

The Social Learning Barrier

Florian Brandl[†]

Institute for Microeconomics, University of Bonn

We consider long-lived agents who interact repeatedly in a social network. In each period, each agent observes a private signal, chooses an action, and then observes the actions of her neighbors. We show that the learning rate of the slowest-learning agent is capped by a constant that depends only on the distribution of the most informative private signal, but not on the number of agents, the network structure, correlations between the private signals, or 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 a broad class of conditions. The result is driven by an information-theoretic tradeoff between optimal action choices and information revelation.

Keywords: Social learning, social networks, Bayesian learning.

1. Introduction

Do large social networks lead to fast learning when people repeatedly watch one another's actions? The pool of private information grows with the network size, thus learning should accelerate as networks grow. This simple intuition turns out to be false. We identify a social learning barrier: under a broad class of conditions, the rate at which the slowest-learning agent learns is capped by a constant that depends only on the distribution of the most informative private signal. The cap does not depend on how many agents there are, how they are connected, whether signals are correlated across agents, or how agents behave. The barrier is thus not driven by a lack of links in the network or by incentives, but by a fundamental information-theoretic tradeoff. Observational learning in social networks fits economic applications such as product choice, technology adoption,

[†]florian.brandl@uni-bonn.de

and opinion formation. Designers of such networks can structure who observes whom or recommend actions. Our results show that attempts to accelerate learning through such interventions face a hard limit.

In our model, long-lived agents interact with one another in a network over infinitely many periods. The state of the world is fixed but unknown. In each period, each agent receives a private signal about the state, chooses an action, and then observes the actions of her neighbors. An agent’s flow utility in a period depends on her action and the state, but not on the actions of other agents, and is unobserved. We quantify learning by how fast the probability of suboptimal action choices vanishes.

In large networks, the total amount of private information is vast. Under fully efficient information sharing, all agents would surpass any desired learning rate as the network size grows. However, our main result (Theorem 1) shows that information aggregation fails: there exists a constant such that for any number of agents, any network structure, and any strategies, some agent learns no faster than this constant. The constant depends only on the distribution of the most informative private signal. Thus, even a social planner who can dictate the agents’ strategies cannot achieve uniformly fast learning. For example, consider agents who aim to match a binary action to a binary state, and in each period, receive conditionally independent binary signals that equal the state with probability 75%. Then, among 100 agents who observe each other’s actions, there is always some agent who learns more slowly than each one of 5 agents who observe each other’s signals, regardless of how the 100 agents behave. This means that observing others’ actions aggregates a negligible fraction of the private information in the network at best.

To see the mechanism behind this result, take the perspective of a social planner who can design the agents’ strategies.¹ The planner’s goal is to maximize the rate at which the probability of suboptimal action choices vanishes. In our model, this corresponds to maximizing the learning rate of the slowest-learning agent. The strategies must trade off two competing objectives: on the one hand, the probability that an agent’s action is suboptimal must vanish rapidly; on the other hand, an agent’s action choice must contain information about her private signals. To achieve significantly faster learning than a single, isolated agent, the first objective requires that an agent’s action rarely depends on her private signals. But this is incompatible with the second objective, which requires the opposite. In fact, strategies that depend on the agents’ private signals too infrequently result in information cascades similar to the “rational herds” of [Bikhchandani](#),

¹For example, in online social networks, the platform can significantly influence the users’ behavior, similar to our imaginary social planner, by curating the information shown to users and making personalized recommendations.

Hirshleifer, and Welch (1992) or the “rational groupthink” event of Harel, Mossel, Strack, and Tamuz (2021).

However, the above tradeoff is not absolute. We show that a social planner can design strategies for which each agent learns faster than an isolated agent. This learning rate approaches the universal upper bound as networks grow in size (Theorem 2).² For complete networks, the strategies prescribed by the planner are simple: each agent follows her past private signals if those decisively favor one of the states, and otherwise she copies the action that was most popular among all agents in the previous period. This allows all agents to learn faster than an isolated agent since most agents’ actions are informative about their private signals in most periods.

We then turn to equilibrium learning by rational agents who geometrically discount future payoffs. An imitation argument of Huang et al. (2024) shows that all agents learn at the same rate in any equilibrium and any strongly connected network.³ Hence, there is no difference between a bound for the slowest-learning agent and a bound for all agents. 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 for any number of agents in any strongly connected network—under a broader class of conditions.⁴ More importantly, the main new conceptual insight is that the bound on the learning rate arises not from equilibrium strategy restrictions, but from an information-theoretic tradeoff. Since our bound applies to any strategies, it also holds for equilibrium learning under different behavioral and informational assumptions. Examples include misspecified agents and settings where a social planner moderates the information agents receive about other agents’ actions.

Finally, we demonstrate that the failure of information aggregation persists even when agents observe their neighbors’ signals in addition to their actions. In complete networks, all information is then public, and rational agents learn at the first-best rate. However, a common feature of social networks is that each agent has a small number of neighbors compared to the network size. As a corollary of our main result, we show that the size of the largest neighborhood controls the equilibrium learning rate. That is, in any equilibrium for any number of agents in any strongly connected network, each agent’s learning

²Theorem 2 requires that the signals are conditionally independent and identically distributed across periods and agents and that the network is strongly connected (i.e., there is a directed path from any agent to any other agent).

³See 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).

⁴For example, we do not assume that the set of actions is finite, that the signals are conditionally independent across agents, or that each signal has bounded informativeness.

rate is capped by a constant that only depends on the size of the largest neighborhood and the distribution of the most informative private signal.⁵

These results lead to four main insights. First, observational learning in networks fails to efficiently aggregate information, even when a social planner designs the agents’ strategies. Second, a small amount of information is aggregated: suitably designed strategies allow each agent to learn faster than an isolated agent. Third, limits to the equilibrium learning rate are not primarily driven by incentives. And fourth, even observing neighbors’ signals improves learning only immaterially in networks with small neighborhoods.

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 outlines the proofs of the main results. Section 8 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 sequentially arriving agents.

2.1. Equilibrium Learning in Networks

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. In recent work, [Huang et al. \(2024\)](#) show that in the same model as in the present paper, the rate of learning of any agent in any equilibrium is bounded independently of the number of agents in strongly connected networks. This result follows from the following elegant argument: if an agent’s neighbor learns much faster than in autarky, she will ignore her private signals and imitate her neighbor, so that her actions cease to reveal information about her private signals; in a strongly connected network, this reasoning extends inductively to all agents; but then all agents eventually ignore their private signals, learning stops, and a contradiction is reached.

⁵For this result, we assume that the signals are conditionally independent across periods and agents, and agents are rational and geometrically discount future payoffs.

Harel et al. (2021) obtain a similar conclusion in a restricted setting with myopic agents and networks in which each agent observes all other agents’ actions. They use large deviations theory to derive their result, and their work is methodologically closer to ours. Both papers rely on agents being fully rational, playing equilibrium strategies, and either geometrically discounting future payoffs or being myopic—assumptions that our results show are not necessary.

2.2. Non-Bayesian Learning

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 an 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.

2.3. Non-Recurring 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 in 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 it 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 observed market prices. In contrast to our model, agents do not receive private signals in later periods.

2.4. Sequentially Arriving Agents

In the classical herding model (Bikhchandani, Hirshleifer, and Welch, 1992; Banerjee, 1992; Smith and Sorensen, 2000), agents arrive sequentially, observe a private signal

as well as their predecessors' actions, and take an action once. 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. First, since each agent acts only once, informational feedback loops need not be considered. Second, each agent receives only one private signal, so that rational agents may fail to learn the state in the limit.

Rosenberg and Vieille (2017), in an unpublished draft of a later paper (Rosenberg and Vieille, 2019), consider a social planner who can dictate the agents' strategies. They demonstrate that the planner can achieve efficient learning (in the sense that the expected number of agents failing to match the state is finite), even if the distribution of private signals is not informative enough for efficient learning in an equilibrium with rational agents. Hence, the inefficiency of equilibrium learning arises from strategic incentives rather than a more basic, information-theoretic tradeoff, which contrasts with our results. Relatedly, Arieli, Babichenko, Müller, Pourbabee, and Tamuz (2025) consider condescending agents who underestimate the quality of the other agents' private signals. This misspecification decreases the probability of herding on the wrong action and improves learning compared to the well-specified case if condescension is mild. Thus, they recover the result of Rosenberg and Vieille (2017) since a social planner could instruct agents to behave as if they were condescending. Harel et al. (2021) conjecture that the same misspecification improves learning in the present model as well. Our results show that its benefit can be small at best. Smith et al. (2021) study a social planner who aims to maximize the discounted social welfare. These works are similar to our approach in spirit.

In a variant of the herding model, the state changes stochastically over time (Moscarini, Ottaviani, and Smith, 1998; Lévy, Peški, 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 sequentially arriving agents.

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 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 the t -th coordinate, $x_{\leq t}$ is its restriction to the periods $\{1, \dots, t\}$, and $x_{< t}$ is its restriction to the periods $\{1, \dots, t-1\}$. There is an unknown state ω , taking values in a finite set Ω . Generic elements of Ω are denoted by

f, g . We denote by $\pi_0 \in \Delta(\Omega)$ the distribution of ω , and assume that it has full support. For a state $f \in \Omega$, we write $\mathbb{P}_f(\cdot) = \mathbb{P}(\cdot | f)$ and $\mathbb{E}_f(\cdot) = \mathbb{E}(\cdot | f)$ for the corresponding conditional probability and conditional expectation.

3.1. Agents' Payoffs

All agents have the same utility function $u: A \times \Omega \rightarrow \mathbb{R}$ 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_f for every state $f \in \Omega$ exists and is unique.

$$\{a_f\} = \arg \max_{a \in A} u(a, f)$$

We also assume that $a_f \neq a_g$ for any two distinct states $f, g \in \Omega$.⁶ 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 precludes agents from encoding information in the selection of the optimal action. 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. On the other hand, a single action choice can contain an arbitrary amount of information since A can be infinite, which makes our result establishing an upper bound on the learning rate stronger. 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 only become relevant when 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 f , \mathfrak{s}_t^i has distribution $\mu_f^i \in \Delta(S)$ and $\mathfrak{s}_t = (\mathfrak{s}_t^i)_{i \in N}$ has distribution $\mu_f \in \Delta(S^N)$ for each $t \in T$, 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 μ_f, μ_g are mutually absolutely continuous for any two states f, g ,

⁶The 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).

so that no signal excludes any state with certainty. Observing the signal realization $s \in S$ changes the log-likelihood ratio of the observing agent i between the states f and g by

$$\ell_{f,g}^i(s) = \log \frac{d\mu_f^i}{d\mu_g^i}(s)$$

Let $\ell_{f,g}^i = \ell_{f,g}^i(\mathbf{s}_1^i)$ for any pair of states f, g and any agent i . We assume that $\ell_{f,g}^i$ is not identically zero and that its cumulant generating function $\lambda_{f,g}^i(z) = \log \mathbb{E} \left(e^{z\ell_{f,g}^i} \right)$ is finite for each $z \in \mathbb{R}$ and any two distinct states f, g . This assumption allows us to liberally use results from large deviations theory, and is satisfied, for example, whenever $\ell_{f,g}^i$ is bounded or if $S = \mathbb{R}$ and μ_f^i is a normal distribution with the same variance for each state f .

The private signals of agent i up to any period t induce the private log-likelihood ratio

$$L_{f,g,t}^i = \log \frac{\pi_0(f)}{\pi_0(g)} + \sum_{r \leq t} \ell_{f,g}^i(\mathbf{s}_r^i)$$

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 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 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.⁷ In Remark 4, we explain that the assumption of unobserved utilities can be weakened substantially.

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 .

⁷Formally, consider the case that agent i 's flow utility for action a in period t is $\tilde{u}(a, \mathbf{s}_{t+1}^i)$ for an action and signal-dependent utility function $\tilde{u}: A \times S \rightarrow \mathbb{R}$. Agents thus observe their flow utility in period t through their signal in the next period. If we define $u(a, f) = \mathbb{E}_f(\tilde{u}(a, \mathbf{s}_1^i))$, 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 $\sigma^i = (\sigma_t^i)_{t \in T}$ of measurable functions $\sigma_t^i: \mathcal{I}_{\leq t}^i \rightarrow A$ from i 's information sets in period t to the set of actions, and a pure strategy profile $\sigma = (\sigma^1, \dots, \sigma^n)$ consists of a strategy for each agent. A pure strategy profile σ induces a random sequence of action profiles: for each i , $a_1^i = \sigma_1^i(\mathfrak{s}_1^i)$, and for each $t > 1$, $a_t^i = \sigma_t^i(\mathfrak{s}_{\leq t}^i; (a_{< t}^j)_{j \in N^i})$, where i 's public history $(a_{< t}^j)_{j \in N^i}$ 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 $\mathcal{S}^i(H_{\leq t}) = \{s_{\leq t}^i \in S^t: \forall r \leq t, \sigma_r^i(s_{\leq r}^i; (H_{< r}^j)_{j \in N^i}) = H_r^i\}$ 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 $a_t^i \neq a_\omega$, and that i learns at rate r if

$$r = \liminf_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}(a_t^i \neq a_\omega)$$

If the limit exists, the probability of a mistake in period t is $e^{-rt+o(t)}$.⁸ Note that learning twice as fast corresponds to achieving the same mistake probability in half of the time. This definition of the learning rate is standard 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](#)). We emphasize that, for finitely many agents, it coincides with a quantitative definition that measures the expected utility difference between the correct action and the chosen action.⁹ For any fixed number of agents, the lowest rate of any agent determines how fast the expected number of mistakes per period vanishes, since the exponential with the smallest negative exponent dominates in the long run. Thus, the rate of the slowest-learning agent measures the rate of learning in the network. Moreover, any bound on the rate of learning for pure strategies entails the same bound for

⁸We use the asymptotic notation $o(t)$ for a function that grows slower than t as t goes to infinity. That is, $\phi(t) \in o(t)$ if $\lim_{t \rightarrow \infty} \frac{\phi(t)}{t} = 0$.

⁹An alternative, more quantitative 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.,

$$\liminf_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{E}(u(a_\omega, \omega) - u(a_t^i, \omega))$$

For finite action sets, both definitions coincide. Indeed, $\liminf_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{E}(u(a_\omega, \omega) - u(a_t^i, \omega)) = \liminf_{t \rightarrow \infty} -\frac{1}{t} \log c_t \mathbb{P}(a_t^i \neq a_\omega) = \liminf_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}(a_t^i \neq a_\omega)$ for some c_t between the minimum and the maximum of the set $\{u(a_f, f) - u(a, f): f \in \Omega, a \in A \setminus \{a_f\}\}$ for each $t \in T$. For infinite action sets and a bounded utility function, the learning rate is weakly higher for the quantitative definition and may be arbitrarily high even when the qualitative learning rate is 0. Hence, the qualitative definition of the learning rate is most reasonable with finitely many actions.

mixed strategies for the same instance with a larger signal space and an additional signal component that is uninformative about the state.¹⁰ Hence, restricting to pure strategies entails no loss in generality, and we do so throughout.

All notation is summarized in Appendix D.

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 periods and agents, and each agent's signal in each period matches the state with probability p for some $p \in (\frac{1}{2}, 1)$ (i.e., $\mu_f^i(s_f) = p$ and $\mu_g^i(s_g) = p$). 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.

For a single agent i who chooses actions optimally, it follows from classical large deviations results that

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

for some $r_{\text{aut}}^i > 0$ depending on the signal distributions $(\mu_f^i)_{f \in \Omega}$.¹¹ Thus, the limit $\lim_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}(a_t^i \neq a_\omega)$ exists and is equal to r_{aut}^i , which we call i 's autarky learn-

¹⁰More precisely, replace the signal space S by $\tilde{S} = S \times [0, 1]$, and for each state f , let $\tilde{\mu}_f$ be the product distribution on \tilde{S}^N with marginal μ_f 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 .

¹¹For 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.

ing rate. Note that the agent learns the state in the limit since the autarky learning rate is positive.

Now consider any number of agents with publicly observed signals. For signals that are conditionally independent and identically distributed across periods and agents, 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 $r_{\text{aut}} = r_{\text{aut}}^i$ does not depend on i .

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

Note that the learning rate grows linearly with the number of agents and thus becomes arbitrarily large as the number of agents grows. With signals that 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 a mismatched signal is sufficiently low, 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. 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.

We show that the learning rate of the slowest-learning agent is bounded by a constant that only depends on the distribution of the most informative private signal. It is, however, independent of the number of agents, the network structure, correlations between the signals, and the strategy profile imposed by the social planner. Hence, information aggregation breaks down even in the favorable case when strategies are not constrained by incentives. The driving force is a fundamental tradeoff between choosing the correct action and using actions to communicate information. More precisely, each agent's strategy must trade off between choosing the action that is most likely to be correct today and using actions to inform other agents about her private signals to lower their mistake rates tomorrow. The result shows that it is impossible to achieve both objectives to a high degree simultaneously.

Theorem 1 (Learning is bounded). *For any number of agents n , any network, and any strategies $\sigma^1, \dots, \sigma^n$, some agent learns at a rate of at most $r_{\text{bdd}} = \min_{f \neq g} \max_{i \in N} \mathbb{E}_f(\ell_{f,g}^i)$.*

In other words, for any strategy profile, there is some agent i such that for each $\epsilon > 0$, we have $\mathbb{P}(a_t^i \neq a_\omega) \geq e^{-(r_{\text{bdd}} + \epsilon)t}$ 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_f^i and is determined by the two states that are hardest to distinguish. Section 7 provides an outline of the proof of Theorem 1.

To provide some intuition, we discuss two policies that a social planner may try to use to achieve a uniformly high learning rate. In the first, each agent randomizes each period: with probability 10%, the agent's action choice reveals all of her information to her neighbors; with probability 90%, she chooses myopically optimal given her information.¹² These strategies provide each agent with all information (including everyone's private signals) about all but the most recent periods with high probability. Hence, the mistake probability is close to the public signal case in 90% of periods in the limit. However, each agent is likely to make a mistake in the remaining 10% of periods, so that the probability of mistakes does not converge to 0 and the learning rate is 0. The second planner policy tries to improve on the first through less frequent information revelation. Now the probability that an agent's action in period t reveals all of her information is, say, e^{-10t} ; otherwise the strategies are as above. Thus, revelation periods are exponentially unlikely and do not prohibit learning at a rate above r_{bdd} (assuming $10 > r_{\text{bdd}}$). However, the total number of revelation periods is almost surely finite under these strategies by the Borel-Cantelli lemma. So, information revelation always ceases eventually, and the agents eventually behave purely myopically. In the first example, enough information is revealed to allow for fast learning, but the revelation induces too many mistakes; the opposite holds for the second example. The proof of Theorem 1 shows that no planner policy can provide enough information revelation to facilitate fast learning while avoiding too many mistakes in the process.

We remark briefly that Theorem 1 can be upgraded to a statement about the fastest-learning agent's rate in symmetric situations. Indeed, for any symmetric distribution of signal profiles, any symmetric network, and any symmetric strategy profile, all agents learn at the same rate.¹³ The canonical example for this situation is the complete network with signals that are conditionally i.i.d. across periods and agents and the same strategy for all agents. It then follows directly from Theorem 1 that no agent can exceed the rate r_{bdd} . This gives one answer to a conjecture of Harel et al. (2021), who speculate that

¹²Recall that we allow the set of actions to be infinite, so that a single action choice can encode all information available to the agent.

¹³An automorphism of a network is a permutation of agents that preserves neighborhood relations. Then, a network is symmetric if for any two agents i and j , there exists an automorphism mapping i to j .

a social planner can improve learning by instructing agents to overweight their private signals. The intuition is that the actions become more informative about private signals and so provide more information for future periods.¹⁴ The preceding remarks show that, assuming symmetry, the benefit of such a planner policy cannot exceed r_{bdd} . In Section 6 we discuss another case in which Theorem 1 can be upgraded to a statement about the fastest-learning agent’s rate.

Theorem 1 is tight in two respects. First, without additional assumptions on the strategies, the statement cannot be lifted from one slow-learning agent to all agents learning slowly. Indeed, it is easy to construct strategies for which a large fraction of agents learn very fast if there are many agents. Consider the leading example from Section 3.4 with the following strategies: in each period, each odd-numbered agent chooses the action that matches her private signal in that period, 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 a rate of $\frac{n}{2}r_{\text{aut}}$, which grows linearly with the number of agents.

Second, the upper bound on the learning rate r_{bdd} of the slowest-learning agent cannot be improved. This is the content of our second result. In its proof, we construct strategies for which all agents learn at a rate close to r_{bdd} or faster, provided the agents’ signals are conditionally independent and identically distributed, the network is strongly connected, and the number of agents is large. Since $r_{\text{bdd}} > r_{\text{aut}}$, this shows that coordination can improve the learning rate of all agents.

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

In other words, for any positive ϵ , there exist strategies so that $\mathbb{P}(a_t^i \neq a_\omega) \leq e^{-(r_{\text{bdd}} - \epsilon)t + o(t)}$ for each agent i if the number of agents is large enough. The assumptions that the signals are conditionally independent and the network is strongly connected are both needed.¹⁵ We sketch the proof of Theorem 2 in Section 7.

¹⁴The conjecture is also supported by results of Arieli et al. (2025), who show that, in the classical herding model where agents arrive sequentially, moderate condescension—a misspecification where agents overweight their private signal but are otherwise rational—can significantly improve the efficiency of learning.

¹⁵First, if the agents’ signals were perfectly correlated instead of conditionally independent, no agent could exceed the autarky learning rate since even observing other agents’ signals would give no information beyond the agent’s private signals. Second, in networks that are not strongly connected, some agents may have no neighbors, in which case they have no more information than a single agent in autarky.

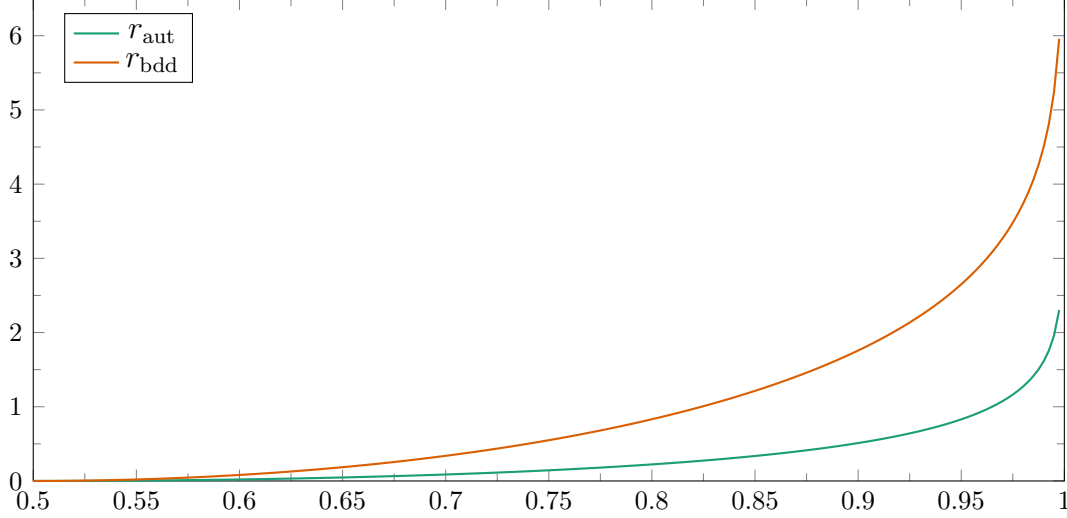


Figure 1: Bounds on the learning rates for the example in Section 3.4. The horizontal axis corresponds to the parameter p determining the informativeness of the signal distributions. Higher values of p correspond to more informative signals. The vertical axis indicates the rate of learning: r_{aut} is 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 in the proof of Theorem 2 (and sketched in Section 7), each agent learns at a rate close to r_{bdd} .

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

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

for each agent i . 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 factor $\delta \in [0, 1)$.¹⁶ The expected utility

¹⁶The assumption that the discount factor is the same for all agents is purely for notational simplicity. All results remain valid with heterogeneous discount factors.

of agent i for a strategy profile σ is

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

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 10), 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 implies that any upper bound on the slowest-learning agent's rate is also an upper bound on the fastest-learning agent's rate.¹⁷

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 in any strongly connected network.

Corollary 1. *For any number of agents, any strongly connected network, and any discount factor $\delta \in [0, 1)$, each agent learns at a rate of 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 tighter bound on the learning rate.¹⁸ Harel et al. (2021) obtain a smaller upper bound than Corollary 1 for two states and complete networks when the signals are conditionally independent and identically distributed across periods and agents, 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

¹⁷The 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.

¹⁸The bound on the learning rate obtained by Huang et al. (2024) is $2 \max_{f \neq g} \max_{i \in N} \sup_{s \in S} |\ell_{f,g}^i(s)| \geq \tilde{r}_{\text{bdd}} > 2r_{\text{bdd}}$. However, their model differs slightly from ours. They allow the signal space and the distribution of signals to vary across periods and agents 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 periods and agents.

information is lost in large networks. Giving a nontrivial lower bound on the equilibrium learning rate analogous to Theorem 2 is an open problem (cf. Remark 3).

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. In this alternative model, any nontrivial bound on the learning rate has to depend on the network structure. Indeed, if a rational agent observes everyone's signals, she learns at a rate of nr_{aut}^i , 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 periods and agents.

Corollary 2. *Assume the signals are conditionally independent across periods and agents. For any number of agents, any strongly connected network, and any discount factor $\delta \in [0, 1)$, each agent learns at a rate of 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 $\tilde{\mathbf{s}}_t^i = (\mathbf{s}_t^j)_{j \in N^i}$ be the vector with the signals of all of i 's neighbors in period t . The distribution of $\tilde{\mathbf{s}}_t^i$ in state $f \in \Omega$ is the $|N^i|$ -fold product $\tilde{\mu}_f^i = (\mu_f^i)^{\otimes |N^i|}$ of μ_f^i . Thus, defining $\tilde{\ell}_{f,g}^i = \log \frac{d\tilde{\mu}_f^i}{d\tilde{\mu}_g^i}(\tilde{\mathbf{s}}_1^i)$ for two states $f, g \in \Omega$, we have $\tilde{\ell}_{f,g}^i = \sum_{j \in N^i} \ell_{f,g}^j$ since the signals $(\mathbf{s}_1^j)_{j \in N^i}$ are conditionally independent. Thus,

$$\min_{f \neq g} \max_{i \in N} \mathbb{E}_f \left(\tilde{\ell}_{f,g}^i \right) = \min_{f \neq g} \max_{i \in N} \sum_{j \in N^i} \mathbb{E}_f \left(\ell_{f,g}^j \right) \leq \max_{i \in N} |N^i| \min_{f \neq g} \max_{i \in N} \mathbb{E}_f \left(\ell_{f,g}^i \right) = \Delta r_{\text{bdd}}$$

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

7. Proof Outlines

In this section, we sketch the proofs of the two main results.

7.1. About the Proof of Theorem 1

In contrast to earlier results, Theorem 1 provides an upper bound on the learning rate for arbitrary strategy profiles rather than only equilibria and is thus, in principle, harder to establish. However, the greater generality makes clear that the argument cannot rely on analyzing belief dynamics or incentives. Moreover, it is clearly without loss to assume that the network is complete, 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, requiring that the signals' log-likelihood ratios are bounded, and loosening the bound on the learning rate. This case already captures the main tension in the model. We sketch its proof and then elaborate on how to adapt it to get Theorem 1.

Theorem 1' (Learning is bounded, weak form). Assume the signals are conditionally independent across periods and agents, and the log-likelihood ratio $\ell_{f,g}^i(s)$ is bounded for each agent i and any two states f, g . For any number of agents n , any network, and any strategies $\sigma^1, \dots, \sigma^n$, some agent learns at a rate of at most $\tilde{r}_{\text{bdd}} = 2 \min_{f \neq g} \max_{i \in N} \sup_{s \in S} |\ell_{f,g}^i(s)|$.

We focus on the case of two states $\{f, g\}$. The intuition is as follows. Suppose all agents learn faster than the prescribed rate. Then in state g , each agent chooses a_g except for a set of signal trajectories with very small probability in any sufficiently late period. But then in state f , these sets still have a moderately small probability, and so with positive probability, all agents incorrectly choose a_g in each sufficiently late period. This contradicts that agents learn at all.

In more detail, assume for contradiction that all agents learn at a rate of at least $\tilde{r}_{\text{bdd}} + 3\epsilon$ 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

$$\mathbb{P}_g(a_t^i \neq a_g \mid H_{\leq t_0}) \leq e^{-(\tilde{r}_{\text{bdd}} + 2\epsilon)t}$$

for each agent i and each period $t > t_0$, and t_0 may be chosen arbitrarily large. In the proof, we show that there exists a set of signal profile trajectories $\mathcal{S}_{\leq t_0} \subset S^{N \times t_0}$ such that (i) $\mathbb{P}_g(\mathcal{S}_{\leq t_0}) > 0$, (ii) each $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$ is consistent with $H_{\leq t_0}$, and (iii) for each $i \in N$, each $t \in T$, and almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$,

$$\mathbb{P}_g(a_t^i \neq a_g \mid s_{\leq t_0}) \leq e^{-(\tilde{r}_{\text{bdd}} + \epsilon)t}$$

Denote by H the infinite 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 . We have that $\mathbb{P}_g(H \mid s_{\leq t_0}) \geq \frac{1}{2}$ for almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$ if t_0 is large enough, since the probability that some agent chooses an action different from a_g in any period after t_0 is small by the preceding inequality. We say that agent i defects in period $t > t_0$ if i is the first agent to deviate from H , i.e., if play follows $H_{< t}$ in the first $t - 1$ periods and $a_t^i \neq a_g$. Consider the set D_t^i of those trajectories of i 's signals for which i defects in period t . We aim to bound the probability of each D_t^i conditional on state f and signal profile trajectories in $\mathcal{S}_{\leq t_0}$.

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

$$\mathbb{P}_g(D_t^i, \forall j \neq i, H_{< t}^j \mid s_{\leq t_0}) \leq \mathbb{P}_g(a_t^i \neq a_g \mid s_{\leq t_0}) \leq e^{-(\tilde{r}_{\text{bdd}} + \epsilon)t}$$

for almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$. Since signals are conditionally independent across periods and agents, the probability on the left-hand side factors.

$$\mathbb{P}_g(D_t^i, \forall j \neq i, H_{< t}^j \mid s_{\leq t_0}) = \mathbb{P}_g(D_t^i \mid s_{\leq t_0}) \mathbb{P}_g(\forall j \neq i, H_{< t}^j \mid s_{\leq t_0})$$

for almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$. But

$$\mathbb{P}_g(\forall j \neq i, H_{< t}^j \mid s_{\leq t_0}) \geq \mathbb{P}_g(H \mid s_{\leq t_0}) \geq \frac{1}{2}$$

and so $\mathbb{P}_g(D_t^i \mid s_{\leq t_0}) \leq 2e^{-(\tilde{r}_{\text{bdd}} + \epsilon)t}$ for almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$. Since signals are conditionally independent across agents, we have $\mathbb{P}_g(D_t^i \mid s_{\leq t_0}) = \mathbb{P}_g(D_t^i \mid s_{\leq t_0}^i)$ and similarly for f instead of g . Hence, from the definition of \tilde{r}_{bdd} and the conditional independence of signals across periods, we see that

$$\mathbb{P}_f(D_t^i \mid s_{\leq t_0}) = \mathbb{P}_f(D_t^i \mid s_{\leq t_0}^i) \leq e^{\tilde{r}_{\text{bdd}} t} \mathbb{P}_g(D_t^i \mid s_{\leq t_0}^i) = e^{\tilde{r}_{\text{bdd}} t} \mathbb{P}_g(D_t^i \mid s_{\leq t_0}) \leq 2e^{-\epsilon t}$$

for almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$, 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 $s_{\leq t_0}^i$ 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, then $2n \sum_{t > t_0} e^{-\epsilon t} \leq \frac{1}{2}$, and so the probability that some agent defects after a signal profile trajectory in $\mathcal{S}_{\leq t_0}$ is thus at most $\frac{1}{2}$ 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 . Thus, in state f , each agent's mistake probability in each period after t_0 is bounded below by a positive constant, so her learning rate is 0, contradicting the assumption.

The proof of Theorem 1 needs to address two further points. First, it needs an argument that only loses $e^{r_{\text{bdd}} t}$ instead of $e^{\tilde{r}_{\text{bdd}} t}$ in the last inequality above. This is taken care of with a simple trick. Second, one cannot use the conditional independence of the agents' private signals. This is a substantial hurdle that requires a delicate argument.

7.2. About the proof of Theorem 2

To prove Theorem 2, we construct strategies for which each agent learns at a rate close to r_{bdd} or faster. For complete networks, such strategies 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.

In more detail, consider again the case of two states f and g . Agent i 's log-likelihood ratio for f over g in period t given the agent's private signals $\mathbf{s}_{\leq t}^i$ is close to $\mathbb{E}_f(\ell_{f,g}^i)t$ with high probability in state f , and similarly in state g . If this log-likelihood ratio for f over g is larger than $(\mathbb{E}_f(\ell_{f,g}^i) - \epsilon)t$, agent i chooses a_f independently of the other agents' actions; similarly, if i 's log-likelihood ratio for g over f is larger than $(\mathbb{E}_g(\ell_{g,f}^i) - \epsilon)t$, she chooses a_g ; otherwise, she chooses the action that most agents chose in period $t - 1$.

These strategies lead to faster learning than 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. 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 an equilibrium for rational and geometrically discounting agents.

For a strongly connected but not necessarily complete network, one can use a similar idea. The only obstacle is that agents do not know the most popular action of the previous period if they cannot see all other agents' actions. We deal with this issue by dedicating some periods to propagating information about the agents' actions in previous periods through the network. More precisely, in these propagation periods, each agent mimics the action of a neighbor in a recent previous period, or of a neighbor's neighbor (which they know through their neighbors' mimicking behavior), and so on for any distance level. In the remaining periods, agents act as before: they follow their private signals if those are sufficiently decisive, and otherwise they follow the action taken by most agents in the latest non-propagation period, which they have learned in the intervening propagation periods.

8. Discussion

We conclude with several remarks about assumptions, 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 no action is optimal in two different states and there is a unique optimal action in each state. The first 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 second assumption forces a tradeoff between correct action choices and using actions as messages. By contrast, if there are two optimal actions in each state and signals are binary, 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.¹⁹

Remark 2 (Arbitrary signal correlation). We have assumed that the conditional distributions of signal profiles are mutually absolutely continuous 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 from the second period onward 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 in the first 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 whether there is any equilibrium for any network in

¹⁹This 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.

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 explicitly constructing an equilibrium in which learning exceeds the autarky benchmark.

A related question is whether the 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 (Learning with experimentation). We shut down experimentation motives by assuming that agents do not observe their flow utilities. In Footnote 7, we explain that this allows for noisy observations of utilities as long as those do not provide information beyond next period's private signal. Theorem 1 survives (with a different constant replacing r_{bdd}) even if agents make noisy observations of their utilities for which future signals are not a sufficient statistic. Suppose whenever agent i chooses action a in state f , her observed utility, denoted $u_t^i(a)$, follows a distribution $\nu_{a,f}^i$ on \mathbb{R} . While this model allows for experimentation, it embeds into ours by replacing each private signal \mathbf{s}_t^i with $\tilde{\mathbf{s}}_t^i = (\mathbf{s}_t^i, (u_t^i(a))_{a \in A})$. Assuming the augmented signals $(\tilde{\mathbf{s}}_t^i)_{i \in N, t \in T}$ satisfy our hypothesis, Theorem 1 still yields an upper bound on the learning rate. This is also an upper bound for the original instance with observed utilities $u_t^i(a)$ since the embedding provides each agent with weakly more information.

Remark 5 (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 28 in the proof of Theorem 2.

Acknowledgements

The author thanks seminar audiences in Berlin, Bonn, and Warwick, as well as Krishna Dasaratha, Johannes Hörner, Kate Huang, and Ilan Kremer for comments on the paper. The author also 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.

The cumulant generating function of a random variable ℓ is

$$\lambda(z) = \log \mathbb{E} (e^{z\ell})$$

We assume that ℓ is non-degenerate and that $\lambda(z)$ is finite for each $z \in \mathbb{R}$, which implies that $\mathbb{E}(\ell)$ is finite. 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 $I^* = \{\eta \in \mathbb{R} : \exists z \in \mathbb{R}, \lambda'(z) = \eta\}$. Then,²⁰*

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

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

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

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

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

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

(iii) *λ is differentiable with*

$$\lambda'(z) = \frac{\mathbb{E}(\ell e^{z\ell})}{\mathbb{E}(e^{z\ell})}$$

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

²⁰Since we assume that ℓ is non-degenerate and has finite cumulants, 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. See also Lemma 5 of Harel et al. (2021).

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

Let ℓ_1, ℓ_2, \dots 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 $\mathbb{E}(\ell)$, and similarly if η is larger than $\mathbb{E}(\ell)$.

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

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

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

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

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 $(-\infty, \mathbb{E}(\ell)]$ and non-decreasing on $[\mathbb{E}(\ell), \infty)$, or by applying Theorem 6 of Harel et al. (2021) to ℓ and $-\ell$.

For $f, g \in \Omega$ with $f \neq g$ and $i \in N$, let $\lambda_{f,g}^i$ be the cumulant generating function of $\ell_{f,g}^i$, and denote by $(\lambda_{f,g}^i)^*$ its Fenchel-Legendre transform. It is not hard to show that $\lambda_{f,g}^i(z) = \lambda_{g,f}^i(-(z+1))$ and $(\lambda_{f,g}^i)^*(\eta) = (\lambda_{g,f}^i)^*(-\eta) - \eta$.²¹ Hence, by Lemma 2(ii),

$$(\lambda_{f,g}^i)^*(-\mathbb{E}_g(\ell_{g,f}^i)) = (\lambda_{g,f}^i)^*(\mathbb{E}_g(\ell_{g,f}^i)) + \mathbb{E}_g(\ell_{g,f}^i) = \mathbb{E}_g(\ell_{g,f}^i) \quad (1)$$

Since $\lambda_{f,g}^i(0) = 0$, $\lambda_{f,g}^i(-1) = \lambda_{g,f}^i(0) = 0$, and since $\lambda_{f,g}^i$ is strictly convex, $\lambda_{f,g}^i$ attains its minimum on $(-1, 0)$. Then, using again that $\lambda_{f,g}^i$ is strictly convex,

$$(\lambda_{f,g}^i)^*(0) = \sup_{z \in \mathbb{R}} -\lambda_{f,g}^i(z) = -\min_{z \in (-1, 0)} \lambda_{f,g}^i(z) < (\lambda_{f,g}^i)'(0) = \mathbb{E}_f(\ell_{f,g}^i) \quad (2)$$

B. Single-Agent Learning

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

²¹These relations appear in Lemma 6 of Harel et al. (2021). Since they use different sign conventions to define $\lambda_{f,g}^i$ and $(\lambda_{f,g}^i)^*$, we reproduce the argument here. First,

$$\lambda_{f,g}^i(z) = \log \int_S e^{z \log \frac{d\mu_f^i}{d\mu_g^i}(s)} d\mu_f^i(s) = \log \int_S \left(\frac{d\mu_f^i}{d\mu_g^i}(s) \right)^z d\mu_f^i(s) = \log \int_S \left(\frac{d\mu_g^i}{d\mu_f^i}(s) \right)^{-(z+1)} d\mu_g^i(s) = \lambda_{g,f}^i(-(z+1))$$

Thus,

$$(\lambda_{f,g}^i)^*(\eta) = \sup_{z \in \mathbb{R}} \eta z - \lambda_{f,g}^i(z) = \sup_{z \in \mathbb{R}} \eta z - \lambda_{g,f}^i(-(z+1)) = \sup_{z \in \mathbb{R}} (-\eta)z - \lambda_{g,f}^i(z) - \eta = (\lambda_{g,f}^i)^*(-\eta) - \eta$$

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).²²

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

$$\mathbb{P}(a_t^i \neq a_\omega) = e^{-(\lambda_{f,g}^i)^*(0)t + o(t)}$$

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*

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

where

$$r_{\text{aut}}^i = \min_{f \neq g} (\lambda_{f,g}^i)^*(0)$$

Proof. Let

$$r = \sup_{\sigma^i} \liminf_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}(a_t^i \neq a_\omega)$$

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

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

$$\liminf_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}(a_t^i \neq a_\omega) > r_{\text{aut}}^i$$

²²Note that $0 \in (I_{f,g}^i)^*$ by the remarks after Lemma 2 and the fact that $\sup \ell_{f,g}^i > 0$ and $\inf \ell_{f,g}^i < 0$.

Then, there are $f, g \in \Omega$ with $f \neq g$ such that $\liminf_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}(a_t^i \neq a_\omega) > (\lambda_{g,f}^i)^*(0)$. 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

$$\liminf_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}(a_t^i \neq a_\omega \mid \omega \in \{f, g\}) \geq \liminf_{t \rightarrow \infty} -\frac{1}{t} \log \mathbb{P}(a_t^i \neq a_\omega) > (\lambda_{g,f}^i)^*(0)$$

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

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

$$a_t^i = \sigma_t^i(\mathfrak{s}_1^i, \dots, \mathfrak{s}_t^i) \in \{a_f \in A : f \in \Omega, L_{f,g,t}^i \geq 0 \forall g \in \Omega\}$$

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.²³ Then, for all $f \in \Omega$,

$$\mathbb{P}_f(a_t^i \neq a_f) \leq \sum_{g \in \Omega} \mathbb{P}_f(L_{f,g,t}^i \leq 0) = \sum_{g \in \Omega} e^{-(\lambda_{f,g}^i)^*(0)t + o(t)} \leq e^{-r_{\text{aut}}^i + o(t)}$$

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,

$$\mathbb{P}(a_t^i \neq a_\omega) = \sum_{f \in \Omega} \mathbb{P}_f(a_t^i \neq a_f) \mathbb{P}(\omega = f) \leq e^{-r_{\text{aut}}^i + o(t)}$$

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. Heuristically, the first 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.

²³Note that σ^i is well-defined since $L_{f,g,t}^i + L_{g,g',t}^i = L_{f,g',t}^i$ ensuring that the right-hand side in the definition of σ^i is always non-empty.

Lemma 3. Let ℓ_1, ℓ_2, \dots be non-degenerate i.i.d. random variables such that $\mathbb{E}(e^{z\ell_1})$ is finite for each $z \in \mathbb{R}$ and let $a^+, a^- \in \mathbb{R}$ with $a^+ > \mathbb{E}(\ell_1)$ and $a^- < \mathbb{E}(\ell_1)$. 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 (with respect to the distribution of (L_1, \dots, L_{t_0})) $x_{\leq t_0} \in \mathbb{R}^{t_0}$ with $b - \frac{K}{2} \leq x_{t_0} \leq b + \frac{K}{2}$,

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

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 $\mathbb{E}(\ell_1)$. Hence, we may and will assume that $\mathbb{E}(\ell_1) < a^+ < \sup_{z \in \mathbb{R}} \lambda'(z)$ and $\inf_{z \in \mathbb{R}} \lambda'(z) < a^- < \mathbb{E}(\ell_1)$, 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 $\tilde{a}^+ = \frac{1}{2}(\mathbb{E}(\ell_1) + a^+)$ and $\tilde{a}^- = \frac{1}{2}(\mathbb{E}(\ell_1) + a^-)$. By Lemma 2(ii), $\lambda^*(\tilde{a}^+) > 0$ and $\lambda^*(\tilde{a}^-) > 0$. Thus, for all $t \geq t_0$, we have by Theorem 3 that

$$\begin{aligned} \mathbb{P}(L_t - L_{t_0} \geq \tilde{a}^+(t - t_0) + o(t - t_0)) &= e^{-\lambda^*(\tilde{a}^+)(t - t_0) + o(t - t_0)}, \text{ and} \\ \mathbb{P}(L_t - L_{t_0} \leq \tilde{a}^-(t - t_0) + o(t - t_0)) &= e^{-\lambda^*(\tilde{a}^-)(t - t_0) + o(t - t_0)} \end{aligned}$$

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

$$\mathbb{P}(\forall t \geq t_1, a^-(t - t_0) \leq L_t - L_{t_0} \leq a^+(t - t_0)) \geq 1 - \frac{\delta}{2}$$

Let $K > 0$ such that $\mathbb{P}(-\frac{K}{2} \leq \ell_1 \leq \frac{K}{2}) \geq \frac{\delta}{2(t_1 - t_0)}$. Then it follows that

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

But ℓ_1, ℓ_2, \dots are independent and $b - \frac{K}{2} \leq x_{t_0} \leq b + \frac{K}{2}$ and so

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

which finishes the proof. \square

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 $\sigma^1, \dots, \sigma^n$ be strategies such that for each $i \in N$,

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

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

$$\mathbb{P}_g(a_t^i \neq a_g \mid H_{\leq t_0}) \leq e^{-(r-\epsilon)t}$$

Moreover, t_0 may be chosen arbitrarily large.

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

$$\mathbb{P}_g(a_t^i \neq a_g) \leq e^{-rt+o(t)}$$

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

$$\sum_{i \in N, t > \tilde{t}_0} \mathbb{P}_g(a_t^i \neq a_g) \leq \frac{1}{2}$$

and

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

and, thus, there is $H_{\leq \tilde{t}_0}^* \in A^{N \times \tilde{t}_0}$ such that $\mathbb{P}_g(H_{\leq \tilde{t}_0}^*) > 0$ and

$$\sum_{i \in N, t > \tilde{t}_0} \mathbb{P}_g(a_t^i \neq a_g \mid H_{\leq \tilde{t}_0}^*) \leq \frac{1}{2} \quad (3)$$

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 $t > \tilde{t}_0$, $H_t^i = a_g$. Note that, by (3),

$$\mathbb{P}_g(H \mid H_{\leq \tilde{t}_0}) \geq \frac{1}{2} \quad (4)$$

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

$$\mathbb{P}_g(a_t^i \neq a_g \mid H_{\leq \tilde{t}_0}) \leq \frac{1}{2} e^{-(r-\epsilon)t}$$

and, thus, by (4),

$$\mathbb{P}_g(a_t^i \neq a_g \mid H_{\leq t_0}) \leq 2\mathbb{P}_g(a_t^i \neq a_g \mid H_{\leq \tilde{t}_0})\mathbb{P}_g(H_{\leq t_0} \mid H_{\leq \tilde{t}_0}) \leq 2\mathbb{P}_g(a_t^i \neq a_g \mid H_{\leq \tilde{t}_0}) \leq e^{-(r-\epsilon)t}$$

which proves the claim. \square

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.²⁴ 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 $f \in \Omega$, denote by $\mu_f^{\otimes t}$ the t -fold product measure of μ_f on $S^{N \times t}$. Then, for each $t \in T$ and each $f, g \in \Omega$, $\mu_f^{\otimes t}$ and $\mu_g^{\otimes t}$ 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 $\sigma^1, \dots, \sigma^n$, some agent learns at a rate of at most $r_{\text{bdd}} = \min_{f \neq g} \max_{i \in N} \mathbb{E}_f(\ell_{f,g}^i)$.*

Proof. For each $g \in \Omega$, we denote by $\mathbb{P}_g(\cdot \mid \mathfrak{s} = \cdot)$ a regular conditional probability, and write $\mathbb{P}_g(\cdot \mid \mathfrak{s} = \cdot) = \mathbb{P}_g(\cdot \mid \cdot)$ again for convenience.²⁵ Let $f, g \in \Omega$ such that $r_{\text{bdd}} = \max_{i \in N} \mathbb{E}_f(\ell_{f,g}^i)$. Fix strategies $\sigma^1, \dots, \sigma^n$. 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 $t \in T$,

$$\mathbb{P}(a_t^i \neq a_\omega) \leq e^{-(r_{\text{bdd}} + 6\epsilon)t + o(t)} \quad (5)$$

We write $r = r_{\text{bdd}}$ 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) $\mathbb{P}_g(H_{\leq t_0}) > 0$,
- (ii) for each $i \in N$ and each $t > t_0$, $\mathbb{P}_g(a_t^i \neq a_g \mid H_{\leq t_0}) \leq e^{-(r+5\epsilon)t}$,
- (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 $t > t_0$, $H_t^i = a_g$,

²⁴A proof of this fact can be found here: <https://math.stackexchange.com/q/1042323>.

²⁵That is, (i) for each $s \in S^T$, $\mathbb{P}_g(\cdot \mid \mathfrak{s} = s)$ is a probability measure on the underlying probability space, (ii) for each event E , $\mathbb{P}_g(E \mid \mathfrak{s} = \cdot)$ is measurable, and (iii) for each event E and each measurable set $\mathcal{S} \subset S^{N \times T}$, $\mathbb{P}_g(E \cap \mathfrak{s}^{-1}(\mathcal{S})) = \int_{\mathcal{S}} \mathbb{P}_g(E \mid \mathfrak{s} = s) d\mathbb{P} \circ \mathfrak{s}^{-1}(s)$.

where K is the constant obtained from applying Lemma 3 to i.i.d. random variables ℓ_1, ℓ_2, \dots each with distribution $\mathbb{P}_f(\ell_{f,g}^i \in \cdot)$, $a^+ = \mathbb{E}_f(\ell_{f,g}^i) + \epsilon$, $a^- = \mathbb{E}_f(\ell_{f,g}^i) - \epsilon$, and $\delta = \frac{1}{2n}$ for each $i \in N$ separately and taking the maximum.²⁶

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 $m^1, \dots, m^n \in \mathbb{R}$ 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) $\mathbb{P}_g(\mathcal{S}_{\leq t_0}) > 0$, (iii) for each $i \in N$, each $t > t_0$, and almost every $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$, $\mathbb{P}_g(a_t^i \neq a_g \mid s_{\leq t_0}) \leq e^{-(r+4\epsilon)t}$, 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 $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}(i, t)$,

$$\mathbb{P}_g(a_t^i \neq a_g \mid s_{\leq t_0}) \geq e^{-(r+4\epsilon)t}$$

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 $t > t_0$,

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

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

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

for which we have

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

²⁶Here, $\mathbb{P}_f(\ell_{f,g}^i \in \cdot)$ denotes the distribution of $\ell_{f,g}^i$ conditional on state f .

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

$$\mathbb{P}_g(a_t^i \neq a_g \mid s_{\leq t_0}) \leq e^{-(r+4\epsilon)t}$$

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 $m^1, \dots, m^n \in \mathbb{R}$ such that $\mathbb{P}_g(\mathcal{S}_{\leq t_0}) > 0$ 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

$$\mathcal{W} = \{s \in S^{N \times T} : \forall i \in N, \forall t \geq t_0,$$

$$(m^i - \epsilon)t_0 - (\mathbb{E}_f(\ell_{f,g}^i) + \epsilon)(t - t_0) \leq L_{g,f}^i(s_{\leq t}^i) \leq (m^i + \epsilon)t_0 - (\mathbb{E}_f(\ell_{f,g}^i) - \epsilon)(t - t_0)\}$$

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 $-\mathbb{E}_f(\ell_{f,g}^i) + \epsilon$ and $-\mathbb{E}_f(\ell_{f,g}^i) - \epsilon$ 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 $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$, $\mathbb{P}_f(\mathcal{W} \mid s_{\leq t_0}) \geq 1 - \frac{n}{2n} = \frac{1}{2}$. Note that for each $i \in N$, each $t \geq t_0$, and each $\mathcal{V}_{\leq t}^i \subset \mathcal{W}_{\leq t}^i$,

$$e^{-(m^i + \epsilon)t_0 + (\mathbb{E}_f(\ell_{f,g}^i) - \epsilon)(t - t_0)} \mathbb{P}_g(\mathcal{V}_{\leq t}^i) \leq \mathbb{P}_f(\mathcal{V}_{\leq t}^i) \leq e^{-(m^i - \epsilon)t_0 + (\mathbb{E}_f(\ell_{f,g}^i) + \epsilon)(t - t_0)} \mathbb{P}_g(\mathcal{V}_{\leq t}^i) \quad (6)$$

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

$$\mathbb{P}_g(a_t^i \neq a_g \mid s_{\leq t_0}, H_{< t}) \leq 2e^{-(r+4\epsilon)t} \quad (7)$$

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 $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$,

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

and, moreover,

$$\mathbb{P}_g(a_t^i \neq a_g \mid s_{\leq t_0}, H_{< t}) \mathbb{P}_g(H_{< t} \mid s_{\leq t_0}) \leq \mathbb{P}_g(a_t^i \neq a_g \mid s_{\leq t_0})$$

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

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

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 $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$,

$$\mathbb{P}_g(D_t^i \mid s_{\leq t_0}, H_{<t}) = \mathbb{P}_g(a_t^i \neq a_g \mid s_{\leq t_0}, H_{<t}) \leq 2e^{-(r+4\epsilon)t} \quad (8)$$

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 $\mathbb{P}_f(H) > 0$, 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 $t > t_0$, $\mathbb{P}_f(D_t^i) \leq e^{-\epsilon t}$, which would follow from $\mathbb{P}_g(D_t^i) \leq e^{-(r+3\epsilon)t}$ by the definition of r . This is close to (8), except that (8) involves conditioning on $H_{<t}$. Unfortunately, $\mathbb{P}_g(D_t^i) \leq e^{-(r+3\epsilon)t}$ does not follow from (8) since $\mathbb{P}_g(D_t^i \mid H_{<t})$ can be arbitrarily small compared to $\mathbb{P}_g(D_t^i)$. 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 $\mathbb{P}_g(D_t^j \mid s^i, H_{<t})$ 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 $\mathbb{P}_g(D_t^j \mid s^i, H_{<t})$ is small for all j and t . For a trajectory $s^i \in E_t^i \cup F^i$, $\mathbb{P}_g(H_{<t} \mid s^i)$ 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 $t > t_0$,

$$\mathbb{P}_g(a_t^j \neq a_g \mid \mathbf{s}_{\leq t_0}, \mathbf{s}^i, H_{<t}) = \mathbb{P}_g(a_t^j \neq a_g \mid \mathbf{s}_{\leq t_0}, \mathbf{s}_{\leq t}^i, H_{<t}) \quad (9)$$

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

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

and define

$$F^i(s_{\leq t_0}) = \left\{ s^i \in \{s_{\leq t_0}^i\} \times S^{\{t_0+1, t_0+2, \dots\}} : \forall j \in N, \forall t > t_0, \mathbb{P}_g(a_t^j \neq a_g \mid s_{\leq t_0}, s^i, H_{<t}) < e^{-\epsilon t} \right\}$$

and note that $\{E_t^i(s_{\leq t_0}) : t > t_0\} \cup \{F^i(s_{\leq t_0})\}$ is a partition of $\{s_{\leq t_0}^i\} \times S^{\{t_0+1, t_0+2, \dots\}}$. Intuitively, $E_t^i(s_{\leq t_0})$ is the set of infinite trajectories of i 's signals with prefix $s_{\leq t_0}^i$ such that conditional on $s_{\leq t_0}$, 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 $F^i(s_{\leq t_0})$ 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 $E_t^i(s_{\leq t_0})$ as “empty” in the sense that some other agent exhausts an appreciable fraction of them through defections, and to think of the sets $F^i(s_{\leq t_0})$ 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 $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$, and almost every $s^i \in E_t^i(s_{\leq t_0}) \cup F^i(s_{\leq t_0})$,

$$\begin{aligned} \mathbb{P}_g(H_{<t} \mid s_{\leq t_0}, s^i) &= \prod_{t > t' > t_0} \mathbb{P}_g(H_{<t'+1} \mid s_{\leq t_0}, s^i, H_{<t'}) \\ &= \prod_{t > t' > t_0} (1 - \mathbb{P}_g(\exists j \in N, a_{t'}^j \neq a_g \mid s_{\leq t_0}, s^i, H_{<t'})) \\ &\geq 1 - \sum_{j \in N, t > t' > t_0} \mathbb{P}_g(a_{t'}^j \neq a_g \mid s_{\leq t_0}, s^i, H_{<t'}) \\ &\geq 1 - n \sum_{t > t' > t_0} e^{-\epsilon t'} \\ &\geq \frac{1}{2} \end{aligned} \tag{10}$$

where the fourth step uses that $s^i \notin E_{t'}^i(s_{\leq t_0})$ for $t' < t$. Thus, for each $i \in N$, each

$t > t_0$, and almost every $s_{\leq t_0}$,

$$\begin{aligned}
\mathbb{P}_g(E_t^i(s_{\leq t_0}) \mid s_{\leq t_0}) &= \int_{E_t^i(s_{\leq t_0})} d\mathbb{P}_g(s^i \mid s_{\leq t_0}) \\
&\leq 2 \int_{E_t^i(s_{\leq t_0})} \mathbb{P}_g(H_{<t} \mid s_{\leq t_0}, s^i) d\mathbb{P}_g(s^i \mid s_{\leq t_0}) \\
&= 2\mathbb{P}_g(E_t^i(s_{\leq t_0}) \cap H_{<t} \mid s_{\leq t_0}) \\
&\leq 2\mathbb{P}_g(E_t^i(s_{\leq t_0}) \mid s_{\leq t_0}, H_{<t}) \\
&= 2 \int_{E_t^i(s_{\leq t_0})} d\mathbb{P}_g(s^i \mid s_{\leq t_0}, H_{<t}) \\
&\leq 2e^{\epsilon t} \int_{E_t^i(s_{\leq t_0})} \sum_{j \in N} \mathbb{P}_g(a_t^j \neq a_g \mid s_{\leq t_0}, s^i, H_{<t}) d\mathbb{P}_g(s^i \mid s_{\leq t_0}, H_{<t}) \\
&\leq 2e^{\epsilon t} \sum_{j \in N} \mathbb{P}_g(a_t^j \neq a_g \mid s_{\leq t_0}, H_{<t}) \\
&\leq 4ne^{\epsilon t} e^{-(r+4\epsilon)t} \\
&= 4ne^{-(r+3\epsilon)t}
\end{aligned} \tag{11}$$

where the second step uses (10), the sixth step uses the definition of $E_t^i(s_{\leq t_0})$, and the second to last step uses (7). In words, the left-hand side is the probability that, conditional on g and $s_{\leq t_0}$, agent i observes an infinite trajectory 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 $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$,

$$\begin{aligned}
\mathbb{P}_g\left(\left(D_t^i \cap F^i(s_{\leq t_0})\right)_{\leq t} \mid s_{\leq t_0}\right) &\leq 2\mathbb{P}_g\left(\left(D_t^i \cap F^i(s_{\leq t_0})\right)_{\leq t} \mid s_{\leq t_0}, H_{<t}\right) \\
&\leq 2\mathbb{P}_g(a_t^i \neq a_g \mid s_{\leq t_0}, H_{<t}) \\
&\leq 4e^{-(r+4\epsilon)t}
\end{aligned} \tag{12}$$

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 $s_{\leq t_0}$, agent i observes a trajectory 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 $s_{\leq t_0} \in \mathcal{S}_{\leq t_0}$, with positive probability, no agent defects in any period.

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

$$\begin{aligned}
\mathbb{P}_f(E_t^i(s_{\leq t_0}) \mid s_{\leq t_0}, \mathcal{W}) &\leq 2\mathbb{P}_f(E_t^i(s_{\leq t_0}) \cap \mathcal{W} \mid s_{\leq t_0}) \\
&\leq 2\mathbb{P}_f((E_t^i(s_{\leq t_0}) \cap \mathcal{W}^i)_{\leq t} \mid s_{\leq t_0}) \\
&\leq 2e^{(m^i+\epsilon)t_0} e^{-(m^i-\epsilon)t_0 + (\mathbb{E}_f(\ell_{f,g}^i) + \epsilon)(t-t_0)} \mathbb{P}_g((E_t^i(s_{\leq t_0}) \cap \mathcal{W}^i)_{\leq t} \mid s_{\leq t_0}) \\
&\leq 2e^{2\epsilon t_0 + (\mathbb{E}_f(\ell_{f,g}^i) + \epsilon)(t-t_0)} \mathbb{P}_g((E_t^i(s_{\leq t_0}) \cap \mathcal{W}^i)_{\leq t} \mid s_{\leq t_0}) \\
&\leq 2e^{(r+2\epsilon)t} \mathbb{P}_g((E_t^i(s_{\leq t_0}) \cap \mathcal{W}^i)_{\leq t} \mid s_{\leq t_0}) \\
&\leq 8ne^{-\epsilon t}
\end{aligned}$$

where the first step uses that $\mathbb{P}_f(\mathcal{W} \mid s_{\leq t_0}) \geq \frac{1}{2}$ 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 $E_t^i(s_{\leq t_0})$ only depends on its prefix up to and including t . Thus,

$$\mathbb{P}_f(\cup_{i \in N, t > t_0} E_t^i(s_{\leq t_0}) \mid s_{\leq t_0}, \mathcal{W}) \leq 8n^2 \sum_{t > t_0} e^{-\epsilon t} \leq \frac{1}{4} \quad (13)$$

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

$$\begin{aligned}
\mathbb{P}_f(D_t^i \cap F^i(s_{\leq t_0}) \mid s_{\leq t_0}, \mathcal{W}) &\leq 2\mathbb{P}_f(D_t^i \cap F^i(s_{\leq t_0}) \cap \mathcal{W} \mid s_{\leq t_0}) \\
&\leq 2\mathbb{P}_f((D_t^i \cap F^i(s_{\leq t_0}))_{\leq t} \cap \mathcal{W}^i)_{\leq t} \mid s_{\leq t_0}) \\
&\leq 2e^{(r+2\epsilon)t} \mathbb{P}_g((D_t^i \cap F^i(s_{\leq t_0}) \cap \mathcal{W}^i)_{\leq t} \mid s_{\leq t_0}) \\
&\leq 8e^{-2\epsilon t}
\end{aligned}$$

and thus,

$$\mathbb{P}_f(\cup_{i \in N, t > t_0} D_t^i \cap F^i(s_{\leq t_0}) \mid s_{\leq t_0}, \mathcal{W}) \leq 4n \sum_{t > t_0} e^{-2\epsilon t} \leq \frac{1}{4} \quad (14)$$

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

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

If the infinite history is not H , some agent defects in some period, hence the first step. The second step uses that $\{E_t^i(s_{\leq t_0}) : t > t_0\} \cup \{F^i(s_{\leq t_0})\}$ is a partition of $\{s_{\leq t_0}^i\} \times$

$S^{\{t_0+1, t_0+2, \dots\}}$. The third step is a basic manipulation. The fourth step follows from (13) and (14). Hence, Step 2, $\mathbb{P}_g(\mathcal{S}_{\leq t_0}) > 0$, and Lemma 5 imply that

$$\mathbb{P}_f(H) \geq \int_{\mathcal{S}_{\leq t_0}} \mathbb{P}_f(H \mid s_{\leq t_0}, \mathcal{W}) \mathbb{P}_f(\mathcal{W} \mid s_{\leq t_0}) d\mathbb{P}_f(s_{\leq t_0}) \geq \frac{1}{2} \int_{\mathcal{S}_{\leq t_0}} \mathbb{P}_f(\mathcal{W} \mid s_{\leq t_0}) d\mathbb{P}_f(s_{\leq t_0}) > 0$$

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. \square

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 periods and agents. For any $\epsilon > 0$, there is n_0 such that for all $n \geq n_0$ and any strongly connected network, there exist strategies $\sigma^1, \dots, \sigma^n$ such that each agent learns at a rate of 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 $f, g \in \Omega$, define $\ell_{f,g} = \ell_{f,g}^1$ and $m_{f,g} = \mathbb{E}_f(\ell_{f,g})$, and fix $\epsilon > 0$. First, by (1), $\lambda_{f,g}^*(-m_{g,f}) = m_{g,f}$. Second, since $I_{f,g}^*$ contains $m_{g,f}$ and $\lambda_{f,g}^*$ is continuous on $I_{f,g}^*$ by Lemma 2 and the remarks thereafter, there is $\delta > 0$ such that

$$\min_{f \neq g} \lambda_{f,g}^*(-m_{g,f} + \delta) > \min_{f \neq g} \lambda_{f,g}^*(-m_{g,f}) - \epsilon$$

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

$$\hat{r}_{\text{bdd}} = \min_{f \neq g} \lambda_{f,g}^*(-m_{g,f} + \delta)$$

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

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 $a_{t-1}^{\text{pop}} \in \arg \max_{a \in A} |\{j \in N : a_{t-1}^j = a\}|$ be an action that is most popular among the actions taken in period $t-1$. For a history $H_{<t}^i \in A^{N^i \times (t-1)}$ for agent i , let

$$a_t^i = \sigma_t^i(\mathbf{s}_1^i, \dots, \mathbf{s}_t^i; H_{<t}^i) = \begin{cases} a_f & \text{if } L_{f,g,t}^i \geq (m_{f,g} - \delta)t \ \forall g \neq f, \text{ and} \\ a_{t-1}^{\text{pop}} & \text{otherwise} \end{cases}$$

Note that σ^i is well-defined since $L_{f,g,t}^i = -L_{g,f,t}^i$ and $m_{f,g} - \delta > 0$. 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 $f, g \in \Omega$,

$$\mathbb{P}_g (L_{g,f,t}^i \leq -(m_{f,g} - \delta)t) = e^{-\lambda_{g,f}^* (-m_{f,g} + \delta)t + o(t)}$$

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

$$\begin{aligned} \mathbb{P} (a_t^i \neq a_\omega \mid L_{f,g,t}^i \geq (m_{f,g} - \delta)t \ \forall g \neq f) &\leq \sum_{g \neq f} \mathbb{P}_g (L_{f,g,t}^i \geq (m_{f,g} - \delta)t) \mathbb{P} (\omega = g) \\ &= \sum_{g \neq f} \mathbb{P}_g (L_{g,f,t}^i \leq -(m_{f,g} - \delta)t) \mathbb{P} (\omega = g) \\ &= \sum_{g \neq f} e^{-\lambda_{g,f}^* (-m_{f,g} + \delta)t + o(t)} \mathbb{P} (\omega = g) \\ &= e^{-\hat{r}_{\text{bdd}} t + o(t)} \end{aligned}$$

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

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

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 $f \in \Omega$,

$$\begin{aligned} \mathbb{P}_f (a_{t-1}^{\text{pop}} \neq a_f) &\leq \mathbb{P}_f \left(|\{i \in N : L_{f,g,t-1}^i \geq (m_{f,g} - \delta)(t-1) \ \forall g \neq f\}| \leq \frac{n}{2} \right) \\ &\leq e^{-nD(\frac{1}{2} \parallel 1 - q_{f,t-1})} \end{aligned}$$

where

$$q_{f,t} = \sum_{g \neq f} e^{-\lambda_{f,g}^* (m_{f,g} - \delta)t + o(t)} \mathbb{P} (\omega = g) = e^{-\min_{f \neq g} \lambda_{f,g}^* (m_{f,g} - \delta)t + o(t)}$$

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

$$\begin{aligned}
D\left(\frac{1}{2} \parallel 1 - q_{f,t}\right) &= \frac{1}{2} \left(\log \frac{1}{2(1 - q_{f,t})} + \log \frac{1}{2q_{f,t}} \right) \\
&= \log \frac{1}{2} + \frac{1}{2} \left(\log \frac{1}{1 - q_{f,t}} + \log \frac{1}{q_{f,t}} \right) \\
&= \log \frac{1}{2} + \frac{1}{2} \left(q_{f,t} + o(q_{f,t}) + \min_{f \neq g} \lambda_{f,g}^* (m_{f,g} - \delta) t + o(t) \right) \\
&= \frac{1}{2} \min_{f \neq g} \lambda_{f,g}^* (m_{f,g} - \delta) t + o(t)
\end{aligned}$$

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 quantity that converges to 0 faster than x as $x \rightarrow 0$. Thus,

$$\mathbb{P}_f(a_{t-1}^{\text{pop}} \neq a_f) \leq e^{-\frac{n}{2} \min_{f \neq g} \lambda_{f,g}^* (m_{f,g} - \delta) t + o(t)}$$

Since $\lambda_{f,g}^*(m_{f,g}) = 0$ and $\lambda_{f,g}^*$ is strictly decreasing on $(\inf \ell_{f,g}, m_{f,g}]$ by Lemma 2(ii), $\lambda_{f,g}^*(m_{f,g} - \delta) > 0$ for all $f \neq g$, and so $\min_{f \neq g} \lambda_{f,g}^*(m_{f,g} - \delta) > 0$. 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}_{bdd} . More precisely, we need that

$$n \geq 2 \frac{\min_{f \neq g} \lambda_{f,g}^* (-m_{g,f} + \delta)}{\min_{f \neq g} \lambda_{f,g}^* (m_{f,g} - \delta)}$$

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

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 $d(i, j) = 1$.²⁷ 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

²⁷More 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 $\min\{d(i, i') : i' \in N, d(i, i') \leq k-1, j \in N^{i'}\} = k-1$.

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.²⁸

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

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

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

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

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.

□

²⁸Extending Theorem 2 to random networks as discussed in Remark 5 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 .

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 behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.
- S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.
- S. Bikhchandani, D. Hirshleifer, O. Tamuz, and I. Welch. Information cascades and social learning. Technical report, arXiv.org, 2021.
- 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.
- R. Lévy, M. Peški, 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. 2017. Working paper.
- 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.
- L. Smith, P. N. Sorensen, and J. Tian. Informational herding, optimal experimentation, and contrarianism. *Review of Economic Studies*, 88(5):2527–2554, 2021.
- X. Vives. How fast do rational agents learn? *Review of Economic Studies*, 60(2):329–347, 1993.

D. Notation

The table below summarizes the notation used in the paper.

Symbol [elements]	Name	Mathematical object
N	i, j	agents
T	t	periods
A	a, a_t^i	actions
Ω	f, g	states
ω		random unknown state
π_0		common prior belief
u		utility function
a_f		utility-maximizing action in state f
S	s, s_t^i	signals
\mathfrak{s}_t^i		i 's private signal in period t
μ_f^i		distribution of \mathfrak{s}_t^i in state f
$\ell_{f,g}^i(s)$		i 's log-likelihood ratio given signal s
$\lambda_{f,g}^i(z)$		CGF of $\ell_{f,g}^i(\mathfrak{s}_1^i)$
$(\lambda_{f,g}^i)^*(\eta)$		Fenchel-Legendre transform of $\lambda_{f,g}^i(z)$
$L_{f,g,t}^i$		i 's log-likelihood ratio given $\mathfrak{s}_{\leq t}^i$
N^i	j	i 's neighborhood
$\mathcal{I}_{\leq t}^i$		i 's information sets in period t
σ_t^i		i 's strategy in period t
σ		strategy profile
$\mathcal{S}^i(H_{\leq t})$	$s_{\leq t}^i$	i 's signals compatible with $H_{\leq t}$
$\mathcal{S}(H_{\leq t})$	$s_{\leq t}$	signal profiles compatible with $H_{\leq t}$
r_{aut}^i		i 's learning rate in autarky
r_{bdd}		tight learning rate upper bound
\tilde{r}_{bdd}		relaxed learning rate upper bound