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 independently of the network size and structure and the agents' strategies. This extends recent findings on equilibrium learning by demonstrating that the limitation stems from an inherent tradeoff between optimal action choices and information revelation, rather than strategic considerations. We complement this result by showing that a social planner can design strategies for which each agent learns faster than an isolated individual, provided the network is sufficiently large and strongly connected.

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. Our main result shows that, however, information aggregation fails drastically: the rate of learning of the slowest learning agent is bounded independently of the size and structure of the network and the agents' behavior. This includes learning in equilibrium by rational and forward-looking agents and has important consequences in many 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 social 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 across agents and periods and identically distributed 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 same generic utility function and quantify the learning rate by the asymptotic probability with which agents choose suboptimal actions. Our main result (Theorem 1) shows

[†]`florian.brandl@uni-bonn.de`

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 number of agents who observe each others' signals. This covers recent results on the learning rate in equilibrium with rational agents who are either myopic or geometrically discount future payoffs by Harel, Mossel, Strack, and Tamuz (2021) and Huang, Strack, and Tamuz (2024). Hence, our main insight is that bounds on the learning rate do not rely on the equilibrium constraints on strategies, but come from an inherent tradeoff between optimal action choices and the information contained in these choices.

In networks with many agents, the total amount of private information is vast, so efficient information sharing would lead to fast learning by all agents. The fact that the rate of learning is bounded independently of the network size if agents only observe each others' actions shows that all but a vanishing fraction of the information is lost as the number of agents grows. To see what causes the breakdown of information aggregation, take the perspective of a social planner who can design the agents' strategies to minimize the asymptotic rate at which a suboptimal action is taken. The strategies must trade off two competing objectives: on the one hand, (obviously) 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, we show that either some agent makes mistakes more frequently than fast learning dictates or with positive probability, all agents eventually choose a suboptimal action so that learning breaks down entirely. The first failure mode occurs if some agent's action depends on her private signals too frequently, and, otherwise, suitable signals in the early periods can cause all agents to choose the same wrong action in all future periods. The second case is reminiscent of the information cascades in the herding model of Bikchandani, Hirschleifer, and Welch (1992) and the "rational groupthink" event of Harel et al. (2021), which drives the bound on learning in equilibrium. However, we encounter the additional first failure mode since our agents need not be rational.

We only show that some agents rather than all agents must learn at a bounded rate. Indeed, a large fraction of agents can learn much faster in large networks if they observe the remaining agents' actions and those are used to communicate their private signals. By contrast, Harel et al. (2021) and Huang et al. (2024) show that all agents learn at a bounded rate in any equilibrium. However, an imitation argument shows that all agents learn at the same rate in any equilibrium in networks with an observational path between any two agents, so both types of bounds are equivalent for equilibrium learning.[1] Since our bound on the learning rate 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 to update their beliefs.

[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).

We complement the upper bound on the learning rate by showing that a social planner can design strategies for which all agents learn faster than a single agent in autarky if there is an observational path between any two agents (Theorem 2). 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 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 strongly 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.

To illustrate our results, consider networks with commonly observed actions, binary states, actions, and signals, and each agent's signal in each period coincides with the state with probability 0.75. Our first result shows that independently of the size of the network, some agent learns at a rate that is lower than that of eight agents who share their private signals. On the other hand, the second result exhibits strategies for which all agents learn at almost half the rate of the upper bound in large networks, and thus faster than three agents who observe each others' 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 with publicly observable signals and equilibrium learning. Section 5 states the results and explains the ideas underlying the proofs. Section 6 concludes with a discussion of model variations 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 because agents may choose suboptimal actions today to induce other agents to reveal information tomorrow, and thus 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 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' arguments rely on the assumption that agents are rational, play equilibrium strategies, and are myopic or exponentially discount future payoffs, which we show is 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 others' 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 that agents learn rationally from their private signals and the others' random payoffs but ignore the information contained in their actions, and the agents of Molavi, Tahbaz-Salehi, and Jadbabaie (2018) use a heuristic to combine past beliefs and rationally update the aggregate belief 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 (Bikchandani, Hirschleifer, 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 agents act only once thus informational feedback loops need not be considered. Arieli, Babichenko, Müller, Pourbabee, and Tamuz (2024) 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 Bikchandani, Hirschleifer, 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 others' 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 $N = \{1, \dots, n\}$ be the set of agents and let $T = \{1, 2, \dots\}$ 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 T and $t \in T$, then x_t is its period-t coordinate, $x_{\leq t}$ is the restriction of x to the periods $\{1, \dots, t\}$, and $x_{<t}$ is its restriction to the periods $\{1, \dots, t-1\}$. 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 $\theta, \theta' \in \Omega$.[2] 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

[2] 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 others' utility functions (or at least each others' optimal action in each state).

fine-grained characteristics of the utility functions are only relevant for analyzing equilibria (cf. Section 4).

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. Conditional on each state θ, $\mathfrak{s}_t^i$ has distribution $\mu_\theta \in \Delta(S)$ and signals are independent across agents and periods. We assume that $\mu_\theta, \mu_{\theta'}$ are mutually absolutely continuous and distinct for any two distinct 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 between the states θ and θ' by

$$\ell_{\theta,\theta'}(s) = \log \frac{d\mu_\theta}{d\mu_{\theta'}}(s)$$

We assume that each $\ell_{\theta,\theta'}(s)$ is bounded so that the signals' informativeness about any pair of states is bounded. Let $\ell_{\theta,\theta'} = \ell_{\theta,\theta'}(\mathfrak{s}_1^1)$ for any pair of states θ, θ'. The private signals of Agent i up to any period t induce the private log-likelihood ratio

$$L_{\theta,\theta',t}^i = \log \frac{\pi_0(\theta)}{\pi_0(\theta')} + \sum_{r \leq t} \ell_{\theta,\theta'}(\mathfrak{s}_r^i)$$

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 is strongly connected if there is an observational path from any agent to any other agent, and complete if all neighborhoods comprise all agents.

Agents do not observe each others' 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.[3]

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 $\mathcal{I}_{\leq t}^i = S^t \times A^{|N^i| \times (t-1)}$ is the collection of information sets of Agent i before choosing an action in period t.

[3]Formally, consider the case that Agent i's flow utility for action a in period t is $\tilde{u}(a, \mathfrak{s}_{t+1}^i)$ for an action and state-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 $u(a, \theta) = \mathrm{E}_\theta \left(\tilde{u}(a, \mathfrak{s}_1^i) \right)$, 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).

3.3. Agents' Strategies

A pure strategy for Agent i is a sequence $\sigma^i = (\sigma^i_t)_{t\in T}$ such that $\sigma^i_t \colon \mathcal{I}^i_{\leq t} \to A$ maps each of i's information sets in period t to an action, 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 sequence of action profiles: for each i, $a^i_1 = \sigma^i_1(\mathfrak{s}^i_1)$, and for each $t > 1$, $a^i_t = \sigma^i_t(\mathfrak{s}^i_1, \ldots, \mathfrak{s}^i_t; H^i_{<t})$, where i's public history $H^i_{<t} = (a^i_r)_{i\in N^i, r<t}$ is given by the actions of her neighbors in the periods preceding t. We also write $I^i_{\leq t} = (\mathfrak{s}^i_1, \ldots, \mathfrak{s}^i_t; H^i_{<t})$ for i's information set in period t. The strategy profile is implicit in this notation but will be clear from the context.

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 $e^{-rt+o(t)}$.[4] This definition of the learning rate is common in the literature (see, e.g., Vives, 1993; Hann-Caruthers et al.; Molavi et al., 2018; Rosenberg and Vieille, 2019; Harel et al., 2021; Huang et al., 2024).

Any 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.[5] 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 failing to match and the network is complete. Lastly, signals are binary with $S = \{s_f, s_g\}$, and for some $p \in (\frac{1}{2}, 1)$, μ_f assigns probability p to s_f and μ_g assigns probability p to s_g. We illustrate our results using this example at the end of Section 5.

[4] 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 $\lim_{t\to\infty} \frac{f(t)}{t} = 0$.

[5] 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 $\tilde{S}$ with first marginal μ_θ and the uniform distribution on $[0,1]$ as the second marginal. We may choose the signals' second coordinates so that they are independent of the state and any other signals. Then, the informativeness of the signals does not change and they remain independent across agents and 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.

4. Autarky, Public Signals, and Equilibrium

We revisit three settings as benchmarks: a single agent learning in autarky, several agents observing each others' signals, and equilibrium learning in a network.

If there is only a single agent and this agent chooses actions optimally based on her private signals, it follows from classical large deviations results for random walks that for some $r_{\text{aut}} > 0$,

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

Hence, the limit $\lim_{t\to\infty} -\frac{1}{t}\log P\left(a_t^i \neq a_\omega\right)$ exists and is equal to r_{aut}, which we call the autarky learning rate. In particular, the probability that the agent makes 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. The signals of n agents in a single period provide the same amount of information as those of a single agent over n periods since the signals are independent across agents and periods. Hence n agents observing each others' signals and choosing actions optimally learn n times as fast as a single agent in autarky. Thus, the rate of learning with public signals is nr_{aut}:

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

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.

Third, we turn to learning in equilibrium when the actions of neighbors are observed, but all signals are private. The main insight from previous work on Bayesian learning in networks with sophisticated agents is that information aggregation breaks down in equilibrium. More precisely, suppose that all agents share a common discount rate $\delta \in [0,1)$. 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. A strategy profile is a Nash equilibrium if no agent can increase her expected utility by unilaterally changing her strategy. Standard fixed-point arguments ensure that mixed equilibria always exist. Since mixed strategy profiles can be mapped one-to-one to behaviorally equivalent pure strategy profiles for larger signal spaces (cf. Footnote 5), it suffices to establish a bound on the learning rate for pure strategy equilibria.

The main result of Huang et al. (2024) shows for any number of agents in any strongly connected network, any discount factor, and any utility function, all agents learn at the same rate, and this rate is as most r_{eqm} with[6]

$$r_{\text{eqm}} = 2\sup_{\theta\neq\theta'}\sup_{s\in S}|\ell_{\theta,\theta'}(s)|$$

[6]The model of Huang et al. (2024) differs slightly from ours. They allow the signal space and the distribution

Harel et al. (2021) obtain a tighter bound for two states and complete networks. The main feature of these results is that the learning rate is bounded independently of the number of agents, showing that all but a vanishing fraction of the private information is lost in large networks.

5. Coordinated Learning

The preceding discussion depicts a stark contrast between publicly observable signals and equilibrium behavior: in the first case, learning becomes arbitrarily fast as the number of agents becomes large, whereas in the second case, the rate of learning is bounded independently of the network size and structure. As an intermediate case, we take the perspective of a social planner who can design the agents' strategies and guarantee that all agents follow those strategies, even if they are not in equilibrium, but assume that agents only observe their private signals and their neighbors' actions. Our main result shows that there is a bound on the learning rate of the slowest learning agent that is independent of the number of agents, the network structure, the shared utility function, and the strategy profile imposed by the social planner. Hence, information aggregation breaks down not because of the equilibrium constraints on strategies, but because of a tradeoff between choosing the correct action and using actions to communicate information. More precisely, agents face a tradeoff between choosing the action that is most likely to be correct in the current period and using their action to inform other agents about their private signals to reduce others' probability of mistakes in future periods. The fact that learning is bounded shows that there is no way to meet both objectives at the same time.

Theorem 1 (Learning is bounded)**.** *For any number of agents n and any strategies $\sigma^1, \ldots, \sigma^n$, there is $i \in N$ such that*

$$\liminf_{t\to\infty} -\frac{1}{t} \log P\left(a_t^i \neq a_\omega\right) \leq r_{\text{bdd}}$$

where $r_{\text{bdd}} = \min_{\theta\neq\theta'}\{\mathrm{E}_\theta\left(\ell_{\theta,\theta'}\right) + \mathrm{E}_{\theta'}\left(\ell_{\theta',\theta}\right)\}$.

In other words, for any strategy profile, there is some agent i such that for each $\epsilon > 0$, we have $P\left(a_t^i \neq a_\omega\right) \geq e^{-(r_{\text{bdd}}+\epsilon)t+o(t)}$ 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 signal distributions and not on the network structure or the agents' shared utility function. The constant r_{bdd} is bounded by twice the maximal log-likelihood ratio:

$$r_{\text{bdd}} \leq 2 \sup_{\theta\neq\theta'} \sup_{s\in S} |\ell_{\theta,\theta'}(s)| = r_{\text{eqm}}$$

The right-hand side is the bound on the rate of learning in equilibrium obtained by Huang et al. (2024).[7] An important difference is that their result gives the same bound for all agents, while

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.

[7] Relatedly, Harel et al. (2021) show that for two states f, g and complete networks, all agents learn at a rate of at most $\min\{\mathrm{E}_g\left(\ell_{g,f}\right), \mathrm{E}_f\left(\ell_{f,g}\right)\} \leq r_{\text{bdd}}$ in equilibrium.

Theorem 1 only asserts that some agent's learning rate meets the bound. In fact, it is easy to construct strategies for which a large fraction of agents learn very fast if there are many agents.[8] By contrast, different learning rates cannot be sustained in equilibrium since this would violate the equilibrium condition for slow learning agents who could benefit from imitating one of their neighbors (cf. Footnote 1).

While it is a priori harder to prove a result for all strategy profiles rather than just equilibria, the greater generality makes clear that the argument cannot rely on analyzing the agents' beliefs. Moreover, it is clearly without loss to prove the result for complete networks, which is not obvious for equilibrium learning. We sketch the proof of Theorem 1 for the case of two states f and g and the bound r_{eqm} instead of r_{bdd} to avoid technicalities. Assume then that all agents learn at a rate of at least $r_{\text{eqm}} + \epsilon$ for some positive ϵ. The main argument distinguishes two cases: in the first case, we show that some agent learns at a rate of less than $r_{\text{eqm}} + \epsilon$ in state g; in the second case, we show that eventually all agents choose a_g in all periods with positive probability in state f, and so the rate of learning is 0. Either conclusion contradicts the assumption.

Ignoring an initial clean-up step, we may assume that in state g, there is a non-zero probability that all agents choose a_g in all periods. (This step holds for any positive learning rate.) We say that Agent i defects in period t if all agents chose a_g in all periods before t and i's action is a_f in period t. The first case assumes that there are infinitely many periods t for which in state g the probability that a fixed agent defects (and thus makes a mistake) in period t conditional on no defection by any agent in any earlier period is larger than $e^{-(r_{\text{eqm}}+\frac{\epsilon}{2})t}$. A simple argument using that the probability that no agent ever defects in state g is positive shows that this agent learns at a rate of at most $r_{\text{eqm}} + \frac{\epsilon}{2}$, giving the desired contradiction.

If the first case is not obtained, then, each agent's defection probability in state g and period t conditional on no earlier defection is at most $e^{-(r_{\text{eqm}}+\frac{\epsilon}{2})t}$ for all large enough t. Heuristically, this entails that agents defect only if their private signals are very pessimistic about g being the true state. Thus, even if the state is f, with non-zero probability no agent ever defects, and the learning rate is 0. In more detail, the probability of any sequence of signal realizations for a single agent up to period t changes by a factor of at most $e^{\frac{r_{\text{eqm}}}{2}t}$ if the state is f rather than g (by definition of the log-likelihood ratio). Thus, the probability that Agent i defects in period t conditional on no earlier defections increases by a factor of at most $e^{r_{\text{eqm}}t}$ if the state is f rather than g.[9] So Agent i's defection probability in state f and period t conditional on

[8] Consider the instance in Section 3.4 with the following strategies: in each period, each odd-numbered agent 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 $\frac{n}{2}r_{\text{aut}}$, which grows linearly with the number of agents.

[9] Consider the sequences of Agent i's private signals for which she does not defect before period t as long as no other agent defects, and among those, consider the sequences of private signals for which she defects in period t. The probability ratio of these two sets is the probability that Agent i defects in period t conditional on no earlier defections. Hence, the latter probability increases by a factor of at most $e^{r_{\text{eqm}}t}$ when conditioning on $\omega = f$ rather than $\omega = g$.

no earlier defections is at most $e^{-\frac{\epsilon}{2}t}$ for late enough periods t. Since this sequence is summable and the signals are independent across periods, Agent i receives private signals for which she never defects as long as no other agent defects with positive probability in state f. Lastly, the agents' signals are independent so that with positive probability no agent ever defects even if the state is f, finishing the argument. Note that the independence of the signals across agents and periods is crucial for the proof (cf. Remark 2).

Our second result shows that coordination can improve the rate of learning in networks compared to a single agent learning in autarky. Hence, a social planner can facilitate learning through appropriately designed strategies. This holds for any strongly connected network as long as there are sufficiently many agents. By contrast, it is to the best of our knowledge open whether there is any equilibrium for which the learning rate is larger than r_{aut} (cf. Remark 3).

Theorem 2 (Coordination improves learning)**.** *Assume the network is strongly connected. For any $\epsilon > 0$, there is n_0 such that for all $n \geq n_0$, there exist strategies $\sigma^1, \ldots, \sigma^n$ such that for all $i \in N$,*

$$\liminf_{t\to\infty} -\frac{1}{t} \log P\left(a_t^i \neq a_\omega\right) \geq r_{\text{crd}} - \epsilon$$

where $r_{\text{crd}} = \min_{\theta\neq\theta'} \mathrm{E}_\theta\left(\ell_{\theta,\theta'}\right) > r_{\text{aut}}$.

In other words, for any positive ϵ, there exist strategies so that $P(a_t^i \neq a_\omega) \leq e^{-(r_{\text{crd}}-\epsilon)t+o(t)}$ for each agent i if the number of agents is large enough. The lower bound r_{crd} only depends on the signal distributions and coincides with the upper bound on the equilibrium learning rate obtained by Harel et al. (2021) for two states and complete networks. Thus, in that case, the learning rate for the strategies from Theorem 2 is at least as high as in any equilibrium.

For complete networks, the strategies we construct are easy to describe: each agent follows her private signals as long as those a 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-liklihod 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 be correct if there are many agents. So either case improves upon learning in autarky. The result cannot be improved by increasing the cutoff for following one's private signals above $\mathrm{E}_g\left(\ell_{g,f}\right)t$ since then with high probability, eventually no agent decides based on her private signals and so information aggregation breaks down.

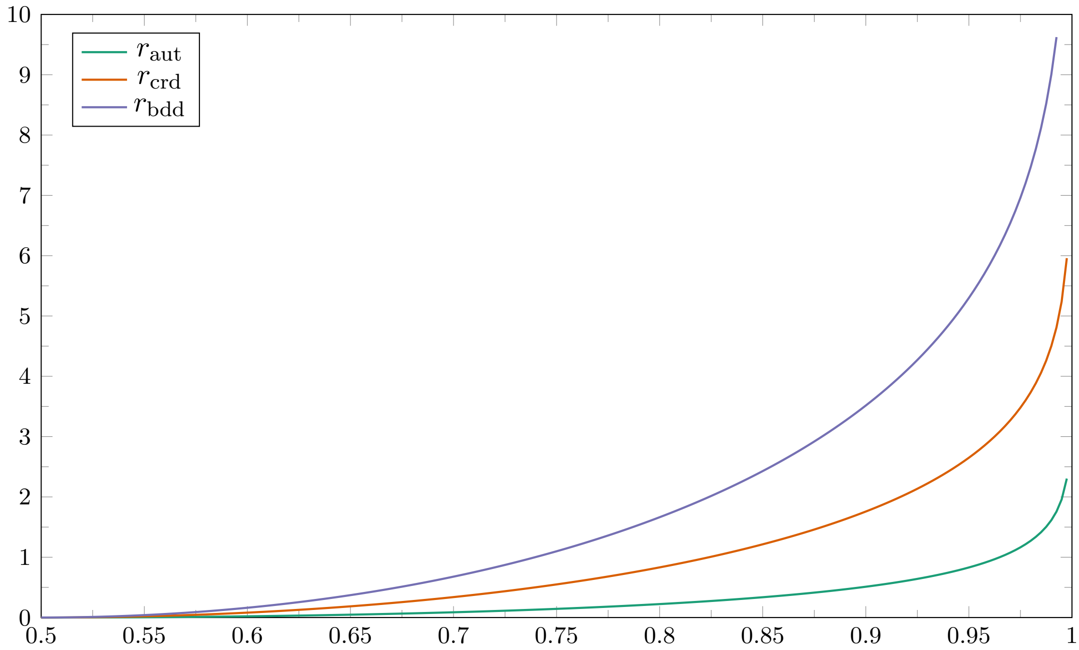

Figure 1: Bounds on the learning rates for the example in Section 3.4. The horizontal axis corresponds to the parameter p determining the informativeness of the signal distributions. Higher values of p correspond to more informative signals. The vertical axis indicates the rate of learning: r_{aut} is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

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 the signal distributions are $\mu_g = (p, 1-p)$ and $\mu_f = (1-p, p)$ for some $p \in (\frac{1}{2}, 1)$, where the two coordinates correspond to the two signals. A calculation shows that

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

Hence, $r_{\text{bdd}} = 2r_{\text{crd}} = 2(2p-1)\log\frac{p}{1-p}$. The expression for r_{aut} involves a minimization problem and cannot be stated in closed form. For $p = 0.75$, we have numerically that $r_{\text{aut}} = 0.14$, $r_{\text{crd}} = 0.55$ and $r_{\text{bdd}} = 1.10$. Figure 1 illustrates the bounds on the learning rates for other values of p.

6. 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.[10]

Remark 2 (Correlated signals). Our results break down emphatically if signals are allowed to be correlated across agents or periods (conditional on the state). With perfect correlation across agents, the private signals of a single agent convey the same information as all agents' signals and no agent in the network can learn faster than a single agent in autarky. Perfect correlation across periods leads to a complete breakdown of learning since periods after the first provide no new information and so the rate of learning is 0.

By contrast, a strong negative 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 whose correlated signal distribution 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 in any network for which any 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 since this incentivizes agents to act based on their private signals.

[10] This equilibrium is reminiscent of a construction by Heidhues et al. (2015), who consider bandit problems where agents observe each others' 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.

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 19 in the proof of Theorem 2.

Acknowledgements

The author 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 $z \in \mathbb{R}$ 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 1 (Dembo and Zeitouni, 2009, Lemma 2.2.5)**.** *Let $I^* = \{\eta \in \mathbb{R} \colon \exists z \in \mathbb{R}, \lambda'(z) = \eta\}$. Then,*[11]

(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 $\mathrm{E}(\ell)$.

Let $\ell_1, \ell_2, \ldots$ be independent random variables with the same ditribution 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 $\mathrm{E}(\ell)$.

[11] Since we assume that ℓ is bounded and non-degenerate, Lemma 1 avoids some case distinctions compared to Lemma 2.2.5 of Dembo and Zeitouni (2009) and allows us to strengthen convexity of λ to strict convexity.

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

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

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

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

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 $\theta, \theta' \in \Omega$, let $\lambda_{\theta,\theta'}$ be the cumulant generating function of $\ell_{\theta,\theta'}$, and denote by $\lambda^*_{\theta,\theta'}$ its Fenchel-Legendre transform. It is not hard to show that $\lambda_{\theta,\theta'}(z) = \lambda_{\theta',\theta}(-(z+1))$ and $\lambda^*_{\theta,\theta'}(\eta) = \lambda^*_{\theta',\theta}(-\eta) - \eta$.[12] Hence, by Lemma 1(ii),

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

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

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

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).[13]

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

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

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

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

Thus,

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

[13] Note that $0 \in I^*_{g,f}$ by the remarks after Lemma 1 and the fact that $\sup \ell_{g,f} > 0$ and $\inf \ell_{g,f} < 0$.

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 1. *The probability that a single agent learning in autarky and choosing actions optimally makes a mistake in period t is*

$$P\left(a_t^1 \neq a_\omega\right) = e^{-r_{\text{aut}}t + o(t)}$$

where

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

Proof. Let

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

where $a_t^1 = \sigma_t^1(\mathfrak{s}_1^1, \ldots, \mathfrak{s}_t^1)$ and the supremum is taken over all strategies $\sigma^1 = (\sigma_1^1, \sigma_2^1, \ldots)$ of a single agent in autarky. Hence, r is the optimal rate of learning, and we have to show that $r = r_{\text{aut}}$.

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

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

Then, there are $f, g \in \Omega$ such that $\liminf_{t\to\infty} -\frac{1}{t}\log P\left(a_t^1 \neq a_\omega\right) > \lambda^*_{g,f}(0)$. Note that σ^1 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

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

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

Step 2 ($r \geq r_{\text{aut}}$). It suffices to find a strategy that achieves a learning rate of at least r_{aut}. Let $\sigma^1 = (\sigma_1^1, \sigma_2^1, \ldots)$ be a strategy with

$$a_t^1 = \sigma_t^1(\mathfrak{s}_1^1, \ldots, \mathfrak{s}_t^1) \in \{a_\theta \in A \colon \theta \in \Omega, L^i_{\theta,\theta',t} \geq 0 \;\forall \theta' \in \Omega\}$$

for all $t \in T$. Thus, σ^1 is a "maximum-likelihood strategy", i.e., it chooses the optimal action for a most probable state.[14] Then, for all $\theta \in \Omega$,

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

[14]Note that σ^1 is well-defined since $L^i_{\theta,\theta',t} + L^i_{\theta',\theta'',t} = L^i_{\theta,\theta'',t}$ ensuring that the right-hand side in the definition of σ^1 is always non-empty.

where the first inequality follows from the definition of σ^1, 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,

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

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.

An auxiliary lemma heuristically shows the following. Fix a (one-dimensional) random walk with i.i.d. increments and a wedge with affine boundaries and sufficiently large intercepts. Then, there is a constant $\alpha > 0$ such that for any point inside the wedge, the probability that the random walk remains inside the wedge for all future periods after hitting that point is at least α.

Lemma 2. *Let $\ell_1, \ell_2, \ldots$ be bounded i.i.d. random variables with $\mathrm{E}\left(\ell_1\right) = 0$, for $t \in T$, let $L_t = \sum_{r\leq t} \ell_r$, and let $L = (L_1, L_2, \ldots)$. Let $a^+, b^+, a^-, b^- > 0$ and let $\mathcal{W} = \{(x_1, x_2, \ldots) \in \mathbb{R}^T : x_t \in [-a^-t - b^-, a^+t + b^+]\ \forall t \in T\}$. Assume that $P\left(\ell_1 \in [0, b^+]\right) > 0$ and $P\left(\ell_1 \in [-b^-, 0]\right) > 0$. Then, there is $\alpha > 0$ such that for all $t \in T$, almost surely,*[15]

$$P\left(L \in \mathcal{W} \mid L_{\leq t}\right) \geq \alpha \mathbf{1}_{\{L_{\leq t}\in\mathcal{W}_{\leq t}\}}$$

We state Lemma 2 with the assumption $\mathrm{E}\left(\ell_1\right) = 0$ to simplify notation but we will apply it without this restriction and the appropriate modifications.

Proof. First, observe that

$$P\left(L \in \mathcal{W}\right) > 0 \tag{3}$$

This follows from the assumption that $P\left(\ell_1 \in [0, b^+]\right) > 0$ and $P\left(\ell_1 \in [-b^-, 0]\right) > 0$ and Theorem 3.

It suffices to show that there are $\alpha > 0$ and $t_0 \in \mathbb{N}$ such that for all $t \in T$, almost surely,

$$P\left(L_{\leq t+t_0} \in \mathcal{W}_{\leq t+t_0}, L_{t+t_0} \in [-a^-(t+t_0), a^+(t+t_0)] \mid L_{\leq t}\right) \geq \alpha \mathbf{1}_{\{L_{\leq t}\in\mathcal{W}_{\leq t}\}} \tag{4}$$

since then the statement follows from (3) and the fact that $(L_{t+t_0} - L_{t_0})_{t\in T}$ and L have the same distribution. Let $\alpha' = \min\{P\left(\ell_1 \in [0, b^+]\right), P\left(\ell_1 \in [-b^-, 0]\right)\}$. By assumption, $\alpha' > 0$. We prove (4) for t_0 such that $a^+t_0 \geq b^+$ and $a^-t_0 \geq b^-$ and $\alpha = (\alpha')^{t_0}$.

[15] For an event E, we denote by $\mathbf{1}_E$ the indicator function of E.

Let $t \in T$, and let $\overline{x} = a^+t + b^+$ and $\underline{x} = a^-t + b^-$. Then, by the choice of α', almost surely,

$$P\left(L_r \in [\underline{x}, \overline{x}] \ \forall r \in \{t+1, \ldots, t+t_0\} \mid L_{\leq t}\right) \geq (\alpha')^{t_0} \mathbf{1}_{\{L_{\leq t} \in \mathcal{W}_{\leq t}\}} = \alpha \mathbf{1}_{\{L_{\leq t} \in \mathcal{W}_{\leq t}\}}$$

Also, note that $\overline{x} = a^+t + b^+ = a^+(t+t_0) - a^+t_0 + b^+ \leq a^+(t+t_0)$ and similarly $\underline{x} \geq -a^-(t+t_0)$. Hence, $L_{t+t_0} \in [\underline{x}, \overline{x}]$ implies that $L_{t+t_0} \in [-a^-(t+t_0), a^+(t+t_0)]$. This proves (4). □

We prove that the rate of learning is bounded independently of the number of agents and their strategies.

Theorem 1 (Learning is bounded)**.** *For any number of agents n and any strategies $\sigma^1, \ldots, \sigma^n$, there is $i \in N$ such that*

$$\liminf_{t \to \infty} -\frac{1}{t} \log P\left(a_t^i \neq a_\omega\right) \leq r_{\text{bdd}}$$

where $r_{\text{bdd}} = \min_{\theta \neq \theta'}\{\mathrm{E}_\theta\left(\ell_{\theta,\theta'}\right) + \mathrm{E}_{\theta'}\left(\ell_{\theta',\theta}\right)\}$.

Proof. Let $n \in \mathbb{N}$ and let $\sigma^1, \ldots, \sigma^n$ be a strategy profile. We may assume without loss of generality that $N^i = N$ for all $i \in N$ since any strategy profile for smaller neighborhoods may be viewed as one for complete neighborhoods. Let $m_{\theta,\theta'} = \mathrm{E}_\theta\left(\ell_{\theta,\theta'}\right)$ for $\theta, \theta' \in \Omega$ and let $f, g \in \Omega$ such that $r_{\text{bdd}} = m_{g,f} + m_{f,g}$. Assume for contradiction that there is $\epsilon > 0$ such that for all $i \in N$, $\liminf_{t\to\infty} -\frac{1}{t} \log P\left(a_t^i \neq a_\omega\right) \geq r_{\text{bdd}} + \epsilon$, or equivalently, for all $i \in N$ and all $t \in T$,

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

That is, assume that all agents learn at a rate of at least $r_{\text{bdd}} + \epsilon$.

We state some preliminary definitions. For $t \in T$ and $s_{\leq t} \in S^t$, let $L_{g,f}(s_{\leq t})$ be the log-likelihood ratio for g over f of an agent who observed private signals $s_{\leq t}$:

$$L_{g,f}(s_{\leq t}) = \log \frac{\pi_0(g)}{\pi_0(f)} + \sum_{r \leq t} \ell_{g,f}(s_r)$$

For $C \geq \sup_{s \in S} |L_{g,f}(s)|$, let $\mathcal{W} = \{s \in S^T \colon -(m_{f,g} + \frac{\epsilon}{4})t - C \leq L_{g,f}(s_{\leq t}) \leq (m_{g,f} + \frac{\epsilon}{4})t + C \ \forall t \in T\}$ be the private signal realizations of an agent inducing log-likelihood ratios for g over f that remain in a wedge with slopes $m_{g,f} + \frac{\epsilon}{4}$ and $-(m_{f,g} + \frac{\epsilon}{4})$ and intercepts C. Then, $\mathcal{W}^n$ is the corresponding set of signal realization profiles. (For proving a weaker version of Theorem 1 with r_{bdd} replaced by $r_{\text{eqm}} = 2\sup_{s \in S} |\ell_{g,f}(s)|$, one can set $\mathcal{W} = S^T$. The required adjustments in the rest of the proof are deleting all occurrences of $\mathcal{W}$ and replacing $m_{f,g}$ and $m_{g,f}$ by $\sup_{s \in S} |\ell_{g,f}(s)|$.)

Step 1. We construct signal realizations for a finite number of initial periods so that, conditional on $\omega = g$ and these signal realizations, all agents choose a_g in all subsequent periods with high probability. Let $0 < \delta < \frac{1}{2}$. By (5) and the fact that $P(\omega = g) = \pi_0(g) > 0$ (since π_0 has full support), there is $t_0 \in T$ such that

$$P_g\left(a_t^i = a_g \ \forall i \in N, \forall t > t_0\right) \geq 1 - \delta$$

Hence, there is a public history $H_{\leq t_0} \in A^{n\times t_0}$ up to period t_0 with $P_g(H_{\leq t_0}) > 0$, such that

$$P_g\left(a_t^i = a_g \ \forall i \in N, \forall t > t_0 \mid H_{\leq t_0}\right) \geq 1-\delta \tag{6}$$

By Theorem 3, C in the definition of $\mathcal{W}$ can be chosen large enough such that for $\theta \in \{f, g\}$,[16]

$$P_\theta\left(\mathcal{W}^n \mid H_{\leq t_0}\right) \geq 1-\delta \tag{7}$$

In words, conditional on either state and the public history $H_{\leq t_0}$, the probability that each agent's sequence of private signal realizations is in the wedge $\mathcal{W}$ is at least $1-\delta$.

Denote by $H \in A^{n\times T}$ the infinite public history that coincides with $H_{\leq t_0}$ in the first t_0 periods and all agents choose a_g in all periods after t_0. For $i \in N$ and $t \geq t_0$, let $\mathcal{S}^i_{\leq t}$ be the set of signal realizations for Agent i up to period t that are consistent with the public history $H_{<t}$ in the first $t-1$ periods:

$$\mathcal{S}^i_{\leq t} = \{s_{\leq t} \in S^t \colon \sigma^i_r(s_{\leq r}; H_{<r}) = H^i_r \ \forall r < t\}$$

Let $\hat{\mathcal{S}}^i_{\leq t} = \mathcal{S}^i_{\leq t} \cap \mathcal{W}_{\leq t}$, and let $\hat{\mathcal{S}}_{\leq t} = \hat{\mathcal{S}}^1_{\leq t} \times \cdots \times \hat{\mathcal{S}}^n_{\leq t}$. By (6) and (7), we have that

$$P_g\left(\hat{\mathcal{S}}_{\leq t} \ \forall t > t_0 \mid H_{\leq t_0}\right) = P_g\left(H, \mathcal{W}^n \mid H_{\leq t_0}\right) \geq 1-2\delta > 0 \tag{8}$$

In words, conditional on $\omega = g$ and the public history $H_{\leq t_0}$ in the first t_0 periods, with probability at least $1-2\delta$, each agent observes signal realizations for which she chooses a_g in all periods after t_0 and which remain inside the wedge $\mathcal{W}$. Lastly, for $i \in N$, $t > t_0$, and $a \in A$, let $\hat{\mathcal{S}}^i_{\leq t, \neq a} \subset \hat{\mathcal{S}}^i_{\leq t}$ be those signal realizations in $\hat{\mathcal{S}}^i_{\leq t}$ for which i's action in period t after observing $H_{<t}$ is different from a:

$$\hat{\mathcal{S}}^i_{\leq t, \neq a} = \{s_{\leq t} \in \hat{\mathcal{S}}^i_{\leq t} \colon \sigma^i_t(s_{\leq t}; H_{<t}) \neq a\}$$

Step 2. We distinguish two cases. In the first case, we show that some agent makes mistakes too frequently if $\omega = g$. In the second case, we show that the probability that all agents choose a_g indefinitely from some period onward is nonzero if $\omega = f$.

Case 1. There exist $i \in N$ and $t_0 < t_1 < t_2 < \dots$ such that for all $k \in \mathbb{N}$,

$$P_g\left(a^i_{t_k} \neq a_g \mid \hat{\mathcal{S}}_{\leq t_k}\right) \geq e^{-(r_{\text{bdd}} + \frac{3}{4}\epsilon)t_k}$$

In words, there are infinitely many periods t_k such that conditional on $\omega = g$ and signal realizations in $\hat{\mathcal{S}}_{\leq t_k}$, Agent i's probability of mistakes exceeds the rate $r_{\text{bdd}} + \frac{3}{4}\epsilon$. By (8) and the fact that $P_g(H_{\leq t_0}) > 0$, for all $k \in \mathbb{N}$,

$$\begin{aligned} P_g\left(a^i_{t_k} \neq a_g\right) &\geq P_g\left(a^i_{t_k} \neq a_g \mid \hat{\mathcal{S}}_{\leq t_k}\right) P_g\left(\hat{\mathcal{S}}_{\leq t_k} \mid H_{\leq t_0}\right) P_g\left(H_{\leq t_0}\right) \\ &\geq e^{-(r_{\text{bdd}} + \frac{3}{4}\epsilon)t_k}(1-2\delta) P_g\left(H_{\leq t_0}\right) \end{aligned}$$

But then Agent i learns at a rate of at most $r_{\text{bdd}} + \frac{3}{4}\epsilon$, which contradicts (5).

[16] We use the shorthand $P_\theta\left(\mathcal{W}^n \mid H_{\leq t_0}\right) = P_\theta\left((\mathfrak{s}^i_1, \mathfrak{s}^i_2, \dots) \in \mathcal{W} \ \forall i \in N \mid H_{\leq t_0}\right)$ and similarly in the rest of the proof. Also, if $\{\mathcal{S}^\xi\}_{\xi \in \Xi}$ is a collection of sets of signal realizations, we write $P\left(\mathcal{S}^\xi \ \forall \xi \in \Xi\right)$ for the probability of the event that the signal realizations are contained in each of the $\mathcal{S}^\xi$.

Case 2. Assume that we are not in Case 1. Carefully negating all quantifiers, it follows that for all $i \in N$, there is $t^i \geq t_0$ such that for all $t > t^i$,

$$P_g\left(a_t^i \neq a_g \mid \hat{\mathcal{S}}_{\leq t}\right) \leq e^{-(r_{\text{bdd}}+\frac{3}{4}\epsilon)t}$$

In words, conditional on $\omega = g$ and signal realizations in $\hat{\mathcal{S}}_{\leq t}$, Agent i's probability of mistakes decreases exponentially in t at a rate of at least $r_{\text{bdd}} + \frac{3}{4}\epsilon$ for sufficiently late periods. Let $\tilde{t}_0 = \max\{t^i : i \in N\}$. Then, for all $i \in N$ and $t > \tilde{t}_0$,

$$P_g\left(a_t^i \neq a_g \mid \hat{\mathcal{S}}_{\leq t}\right) \leq e^{-(r_{\text{bdd}}+\frac{3}{4}\epsilon)t} \tag{9}$$

We want to show that there is $\bar{t}_0 \geq \tilde{t}_0$ such that

$$P_f\left(a_t^i = a_g \;\forall i \in N, \forall t \geq \bar{t}_0\right) > 0$$

Together with the fact that $P(\omega = f) = \pi_0(f) > 0$, this would give the desired contradiction to (1). For each $t > \tilde{t}_0$, we have $P_f\left(H_{<t}, \mathcal{W}^n_{<t}\right) > 0$ by (8) and the fact that $P_f\left(H_{\leq t_0}\right) > 0$ (since $P_g\left(H_{\leq t_0}\right) > 0$), and so $P_f\left(H_{<t}, \mathcal{W}^n\right) > 0$ by Lemma 2. Moreover, for each $t > \tilde{t}_0$,

$$P_f\left(a_r^i = a_g \;\forall i \in N, \forall r \geq t\right) \geq P_f(H) \geq P_f\left(H \mid H_{<t}, \mathcal{W}^n\right) P_f\left(H_{<t}, \mathcal{W}^n\right)$$

Hence, it suffices to show that there is $\bar{t}_0 \geq \tilde{t}_0$ such that

$$P_f\left(H \mid H_{<\bar{t}_0}, \mathcal{W}^n\right) > 0 \tag{10}$$

For all $i \in N$ and $t > \tilde{t}_0$,

$$P_g\left(\hat{\mathcal{S}}^i_{\leq t, \neq a_g} \mid \hat{\mathcal{S}}^i_{\leq t}\right) = P_g\left(\hat{\mathcal{S}}^i_{\leq t, \neq a_g} \mid \hat{\mathcal{S}}_{\leq t}\right) = P_g\left(a_t^i \neq a_g \mid \hat{\mathcal{S}}_{\leq t}\right) \leq e^{-(r_{\text{bdd}}+\frac{3}{4}\epsilon)t} \tag{11}$$

where the first equality uses that signals are independent across agents, the second equality uses that signal realizations in $\hat{\mathcal{S}}_{\leq t}$ induce the public history $H_{<t}$, and the inequality follows from (9). By Lemma 2, there is $\alpha > 0$ such that for all $i \in N$ and $t \in T$, $P_f\left(\mathcal{W} \mid \hat{\mathcal{S}}^i_{\leq t}\right) \geq \alpha$. Then, for all $i \in N$ and $t > \tilde{t}_0$, we have that

$$\begin{aligned} P_f\left(\hat{\mathcal{S}}^i_{\leq t, \neq a_g} \mid \hat{\mathcal{S}}^i_{\leq t}, \mathcal{W}\right) &= \frac{P_f\left(\hat{\mathcal{S}}^i_{\leq t, \neq a_g}, \mathcal{W}\right)}{P_f\left(\hat{\mathcal{S}}^i_{\leq t}, \mathcal{W}\right)} \leq \frac{P_f\left(\hat{\mathcal{S}}^i_{\leq t, \neq a_g}\right)}{P_f\left(\mathcal{W} \mid \hat{\mathcal{S}}^i_{\leq t}\right) P_f\left(\hat{\mathcal{S}}^i_{\leq t}\right)} \\ &\leq \alpha^{-1} \frac{P_f\left(\hat{\mathcal{S}}^i_{\leq t, \neq a_g}\right)}{P_f\left(\hat{\mathcal{S}}^i_{\leq t}\right)} \leq \alpha^{-1} \frac{P_g\left(\hat{\mathcal{S}}^i_{\leq t, \neq a_g}\right) e^{(m_{f,g}+\frac{\epsilon}{4})t}}{P_g\left(\hat{\mathcal{S}}^i_{\leq t}\right) e^{-(m_{g,f}+\frac{\epsilon}{4})t}} \\ &= \alpha^{-1} P_g\left(\hat{\mathcal{S}}^i_{\leq t, \neq a_g} \mid \hat{\mathcal{S}}^i_{\leq t}\right) e^{(r_{\text{bdd}}+\frac{\epsilon}{2})t} \leq \alpha^{-1} e^{-\frac{\epsilon}{4}t} \end{aligned} \tag{12}$$

The first and the last equality use that $\hat{\mathcal{S}}^i_{\leq t, \neq a_g} \subset \hat{\mathcal{S}}^i_{\leq t}$; the first inequality is a standard manipulation; the second inequality uses the preceding lower bound for $P_f\left(\mathcal{W} \mid \hat{\mathcal{S}}^i_{\leq t}\right)$; the third inequality uses that $\hat{\mathcal{S}}^i_{\leq t} \subset \mathcal{W}_{\leq t}$ and the definition of $\mathcal{W}_{\leq t}$; and the last inequality follows from (11). Let $\bar{t}_0 \geq \tilde{t}_0$ such that $\alpha^{-1} \sum_{t \geq \bar{t}_0} e^{-\frac{\epsilon}{4}t} < 1$. Then,

$$
\begin{aligned}
P_f\left(H \mid H_{<\bar{t}_0}, \mathcal{W}^n\right) &= \prod_{t \geq \bar{t}_0} P_f\left(H_{\leq t} \mid H_{<t}, \mathcal{W}^n\right) \\
&= \prod_{t \geq \bar{t}_0} \prod_{i \in N} \left(1 - P_f\left(a^i_t \neq a_g \mid H_{<t}, \mathcal{W}^n\right)\right) \\
&= \prod_{i \in N} \prod_{t \geq \bar{t}_0} \left(1 - P_f\left(\hat{\mathcal{S}}^i_{\leq t, \neq a_g} \mid \hat{\mathcal{S}}_{\leq t}, \mathcal{W}^n\right)\right) \\
&= \prod_{i \in N} \prod_{t \geq \bar{t}_0} \left(1 - P_f\left(\hat{\mathcal{S}}^i_{\leq t, \neq a_g} \mid \hat{\mathcal{S}}^i_{\leq t}, \mathcal{W}\right)\right) \\
&\geq \prod_{i \in N} \left(1 - \sum_{t \geq \bar{t}_0} P_f\left(\hat{\mathcal{S}}^i_{\leq t, \neq a_g} \mid \hat{\mathcal{S}}^i_{\leq t}, \mathcal{W}\right)\right) \\
&\geq \prod_{i \in N} \left(1 - \alpha^{-1} \sum_{t \geq \bar{t}_0} e^{-\frac{\epsilon}{4}t}\right) > 0
\end{aligned}
$$

The first equality follows from the chain rule for conditional probability and the fact that $H_{<t}$ is a prefix of $H_{\leq t} = H_{<t+1}$; the second and fourth equality use that the signals are independent across agents; the third equality exchanges the products and uses that the events $\{H_{<t}, \mathcal{W}^n\}$ and $\{\hat{\mathcal{S}}_{\leq t}, \mathcal{W}^n\}$ are the same;[17] the first inequality is a standard estimate; and the second inequality follows from (12) and the choice of $\bar{t}_0$. This proves (10) and thus gives the desired contradiction.

□

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 network is strongly connected. For any $\epsilon > 0$, there is n_0 such that for all $n \geq n_0$, there exist strategies $\sigma^1, \ldots, \sigma^n$ such that for all $i \in N$,*

$$
\liminf_{t \to \infty} -\frac{1}{t} \log P\left(a^i_t \neq a_\omega\right) \geq r_{\text{crd}} - \epsilon
$$

where $r_{\text{crd}} = \min_{\theta \neq \theta'} \mathrm{E}_\theta\left(\ell_{\theta,\theta'}\right) > r_{\text{aut}}$.

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

$$
\tilde{r}_{\text{crd}} := \min_{\theta \neq \theta'} \lambda^*_{\theta,\theta'}(-m_{\theta',\theta} + \delta) > \min_{\theta \neq \theta'} \lambda^*_{\theta,\theta'}(-m_{\theta',\theta}) - \epsilon
$$

We may further assume that $\delta < \min_{\theta \neq \theta'} m_{\theta,\theta'}$. Using (1), we get

$$
\tilde{r}_{\text{crd}} > \min_{\theta \neq \theta'} m_{\theta,\theta'} - \epsilon = r_{\text{crd}} - \epsilon
$$

Hence, it suffices to exhibit strategies for which the learning rate is at least $\tilde{r}_{\text{crd}}$.

[17]Exchaning the products in harmless since one of them is a finite product.

Step 1 (Constructing the strategies). We start by inductively defining the strategies. Let $i \in N$. The strategy σ_1^i is arbitrary. Now let $t > 1$. Given $a_{t-1}^1, \dots, a_{t-1}^n$, let $a_{t-1}^{\text{pop}} \in \arg\max_{a \in A} |\{j \in N \colon a_{t-1}^j = a\}|$ be an action that is most popular among the actions taken in period $t-1$. For a public history $H_{<t}^i$ for Agent i, let

$$a_t^i = \sigma_t^i(\mathfrak{s}_1^i, \dots, \mathfrak{s}_t^i; H_{<t}^i) = \begin{cases} a_\theta & \text{if } L_{\theta,\theta',t}^i \geq (m_{\theta,\theta'} - \delta)t \ \forall \theta' \neq \theta, \text{ and} \\ a_{t-1}^{\text{pop}} & \text{otherwise} \end{cases}$$

Note that σ^i is well-defined since $m_{\theta,\theta'} - \delta > 0$. 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 1, we have for any two distinct $\theta, \theta' \in \Omega$,

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

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

$$\begin{aligned} P\left(a_t^i \neq a_\omega \mid L_{\theta,\theta',t}^i \geq (m_{\theta,\theta'} - \delta)t \ \forall \theta' \neq \theta\right) &\leq \sum_{\theta' \neq \theta} P_{\theta'}\left(L_{\theta,\theta',t}^i \geq (m_{\theta,\theta'} - \delta)t\right) P\left(\omega = \theta'\right) \\ &= \sum_{\theta' \neq \theta} P_{\theta'}\left(L_{\theta',\theta,t}^i \leq -(m_{\theta,\theta'} - \delta)t\right) P\left(\omega = \theta'\right) \\ &= \sum_{\theta' \neq \theta} e^{-\lambda^*_{\theta',\theta}(-m_{\theta,\theta'}+\delta)t + o(t)} P\left(\omega = \theta'\right) \\ &= e^{-\tilde{r}_{\text{crd}} t + o(t)} \end{aligned}$$

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 $\theta \in \Omega$,

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

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 equality uses that $\frac{1}{1-x} = 1 + x + o(x)$ and $\log(1 + x + o(x)) = x + o(x)$ for $x \to 0$. Thus,

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

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 1(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 $\tilde{r}_{\text{crd}}$. 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 $\tilde{r}_{\text{crd}} = \min_{\theta\neq\theta'}\lambda^*_{\theta,\theta'}(-m_{\theta',\theta}+\delta)$. Since $\lambda^*_{\theta,\theta'}$ is strictly decreasing on $(\inf \ell_{\theta,\theta'}, m_{\theta,\theta'}]$ and $m_{\theta',\theta} > 0$ for any two distinct $\theta, \theta' \in \Omega$, it follows that $\tilde{r}_{\text{crd}} > r_{\text{aut}} = \min_{\theta\neq\theta'}\lambda^*_{\theta,\theta'}(0)$.

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 $d(i, j) = 1$.[18] 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; otherwise, if i's private signals up to period t are sufficiently decisive, she chooses an action optimally

[18] 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 $\min\{d(i, i') \colon j \in N^{i'}\} = k - 1$.

based on those, and she follows the most popular 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 $t = 1 + j + kn \mod M$ 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.[19]

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 $t \in \{1, 1+M, 1+2M, \dots\}$,

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

provided that n is large enough. In any propagation period t, each agent i imitates the action of an agent from the latest voting period, and so

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

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. Technical report, arXiv.org, 2024.

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

[19]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.

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. Bikchandani, D. Hirschleifer, 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. Bikchandani, D. Hirschleifer, 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.

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.