t_0}, \mathcal{W}\right) \\
&+ P_f\left(\cup_{i\in N, t>t_0} D_t^i \cap F^i(s_{\leq t_0}) \mid s_{\leq t_0}, \mathcal{W}\right) \\
&\leq \frac{1}{2}
\end{aligned}$$

If the infinite history is not H, some agent defects in some period, hence the first step. The second step uses that $\{E_t^i(s_{\leq t_0})\colon t > t_0\} \cup \{F^i(s_{\leq t_0})\}$ is a partition of $\{s^i_{\leq t_0}\} \times S^{\{t_0+1, t_0+2, \dots\}}$. The third step is a basic manipulation. The fourth step follows from (13) and (14). Hence, Step 2, $P_g\left(\mathcal{S}_{\leq t_0}\right) > 0$, and Lemma 5 imply that

$$P_f\left(H\right) \geq \int_{S_{\leq t_0}} P_f\left(H \mid s_{\leq t_0}, \mathcal{W}\right) P_f\left(\mathcal{W} \mid s_{\leq t_0}\right) dP_f\left(s_{\leq t_0}\right) \geq \frac{1}{2} \int_{\mathcal{S}_{\leq t_0}} P_f\left(\mathcal{W} \mid s_{\leq t_0}\right) dP_f\left(s_{\leq t_0}\right) > 0$$

In particular, in state f, the probability that each agent chooses a_g in each period after t_0 is strictly positive, and thus each agent's learning rate is 0. This contradicts (5) and finishes the proof.

□

Theorem 1' (Learning is bounded, weak form)**.** Assume the signals are conditionally independent across agents and periods. For any number of agents n, any network, and any strategies $\sigma^1, \dots, \sigma^n$, some agent learns at rate at most $\tilde{r}_{\text{bdd}} = 2 \min_{\theta,\theta'} \max_{i\in N} \sup_{s\in S} |\ell^i_{\theta,\theta'}(s)|$.

Proof. Let $f, g \in \Omega$ such that $\tilde{r}_{\text{bdd}} = \max_{i\in N} \sup_{s\in S} \ell^i_{g,f}(s)$. Fix strategies $\sigma^1, \dots, \sigma^n$. Assume for contradiction that there is $\epsilon > 0$ such that for each $i \in N$ and each $t \in T$,

$$P\left(a_t^i \neq a_\omega\right) \leq e^{-(r_{\text{bdd}}+2\epsilon)t + o(t)} \tag{15}$$

We write $r = \tilde{r}_{\text{bdd}}$ in the rest of the proof for convenience.

As promised by Lemma 4, there are $t_0 \in T$ and $H \in A^{N\times T}$ such that

(i) $P_g\left(H_{\leq t_0}\right) > 0$,

(ii) for each $i \in N$ and each $t > t_0$, $P_g\left(a_t^i \neq a_g \mid H_{\leq t_0}\right) \leq e^{-(r+\epsilon)t}$,

(iii) $\sum_{t>t_0} e^{-\epsilon t} \leq \frac{1}{4n}$.

(iv) for each $i \in N$ and each $t > t_0$, $H_t^i = a_g$, and

In particular, by (ii) and (iii),

$$P_g\left(H \mid H_{\leq t_0}\right) \geq 1 - \sum_{i\in N, t>t_0} P_g\left(a_t^i \neq a_g \mid H_{\leq t_0}\right) \geq 1 - n \sum_{t>t_0} e^{-(r+\epsilon)t} \geq \frac{1}{2} \tag{16}$$

For each $i \in N$ and each $t > t_0$, let

$$D_t^i = \left\{ s^i \in S^T : \sigma_t^i(s_{\le t}^i; H_{<t}) \neq a_g \right\}$$

contain each infinite trajectory s^i for agent i such that i defects at t for $s_{\le t}^i$, and observe that

$$\begin{aligned} P_g\left(a_t^i \neq a_g \mid H_{\le t_0}\right) &\ge P_g\left(D_t^i, \forall j \neq i, H_{<t}^j \mid H_{\le t_0}\right) \\ &= P_g\left(D_t^i \mid H_{\le t_0}\right) P_g\left(\forall j \neq i, H_{\le t}^j \mid H_{\le t_0}\right) \\ &\ge P_g\left(D_t^i \mid H_{\le t_0}\right) P_g\left(H_{<t} \mid H_{\le t_0}\right) \\ &\ge \frac{1}{2} P_g\left(D_t^i \mid H_{\le t_0}\right) \end{aligned}$$

The first step follows from the definition of D_t^i; the second step uses that $(\mathfrak{s}_t^i)_{i \in N, t \in T}$ are independent conditional on g; the third step is a straightforward estimate; and the fourth step follows from (16). Hence, by (ii),

$$P_g\left(D_t^i \mid H_{\le t_0}\right) \le 2e^{-(r+\epsilon)t} \tag{17}$$

From the conditional independence of signals across agents and periods, (17), and the definition of r, we get

$$\begin{aligned} P_f\left(D_t^i \mid H_{\le t_0}\right) &= P_f\left(D_t^i \mid \mathcal{S}(H_{\le t_0})\right) \\ &= P_f\left(D_t^i \mid \mathcal{S}^i(H_{\le t_0})\right) \\ &\le e^{\frac{1}{2} r(t+t_0)} P_g\left(D_t^i \mid \mathcal{S}^i(H_{\le t_0})\right) \\ &\le e^{\frac{1}{2} r(t+t_0)} P_g\left(D_t^i \mid \mathcal{S}(H_{\le t_0})\right) \\ &\le e^{\frac{1}{2} r(t+t_0)} P_g\left(D_t^i \mid H_{\le t_0}\right) \\ &\le 2e^{-\epsilon t} \end{aligned} \tag{18}$$

Then, by (18), (iii), (i), and Lemma 5,

$$\begin{aligned} P_f(H) &= P_f(H_{\le t_0}) P_f(H \mid H_{\le t_0}) \\ &\ge P_f(H_{\le t_0}) \left(1 - \sum_{i \in N, t > t_0} P_f\left(D_t^i \mid H_{\le t_0}\right)\right) \\ &\ge P_f(H_{\le t_0}) \left(1 - 2n \sum_{t > t_0} e^{-\epsilon t}\right) \\ &\ge \frac{1}{2} P_f(H_{\le t_0}) \\ &> 0 \end{aligned}$$

But this contradicts (15) □

We prove that sufficiently many agents in a strongly connected network can learn faster than a single agent in autarky.

Theorem 2 (Coordination improves learning)**.** *Assume the signals are conditionally independent and identically distributed across agents and periods. For any $\epsilon > 0$, there is n_0 such that for all $n \geq n_0$ and any strongly connected network, there exist strategies $\sigma^1, \ldots, \sigma^n$ such that each agent learns at rate at least $r_{\text{bdd}} - \epsilon$.*

Proof. We assume for now that $N^i = N$ for all $i \in N$ and treat the general case later. For $\theta, \theta' \in \Omega$, define $\ell_{\theta,\theta'} = \ell^1_{\theta,\theta'}$ and $m_{\theta,\theta'} = \mathrm{E}_\theta\left(\ell_{\theta,\theta'}\right)$, and fix $\epsilon > 0$. First, by (1), $\lambda^*_{\theta,\theta'}(-m_{\theta',\theta}) = m_{\theta',\theta}$. Second, since $I^*_{\theta,\theta'}$ contains $m_{\theta',\theta}$ and $\lambda^*_{\theta,\theta'}$ is continuous on $I^*_{\theta,\theta'}$ by Lemma 2 and the remarks thereafter, there is $\delta > 0$ such that

$$\min_{\theta\neq\theta'} \lambda^*_{\theta,\theta'}(-m_{\theta',\theta} + \delta) > \min_{\theta\neq\theta'} \lambda^*_{\theta,\theta'}(-m_{\theta',\theta}) - \epsilon$$

We may further assume that $\delta < \min_{\theta\neq\theta'} m_{\theta,\theta'}$. Hence, when defining

$$\hat{r}_{\text{bdd}} = \min_{\theta\neq\theta'} \lambda^*_{\theta,\theta'}(-m_{\theta',\theta} + \delta)$$

we have $\hat{r}_{\text{bdd}} > \min_{\theta\neq\theta'} m_{\theta,\theta'} = r_{\text{bdd}} - \epsilon$. It thus suffices to exhibit strategies for which the learning rate is at least $\hat{r}_{\text{bdd}}$.

Step 1 (Constructing the strategies)*.* We start by inductively defining the strategies. Let $i \in N$. The strategy σ^i_1 is arbitrary. Now let $t > 1$. Given $a^1_{t-1}, \ldots, a^n_{t-1}$, let $a^{\text{pop}}_{t-1} \in \arg\max_{a\in A} |\{j \in N \colon a^j_{t-1} = a\}|$ be an action that is most popular among the actions taken in period $t-1$. For a history $H^i_{<t} \in A^{N^i\times(t-1)}$ for agent i, let

$$a^i_t = \sigma^i_t(\mathfrak{s}^i_1, \ldots, \mathfrak{s}^i_t; H^i_{<t}) = \begin{cases} a_\theta & \text{if } L^i_{\theta,\theta',t} \geq (m_{\theta,\theta'} - \delta)t \;\forall\theta' \neq \theta, \text{ and} \\ a^{\text{pop}}_{t-1} & \text{otherwise} \end{cases}$$

Note that σ^i is well-defined since $L^i_{\theta,\theta',t} = -L^i_{\theta',\theta,t}$ and $m_{\theta,\theta'} - \delta > 0$. Hence, agents follow their private signals if those are sufficiently decisive and the previous period's most popular action otherwise.

Step 2 (Bounding the probabilities of mistakes)*.* Now we derive the claimed bound on the probability of mistakes.

Case 1*.* First, we consider the probability that agent i makes a mistake if she acts based on her private signals. By Theorem 3 and the remarks after Lemma 2, we have for any two distinct $\theta, \theta' \in \Omega$,

$$P_{\theta'}\left(L^i_{\theta',\theta,t} \leq -(m_{\theta,\theta'} - \delta)t\right) = e^{-\lambda^*_{\theta',\theta}(-m_{\theta,\theta'}+\delta)t+o(t)}$$

Hence, for each $\theta \in \Omega$,

$$\begin{aligned} P\left(a^i_t \neq a_\omega \mid L^i_{\theta,\theta',t} \geq (m_{\theta,\theta'} - \delta)t \;\forall\theta' \neq \theta\right) &\leq \sum_{\theta'\neq\theta} P_{\theta'}\left(L^i_{\theta,\theta',t} \geq (m_{\theta,\theta'} - \delta)t\right) P\left(\omega = \theta'\right) \\ &= \sum_{\theta'\neq\theta} P_{\theta'}\left(L^i_{\theta',\theta,t} \leq -(m_{\theta,\theta'} - \delta)t\right) P\left(\omega = \theta'\right) \\ &= \sum_{\theta'\neq\theta} e^{-\lambda^*_{\theta',\theta}(-m_{\theta,\theta'}+\delta)t+o(t)} P\left(\omega = \theta'\right) \\ &= e^{-\hat{r}_{\text{bdd}}t+o(t)} \end{aligned}$$

which gives the desired bound.

Case 2. Second, we consider the probability that agent i makes a mistake if she follows the previous period's most popular action. We use a standard tail estimate for binomial distributions: if X is binomially distributed with sample size n and success probability q, then

$$P\left(X \leq k\right) \leq e^{-nD(\frac{k}{n} \parallel q)}$$

where

$$D(a \parallel b) = a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}$$

is the Kullback-Leibler divergence of two Bernoulli distributions with success probabilities $a, b \in [0,1]$. Thus, for all $\theta \in \Omega$,

$$\begin{aligned} P_\theta\left(a_{t-1}^{\text{pop}} \neq a_\theta\right) &\leq P_\theta\left(\left|\{i \in N \colon L^i_{\theta,\theta',t-1} \geq (m_{\theta,\theta'} - \delta)(t-1)\ \forall \theta' \neq \theta\}\right| \leq \frac{n}{2}\right) \\ &\leq e^{-nD(\frac{1}{2} \parallel 1 - q_{\theta,t-1})} \end{aligned}$$

where

$$q_{\theta,t} = \sum_{\theta' \neq \theta} e^{-\lambda^*_{\theta,\theta'}(m_{\theta,\theta'} - \delta)t + o(t)} P\left(\omega = \theta'\right) = e^{-\min_{\theta \neq \theta'} \lambda^*_{\theta,\theta'}(m_{\theta,\theta'} - \delta)t + o(t)}$$

is an upper bound for the probability that agent i does not choose a_θ in state θ (obtained from Theorem 3). Hence, for all $\theta \in \Omega$,

$$\begin{aligned} D(\frac{1}{2} \parallel 1 - q_{\theta,t}) &= \frac{1}{2}\left(\log \frac{1}{2(1-q_{\theta,t})} + \log \frac{1}{2q_{\theta,t}}\right) \\ &= \log \frac{1}{2} + \frac{1}{2}\left(\log \frac{1}{1-q_{\theta,t}} + \log \frac{1}{q_{\theta,t}}\right) \\ &= \log \frac{1}{2} + \frac{1}{2}\left(q_{\theta,t} + o(q_{\theta,t}) + \min_{\theta \neq \theta'} \lambda^*_{\theta,\theta'}(m_{\theta,\theta'} - \delta)t + o(t)\right) \\ &= \frac{1}{2} \min_{\theta \neq \theta'} \lambda^*_{\theta,\theta'}(m_{\theta,\theta'} - \delta)t + o(t) \end{aligned}$$

where the third step uses that $\frac{1}{1-x} = 1 + x + o(x)$ and $\log(1 + x + o(x)) = x + o(x)$, where $o(x)$ is a quanitity that converges to 0 faster than x as $x \to 0$. Thus,

$$P_\theta\left(a_{t-1}^{\text{pop}} \neq a_\theta\right) \leq e^{-\frac{n}{2}\min_{\theta \neq \theta'} \lambda^*_{\theta,\theta'}(m_{\theta,\theta'} - \delta)t + o(t)}$$

Since $\lambda^*_{\theta,\theta'}(m_{\theta,\theta'}) = 0$ and $\lambda^*_{\theta,\theta'}$ is strictly decreasing on $(\inf \ell_{\theta,\theta'}, m_{\theta,\theta'}]$ by Lemma 2(ii), $\lambda^*_{\theta,\theta'}(m_{\theta,\theta'} - \delta) > 0$ for all $\theta \neq \theta'$, and so $\min_{\theta \neq \theta'} \lambda^*_{\theta,\theta'}(m_{\theta,\theta'} - \delta) > 0$. So if n is large enough depending on δ, the probability that i makes a mistake in period t by following the most popular action of period $t-1$ decreases at a rate of at least $\hat{r}_{\text{bdd}}$. More precisely, we need that

$$n \geq 2 \frac{\min_{\theta \neq \theta'} \lambda^*_{\theta,\theta'}(-m_{\theta',\theta} + \delta)}{\min_{\theta \neq \theta'} \lambda^*_{\theta,\theta'}(m_{\theta,\theta'} - \delta)}$$

We conclude that if n is large enough depending on δ, then each agent learns at a rate of at least $\hat{r}_{\text{bdd}} = \min_{\theta \neq \theta'} \lambda^*_{\theta,\theta'}(-m_{\theta',\theta} + \delta)$.

Step 3 (Extending to arbitrary networks). It remains to extend the results to arbitrary strongly connected networks. We sketch the argument but omit the details. The main idea is to add periods that are used to propagate the agents' action choices in previous periods through the network.

For two agents i, j, denote by $d(i, j)$ the length of a shortest path from i to j. For example, if $j \in N^i$ and $i \neq j$, then $d(i, j) = 1$.[20] Since the network is strongly connected, $d(i, j)$ is at most $n-1$ for all i, j. We partition the set of periods T into intervals of $M = 1+n(n-2)$ periods. We call the periods $\{1, 1+M, 1+2M, \dots\}$ voting periods, and the remaining periods propagation periods. In each voting period $t \in \{1, 1+M, 1+2M, \dots\}$, each agent i chooses an action similar to the construction in Step 1: if $t = 1$, i chooses an arbitrary action; in any later voting period t, if i's private signals up to period t are sufficiently decisive, she chooses an action optimally based on those, and she follows the most popular action in period $t - M$ otherwise. (The strategies during the propagation periods will ensure that i knows the most popular action in period $t - M$ even though she does not observe all agents' actions directly.) For $j \in N$, in period $t \in \{1+j, 1+j+M, 1+j+2M, \dots\}$, each agent i with $d(i, j) = 1$ imitates j's action in period $t-j$ (i.e., $a_t^i = a_{t-j}^j$), and all other agents repeat their own action in period $t-j$ (i.e., $a_t^i = a_{t-j}^i$). For $j \in N$ and $k \in [n-3]$, in period $t \in \{1+j+kn, 1+j+kn+M, 1+j+kn+2M, \dots\}$, each agent i with $d(i, j) = k+1$ imitates j's action in period $t-j-kn$ (which they observed from some agent with distance k to j in period $t-n$), and all other agents repeat their own action in period $t-j-kn$. Hence, any propagation period t with $t \in \{1+j+kn, 1+j+kn+M, 1+j+kn+2M, \dots\}$ is used to inform agents with distance $k+2$ to j about j's action in the latest voting period by letting an agent with distance $k+1$ to j imitate that action.[21]

Since the agents know the network, they know whether the action of an agent i in any propagation period imitates the action of another agent or agent i's own action in the latest voting period. Thus, in each voting period, each agent knows all other agents' actions in the preceding voting period. By the same arguments as in Step 2, for each $i \in N$ and each voting period $t \in \{1, 1+M, 1+2M, \dots\}$,

$$P\left(a_t^i \neq a_\omega\right) \leq e^{-\hat{r}_{\mathrm{bdd}} t + o(t)}$$

provided that n is large enough. In any propagation period t, each agent i imitates the action

[20] More precisely, $d(i, j)$ is defined inductively by letting $d(i, i) = 0$ and for $k \geq 1$ and $j \neq i$, $d(i, j) = k$ if $\min\{d(i, i') \colon i' \in N, d(i, i') \leq k-1, j \in N^{i'}\} = k-1$.

[21] Extending Theorem 2 to random networks as discussed in Remark 4 requires a minor modification of the strategies: in each period $t \in \{1+j+kn, 1+j+kn+M, 1+j+kn+2M, \dots\}$, each agent who has observed j's action in the latest voting period either directly or indirectly through other agents, imitates j's action in that voting period, and all other agents repeat their own action in that voting period. Then, the set of agents who have (possibly indirectly) observed j's action in the latest voting period grows by at least one agent in period t, unless it already contained all agents before period t. This follows from the assumption that any realization of the network is strongly connected, so that in period t, at least one new agent observes the action of some agent already contained in that set before period t.

of an agent from the latest voting period, and so

$$P\left(a_t^i \neq a_\omega\right) \leq e^{-\hat{r}_{\text{bdd}}(t-M)+o(t)} = e^{-\hat{r}_{\text{bdd}}t+o(t)}$$

since that voting period does not lie more than M periods in the past. So the preceding inequality holds for all periods after replacing the $o(t)$-term.

□

References

I. Arieli, Y. Babichenko, S. Müller, F. Pourbabee, and O. Tamuz. The hazards and benefits of condescension in social learning. *Theoretical Economics*, 20(1):27–56, 2025.

M. Bacharach. Some extensions of a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory*, 37(1):167–190, 1985.

V. Bala and S. Goyal. Learning from neighbors. *Review of Economic Studies*, 65(3):595–621, 1998.

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

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

S. Bikhchandani, D. Hirshleifer, O. Tamuz, and I. Welch. Information cascades and social learning. Technical report, arXiv.org, 2021.

P. Bolton and C. Harris. Strategic experimentation. *Econometrica*, 67(2):349–374, 1999.

H. Cramér. Sur un nouveau théorème-limite de la théorie des probabilités. *In Colloque consacré à la théorie des probabilités*, 736:2–23, 1938.

M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69 (345):118–121, 1974.

A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Applications of Mathematics, 38. Springer, second edition edition, 2009.

D. Gale and S. Kariv. Bayesian learning in social networks. *Games and Economic Behavior*, 45(2):329–346, 2003.

J. D. Geanakoplos and H. M. Polemarchakis. We can't disagree forever. *Journal of Economic Theory*, 28(1):192–200, 1982.

B. Golub and M. O. Jackson. Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–149, 2010.

B. Golub and E. Sadler. Learning in social networks. Technical report, SSRN, 2017.

W. Hann-Caruthers, V. V. Martynov, and O. Tamuz. The speed of sequential asymptotic learning. *Journal of Economic Theory*, 173:383–409, 2018.

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

P. Heidhues, S. Rady, and P. Strack. Strategic experimentation with private payoffs. *Journal of Economic Theory*, 159(Part A):531–551, 2015.

W. Huang. The emergence of fads in a changing world. 2024. Working paper.

W. Huang, P. Strack, and O. Tamuz. Learning in repeated interactions on networks. *Econometrica*, 92(1):1–27, 2024.

G. Keller and S. Rady. Strategic experimentation with Poisson bandits. *Theoretical Economics*, 5(2):275–311, 2010.

G. Keller, S. Rady, and M. Cripps. Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68, 2005.

R. Lévy, M. Pęski, and N. Vieille. Stationary learning in a changing environment. *Econometrica*, 2024. Forthcoming.

P. Molavi, A. Tahbaz-Salehi, and A. Jadbabaie. A theory of non-Bayesian social learning. *Econometrica*, 86(2):445–490, 2018.

G. Moscarini, M. Ottaviani, and L. Smith. Social learning in a changing world. *Economic Theory*, 11:657–665, 1998.

E. Mossel, A. Sly, and O. Tamuz. Asymptotic learning on Bayesian social networks. *Probability Theory and Related Fields*, 158:127–157, 2014.

E. Mossel, A. Sly, and O. Tamuz. Strategic learning and the topology of social networks. *Econometrica*, 83(5):1755–1794, 2015.

R. Parikh and P. Krasucki. Communication, consensus, and knowledge. *Journal of Economic Theory*, 52(1):178–189, 1990.

D. Rosenberg and N. Vieille. On the efficiency of social learning. *Econometrica*, 87(6):2141–2168, 2019.

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

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

X. Vives. How fast do rational agents learn? *Review of Economic Studies*, 60(2):329–347, 1993.