2-2014

Asymptotic Learning on Bayesian Social Networks

Elchanan Mossel
University of Pennsylvania

Allan Sly

Omer Tamuz

Recommended Citation

Mossel, E., Sly, A., & Tamuz, O. (2014). Asymptotic Learning on Bayesian Social Networks. *Probability Theory and Related Fields, 158* (1), 127-157. http://dx.doi.org/10.1007/s00440-013-0479-y

Asymptotic Learning on Bayesian Social Networks

Abstract

Understanding information exchange and aggregation on networks is a central problem in theoretical economics, probability and statistics. We study a standard model of economic agents on the nodes of a social network graph who learn a binary “state of the world” S, from initial signals, by repeatedly observing each other’s best guesses.

Asymptotic learning is said to occur on a family of graphs $G_n = (V_n, E_n)$ with $|V_n| \to \infty$ if with probability tending to 1 as $n \to \infty$ all agents in G_n eventually estimate S correctly. We identify sufficient conditions for asymptotic learning and contruct examples where learning does not occur when the conditions do not hold.

Keywords

Bayesian learning, social networks, aggregation of information, rational expectatons

Disciplines

Statistics and Probability

Asymptotic Learning on Bayesian Social Networks*

Elchanan Mossel   Allan Sly   Omer Tamuz

July 26, 2012

Abstract

Understanding information exchange and aggregation on networks is a central problem in theoretical economics, probability and statistics. We study a standard model of economic agents on the nodes of a social network graph who learn a binary "state of the world" S, from initial signals, by repeatedly observing each other's best guesses.

Asymptotic learning is said to occur on a family of graphs $G_n = (V_n, E_n)$ with $|V_n| \to \infty$ if with probability tending to 1 as $n \to \infty$ all agents in G_n eventually estimate S correctly. We identify sufficient conditions for asymptotic learning and contruct examples where learning does not occur when the conditions do not hold.

1 Introduction

We consider a directed graph G representing a social network. The nodes of the graph are the set of agents V, and an edge from agent u to w indicates that u can observe the actions of w. The agents try to estimate a binary *state of the world* $S \in \{0,1\}$, where each of the two possible states occurs with probability one half.

The agents are initially provided with private signals which are informative with respect to S and i.i.d., conditioned on S: There are two distributions, $\mu_0 \neq \mu_1$, such that conditioned on S, the private signals are independent and distributed μ_S.

In each time period $t \in \mathbb{N}$, each agent v chooses an "action" $A_v(t)$, which equals whichever of $\{0,1\}$ the state of the world is more likely to equal, conditioned on the information available to v at time t. This information includes its private signal, as well as the actions of its social network neighbors in the previous periods.

A first natural question is whether the agents eventually reach consensus, or whether it is possible that neighbors "agree to disagree" and converge to different actions. Assuming that the agents do reach consensus regarding their estimate of S, a second natural question is whether this consensus estimator is equal to S. Certainly, since private signals are independent conditioned on S, a large enough group of agents has, in the aggregation of their private signals, enough information to learn S with high probability. However, it may be the case that this information is

*The authors would like to thank Shachar Kariv for an enthusiastic introduction to his work with Douglas Gale, and for suggesting the significance of asymptotic learning in this model. Elchanan Mossel is supported by NSF NSF award DMS 1106999, by ONR award N000141110140 and by ISF grant 1300/08. Allan Sly is supported in part by an Alfred Sloan Fellowship in Mathematics. Omer Tamuz is supported by ISF grant 1300/08, and is a recipient of the Google Europe Fellowship in Social Computing. This research is supported in part by this Google Fellowship.

not disseminated by the above described process. These and related questions have been studied extensively in economics, statistics and operations research; see Section 1.1.

We say that the agents learn on a social network graph G when all their actions converge to the state of the world S. For a sequence of graphs $\{G_n\}_{n=1}^{\infty}$ such that G_n has n agents, we say that *Asymptotic learning* occurs when the probability that the agents learn on G_n tends to one as n tends to infinity, for a fixed choice of private signal distributions μ_1 and μ_0.

An agent's initial *private belief* is the probability that $S = 1$, conditioned only on its private signal. When the distribution of private beliefs is atomic, asymptotic learning does not necessarily occur (see Example A.1). This is also the case when the social network graph is undirected (see Example 2.7). Our main result (Theorem 3) is that asymptotic learning occurs for non-atomic private beliefs and undirected graphs.

To prove this theorem we first prove that the condition of non-atomic initial private beliefs implies that the agents all converge to the same action, or all don't converge at all (Theorem 1). We then show that for any model in which this holds, asymptotic learning occurs (Theorem 2). Note that it has been shown that agents reach agreement under various other conditions (cf. Ménager [12]). Hence, by Theorem 2, asymptotic learning also holds for these models.

Our proof includes several novel insights into the dynamics of interacting Bayesian agents. Broadly, we show that on undirected social network graphs connecting a *countably infinite* number of agents, if all agents converge to the same action then they converge to the correct action. This follows from the observation that if agents in distant parts of a large graph converge to the same action then they do so *almost* independently. We then show that this implies that for *finite* graphs of growing size the probability of learning approaches one.

At its heart of this proof lies a topological lemma (Lemma 3.13) which may be of independent interest; the topology here is one of rooted graphs (see, e.g., Benjamini and Schramm [5], Aldous and Steele [1]). The fact that asymptotic learning occurs for undirected graphs (as opposed to general strongly connected graphs) is related to the fact that sets of bounded degree, *undirected* graphs are compact in this topology. In fact, our proof applies equally to any such compact sets. For example, one can replace *undirected* with *L-locally strongly connected*: a directed graph $G = (V, E)$ is L-locally strongly connected if, for each $(u, w) \in E$, there exists a path in G of length at most L from w to u. Asymptotic learning also takes place on L-locally strongly connected graphs, for fixed L, since sets of L-locally strongly connected, uniformly bounded degree graphs are compact. See Section 3.7 for further discussion.

1.1 Related literature

1.1.1 Agreement

There is a vast economic literature studying the question of convergence to consensus in dynamic processes and games. A founding work is Aumann's seminal Agreement Theorem [2], which states that Bayesian agents who observe beliefs (i.e., posterior probabilities, as opposed to actions in our model) cannot "agree to disagree". Subsequent work (notably Geanakoplos and Polemarchakis [10], Parikh and Krasucki [14], McKelvey and Page [11], Gale and Kariv [9] Ménager [12] and Rosenberg, Solan and Vieille [15]) expanded the range of models that display convergence to consensus. One is, in fact, left with the impression that it takes a pathological model to feature interacting Bayesian agents who do "agree to disagree".

Ménager [12] in particular describes a model similar to ours and proves that consensus is achieved in a social network setting under the condition that the probability space is finite and ties cannot occur (i.e., posterior beliefs are always different than one half). Note that our asymptotic learning result applies for any model where consensus is guaranteed, and hence in particular applies to models satisfying Ménager's conditions.

1.1.2 Agents on social networks

Gale and Kariv [9] also consider Bayesian agents who observe each other's actions. They introduce a model in which, as in ours, agents receive a single initial private signal, and the action space is discrete. However, there is no "state of the world" or conditionally i.i.d. private signals. Instead, the relative merit of each possible action depends on all the private signals. Our model is in fact a particular case of their model, where we restrict our attention to the particular structure of the private signals described above.

Gale and Kariv show (loosely speaking) that neighboring agents who converge to two different actions must, at the limit, be indifferent with respect to the choice between these two actions. Their result is therefore also an agreement result, and makes no statement on the optimality of the chosen actions, although they do profess interest in the question of *"... whether the common action chosen asymptotically is optimal, in the sense that the same action would be chosen if all the signals were public information... there is no reason why this should be the case."* This is precisely the question we address.

A different line of work is the one explored by Ellison and Fudenberg [8]. They study agents on a social network that use rules of thumb rather than full Bayesian updates. A similar approach is taken by Bala and Goyal [3], who also study agents acting iteratively on a social network. They too are interested in asymptotic learning (or "complete learning", in their terms). They consider a model of bounded rationality which is not completely Bayesian. One of their main reasons for doing so is the mathematical complexity of the fully Bayesian model, or as they state, *"to keep the model mathematically tractable... this possibility [fully Bayesian agents] is precluded in our model... simplifying the belief revision process considerably."* In this simpler, non-Bayesian model, Bala and Goyal show both behaviors of asymptotic learning and results of non-learning, depending on various parameters of their model.

1.1.3 Herd behavior

The "herd behavior" literature (cf. Banerjee [4], Bikhchandani, Hirshleifer and Welch [6], Smith and Sørensen [16]) consider related but fundamentally simpler models. As in our model there is a "state of the world" and conditionally independent private signals. A countably infinite group of agents is exogenously ordered, and each picks an action sequentially, after observing the actions of its predecessors or some of its predecessors. Agents here act only *once*, as opposed to our model in which they act *repeatedly*.

The main result for these models is that in some situations there may arise an "information cascade", where, with positive probability, almost all the agents take the wrong action. This is precisely the opposite of *asymptotic learning*. The condition for information cascades is "bounded private beliefs"; herd behavior occurs when the agents' beliefs, as inspired by their private signals, are bounded away both from zero and from one [16]. In contrast, we show that in our model asymptotic learning occurs even for bounded beliefs.

In the herd behavior models information only flows in one direction: If agent u learns from w then w does not learn from u. This significant difference, among others, makes the tools used for their analysis irrelevant for our purposes.

2 Formal definitions, results and examples

2.1 Main definitions

The following definition of the agents, the state of the world and the private signals is adapted from [13], where a similar model is discussed.

Definition 2.1. *Let $(\Omega, \mathcal{O})$ be a σ-algebra. Let μ_0 and μ_1 be different and mutually absolutely continuous probability measures on $(\Omega, \mathcal{O})$.*

Let δ_0 and δ_1 be the distributions on $\{0,1\}$ such that $\delta_0(0) = \delta_1(1) = 1$.

Let V be a countable (finite or infinite) set of agents, and let

$$\mathbb{P} = \tfrac{1}{2}\delta_0\mu_0^V + \tfrac{1}{2}\delta_1\mu_1^V,$$

be a distribution over $\{0,1\} \times \Omega^V$. We denote by $S \in \{0,1\}$ the **state of the world** *and by W_u the* **private signal** *of agent $u \in V$. Let*

$$(S, W_{u_1}, W_{u_2}, \ldots) \sim \mathbb{P}.$$

Note that the private signals W_u are i.i.d., conditioned on S: if $S = 0$ - which happens with probability half - the private signals are distributed i.i.d. μ_0, and if $S = 1$ then they are distributed i.i.d. μ_1.

We now define the dynamics of the model.

Definition 2.2. *Consider a set of agents V, a state of the world S and private signals $\{W_u : u \in V\}$ such that*

$$(S, W_{u_1}, W_{u_2}, \ldots) \sim \mathbb{P},$$

as defined in Definition 2.1.

Let $G = (V, E)$ be a directed graph which we shall call the **social network**. *We assume throughout that G is simple (i.e., no parallel edges or loops) and strongly connected. Let the set of neighbors of u be $N(u) = \{v : (u,v) \in E\}$. The* **out-degree** *of u is equal to $|N(u)|$.*

For each time period $t \in \{1, 2, \ldots\}$ and agent $u \in V$, denote the **action** *of agent u at time t by $A_u(t)$, and denote by $\mathcal{F}_u(t)$ the information available to agent u at time t. They are jointly defined by*

$$\mathcal{F}_u(t) = \sigma(W_u, \{A_v(t') : v \in N(u), t' < t\}),$$

and

$$A_u(t) = \begin{cases} 0 & \mathbb{P}[S = 1 | \mathcal{F}_u(t)] < 1/2 \\ 1 & \mathbb{P}[S = 1 | \mathcal{F}_u(t)] > 1/2 \\ \in \{0,1\} & \mathbb{P}[S = 1 | \mathcal{F}_u(t)] = 1/2. \end{cases}$$

Let $X_u(t) = \mathbb{P}[S = 1 | \mathcal{F}_u(t)]$ be agent u's **belief** *at time t.*

Informally stated, $A_u(t)$ is agent u's best estimate of S given the information $\mathcal{F}_u(t)$ available to it up to time t. The information available to it is its private signal W_u and the actions of its neighbors in G in the previous time periods.

Remark 2.3. *An alternative and equivalent definition of $A_u(t)$ is the MAP estimator of S, as calculated by agent u at time t:*

$$A_u(t) = \operatorname*{argmax}_{s \in \{0,1\}} \mathbb{P}\left[S = s | \mathcal{F}(t)\right] = \operatorname*{argmax}_{A \in \mathcal{F}(t)} \mathbb{P}\left[A = S\right],$$

with some tie-breaking rule.

Note that we assume nothing about how agents break ties, i.e., how they choose their action when, conditioned on their available information, there is equal probability for S to equal either 0 or 1.

Note also that the belief of agent u at time $t = 1$, $X_u(1)$, depends only on W_u:

$$X_u(1) = \mathbb{P}\left[S = 1 | W_u\right].$$

We call $X_u(1)$ the *initial belief* of agent u.

Definition 2.4. *Let μ_0 and μ_1 be such that $X_u(1)$, the initial belief of u, has a non-atomic distribution ($\Leftrightarrow$ the distributions of the initial beliefs of all agents are non-atomic). Then we say that the pair (μ_0, μ_1)* **induce non-atomic beliefs**.

We next define some limiting random variables: $\mathcal{F}_u$ is the limiting information available to u, and X_u is its limiting belief.

Definition 2.5. *Denote $\mathcal{F}_u = \cup_t \mathcal{F}_u(t)$, and let*

$$X_u = \mathbb{P}\left[S = 1 | \mathcal{F}_u\right].$$

Note that the limit $\lim_{t \to \infty} X_u(t)$ almost surely exists and equals X_u, since $X_u(t)$ is a bounded martingale.

We would like to define the limiting *action* of agent u. However, it might be the case that agent u takes both actions infinitely often, or that otherwise, at the limit, both actions are equally desirable. We therefore define A_u to be the limiting *optimal action set*. It can take the values $\{0\}$, $\{1\}$ or $\{0, 1\}$.

Definition 2.6. *Let A_u, the* **optimal action set** *of agent u, be defined by*

$$A_u = \begin{cases} \{0\} & X_u < 1/2 \\ \{1\} & X_u > 1/2 \\ \{0,1\} & X_u = 1/2. \end{cases}$$

Note that if a is an action that u takes infinitely often then $a \in A_u$, but that if 0 (say) is the only action that u takes infinitely often then it still may be the case that $A_u = \{0, 1\}$. However, we show below that when (μ_0, μ_1) induce non-atomic beliefs then A_u is almost surely equal to the set of actions that u takes infinitely often.

2.2 Main results

In our first theorem we show that when initial private beliefs are non-atomic, then at the limit $t \to \infty$ the optimal action sets of the players are identical. As Example A.1 indicates, this may not hold when private beliefs are atomic.

Theorem 1. *Let (μ_0, μ_1) induce non-atomic beliefs. Then there exists a random variable A such that almost surely $A_u = A$ for all u.*

I.e., when initial private beliefs are non-atomic then agents, at the limit, agree on the optimal action. The following theorem states that when such agreement is guaranteed then the agents learn the state of the world with high probability, when the number of agents is large. This phenomenon is known as *asymptotic learning.* This theorem is our main result.

Theorem 2. *Let μ_0, μ_1 be such that for every connected, undirected graph G there exists a random variable A such that almost surely $A_u = A$ for all $u \in V$. Then there exists a sequence $q(n) = q(n, \mu_0, \mu_1)$ such that $q(n) \to 1$ as $n \to \infty$, and $\mathbb{P}[A = \{S\}] \geq q(n)$, for any choice of undirected, connected graph G with n agents.*

Informally, when agents agree on optimal action sets then they necessarily learn the correct state of the world, with probability that approaches one as the number of agents grows. This holds uniformly over all possible connected and undirected social network graphs.

The following theorem is a direct consequence of the two theorems above, since the property proved by Theorem 1 is the condition required by Theorem 2.

Theorem 3. *Let μ_0 and μ_1 induce non-atomic beliefs. Then there exists a sequence $q(n) = q(n, \mu_0, \mu_1)$ such that $q(n) \to 1$ as $n \to \infty$, and $\mathbb{P}[A_u = \{S\}] \geq q(n)$, for all agents u and for any choice of undirected, connected G with n agents.*

2.3 Note on directed vs. undirected graphs

Note that we require that the graph G not only be strongly connected, but also undirected (so that if $(u, v) \in E$ then $(v, u) \in E$.) The following example (depicted in Figure 1) shows that when private beliefs are bounded then asymptotic learning may not occur when the graph is strongly connected but not undirected[1].

Example 2.7. *Consider the the following graph. The vertex set is comprised of two groups of agents: a "royal family" clique of 5 agents who all observe each other, and $5 - n$ agents - the "public" - who are connected in a chain, and in addition can all observe all the agents in the royal family. Finally, a single member of the royal family observes one of the public, so that the graph is strongly connected.*

Now, with positive probability, which is independent of n, there occurs the event that all the members of the royal family initially take the wrong action. Assuming the private signals are sufficiently weak, then it is clear that all the agents of the public will adopt the wrong opinion of the royal family and will henceforth choose the wrong action.

Note that the removal of one edge - the one from the royal back to the commoners - results in this graph no longer being strongly connected. However, the information added by this edge

[1] We draw on Bala and Goyal's [3] *royal family* graph.

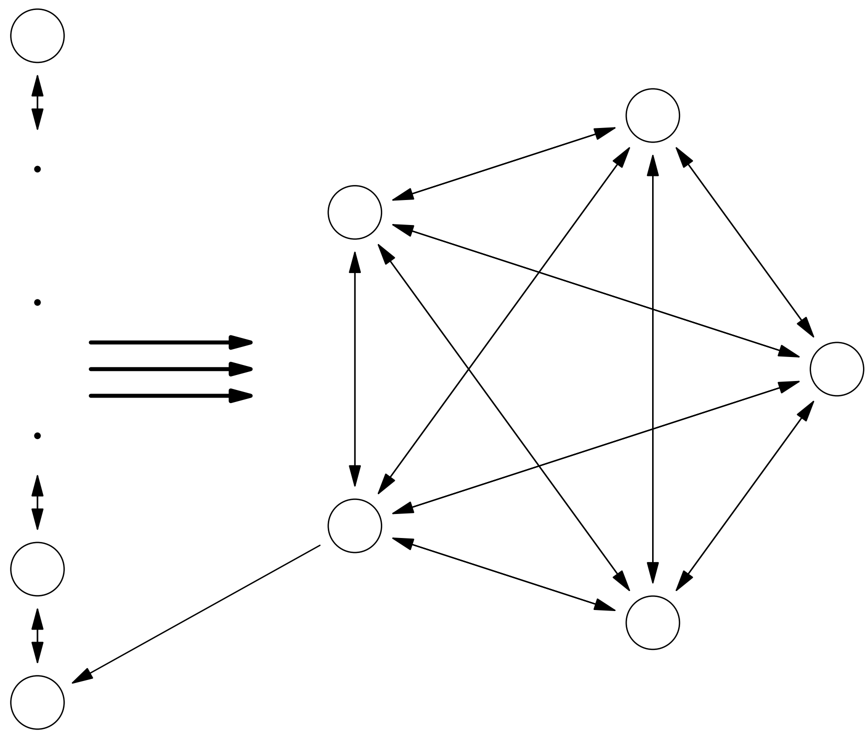

Figure 1: The five members of the royal family (on the right) all observe each other. The rest of the agents - the public - all observe the royal family (as suggested by the three thick arrows in the middle) and their immediate neighbors. Finally, one of the royals observes one of the public, so that the graph is strongly connected. This is an example of how asymptotic learning does not necessarily occur when the graph is undirected.

rarely has an affect on the final outcome of the process. This indicates that strong connectedness is too weak a notion of connectedness in this context. We therefore in seek stronger notions such as connectedness in undirected graphs.

A weaker notion of connectedness is that of L-locally strongly connected graphs, which we defined above. For any L, the graph from Example 2.7 is not L-locally strongly connected for n large enough.

3 Proofs

Before delving into the proofs of Theorems 1 and 2 we introduce additional definitions in subsection 3.1 and prove some general lemmas in subsections 3.2, 3.3 and 3.4. Note that Lemma 3.13, which is the main technical insight in the proof of Theorem 2, may be of independent interest. We prove Theorem 2 in subsection 3.5 and Theorem 1 in subsection 3.6.

3.1 Additional general notation

Definition 3.1. *We denote the* **log-likelihood ratio** *of agent u's belief at time t by*

$$Z_u(t) = \log \frac{X_u(t)}{1 - X_u(t)},$$

and let

$$Z_u = \lim_{t \to \infty} Z_u(t).$$

Note that

$$Z_u(t) = \log \frac{\mathbb{P}\left[S = 1 | \mathcal{F}_u(t)\right]}{\mathbb{P}\left[S = 0 | \mathcal{F}_u(t)\right]}.$$

and that

$$Z_u(1) = \log \frac{d\mu_1}{d\mu_0}(W_u).$$

Note also that $Z_u(t)$ converges almost surely since $X_u(t)$ does.

Definition 3.2. *We denote the set of actions of agent u up to time t by*

$$\bar{A}_u(t) = (A_u(1), \ldots, A_u(t-1)).$$

The set of all actions of u is similarly denoted by

$$\bar{A}_u = (A_u(1), A_u(2), \ldots).$$

We denote the actions of the neighbors of u up to time t by

$$I_u(t) = \{\bar{A}_w(t) :\ w \in N(u)\} = \{A_w(t') :\ w \in N(u), t' < t\},$$

and let I_u denote all the actions of u's neighbors:

$$I_u = \{\bar{A}_w :\ w \in N(u)\} = \{A_w(t') :\ w \in N(u), t' \geq 1\}.$$

Note that using this notation we have that $\mathcal{F}_u(t) = \sigma(W_u, I_u(t))$ and $\mathcal{F}_u = \sigma(W_u, I_u)$.

Definition 3.3. *We denote the probability that u chooses the correct action at time t by*

$$p_u(t) = \mathbb{P}\left[A_u(t) = S\right].$$

and accordingly

$$p_u = \lim_{t \to \infty} p_u(t).$$

Definition 3.4. *For a set of vertices U we denote by $W(U)$ the private signals of the agents in U.*

3.2 Sequences of rooted graphs and their limits

In this section we define a topology on *rooted graphs*. We call convergence in this topology *convergence to local limits*, and use it repeatedly in the proof of Theorem 2. The core of the proof of Theorem 2 is the topological Lemma 3.13, which we prove here. This lemma is a claim related to *local graph properties*, which we also introduce here.

Definition 3.5. *Let $G = (V, E)$ be a finite or countably infinite graph, and let $u \in V$ be a vertex in G. We denote by (G, u) the* **rooted graph** *G with root u.*

Definition 3.6. *Let $G = (V, E)$ and $G' = (V', E')$ be graphs. $h : V \to V'$ is a* **graph isomorphism** *between G and G' if $(u, v) \in E \Leftrightarrow (h(u), h(v)) \in E'$.*

Let (G, u) and (G', u') be rooted graphs. Then $h : V \to V'$ is a **rooted graph isomorphism** *between (G, u) and (G', u') if h is a graph isomorphism and $h(u) = u'$.*

We write $(G, u) \cong (G', u')$ whenever there exists a rooted graph isomorphism between the two rooted graphs.

Given a (perhaps directed) graph $G = (V, E)$ and two vertices $u, w \in V$, the graph distance $d(u, w)$ is equal to the length in edges of a shortest (directed) path between u and w.

Definition 3.7. *We denote by $B_r(G, u)$ the ball of radius r around the vertex u in the graph $G = (V, E)$: Let V' be the set of vertices w such that $d(u, w)$ is at most r. Let $E' = \{(u, w) \in E : u, w \in V'\}$. Then $B_r(G, u)$ is the rooted graph with vertices V', edges E' and root u'.*

We next define a topology on strongly connected rooted graphs (or rather on their isomorphism classes; we shall simply refer to these classes as graphs). A natural metric between strongly connected rooted graphs is the following (see Benjamini and Schramm [5], Aldous and Steele [1]). Given (G, u) and (G', u'), let

$$D((G, u), (G', u')) = 2^{-R},$$

where

$$R = \sup\{r : B_r(G, u) \cong B_r(G', u')\}.$$

This is indeed a metric: the triangle inequality follows immediately, and a standard diagonalization argument is needed to show that if $D((G, u), (G', u')) = 0$ then $(G, u) \cong (G', u')$.

This metric induces a topology that will be useful to us. As usual, the basis of this topology is the set of balls of the metric; the ball of radius 2^{-R} around the *graph* (G, u) is the set of graphs (G', u') such that $B_R(G, u) \cong B_R(G', u')$. We refer to convergence in this topology as convergence to a *local limit*, and provide the following equivalent definition for it:

Definition 3.8. *Let $\{(G_r, u_r)\}_{r=1}^{\infty}$ be a sequence of strongly connected rooted graphs. We say that the sequence converges if there exists a strongly connected rooted graph (G', u') such that*

$$B_r(G', u') \cong B_r(G_r, u_r),$$

for all $r \geq 1$. We then write

$$(G', u') = \lim_{r\to\infty} (G_r, u_r),$$

and call (G', u') the **local limit** *of the sequence $\{(G_r, u_r)\}_{r=1}^{\infty}$.*

Let $\mathcal{G}_d$ be the set of strongly connected rooted graphs with degree at most d. Another standard diagonalization argument shows that $\mathcal{G}_d$ is *compact* (see again [5, 1]). Then, since the space is metric, every sequence in $\mathcal{G}_d$ has a converging subsequence:

Lemma 3.9. *Let $\{(G_r, u_r)\}_{r=1}^{\infty}$ be a sequence of rooted graphs in $\mathcal{G}_d$. Then there exists a subsequence $\{(G_{r_i}, u_{r_i})\}_{i=1}^{\infty}$ with $r_{i+1} > r_i$ for all i, such that $\lim_{i\to\infty}(G_{r_i}, u_{r_i})$ exists.*

We next define *local properties* of rooted graphs.

Definition 3.10. *Let P be property of rooted graphs or a Boolean predicate on rooted graphs. We write $(G, u) \in P$ if (G, u) has the property, and $(G, u) \notin P$ otherwise.*

We say that P is a **local property** *if, for every $(G, u) \in P$ there exists an $r > 0$ such that if $B_r(G, u) \cong B_r(G', u')$, then $(G', u') \in P$. Let r be such that $B_r(G, u) \cong B_r(G', u') \Rightarrow (G', u') \in P$. Then we say that (G, u)* **has property** P **with radius** r*, and denote $(G, u) \in P^{(r)}$.*

That is, if (G, u) has a local property P then there is some r such that knowing the ball of radius r around u in G is sufficient to decide that (G, u) has the property P. An alternative name for a local property would therefore be a *locally decidable* property. In our topology, local properties are nothing but *open sets*: the definition above states that if $(G, u) \in P$ then there exists an element of the basis of the topology that includes (G, u) and is also in P. This is a necessary and sufficient condition for P to be open.

We use this fact to prove the following lemma.

Definition 3.11. *Let $\mathcal{B}_d$ be the set of infinite, connected, undirected graphs of degree at most d, and let $\mathcal{B}_d^r$ be the set of $\mathcal{B}_d$-rooted graphs*

$$\mathcal{B}_d^r = \{(G, u) \,:\, G \in \mathcal{B}_d, u \in G\}.$$

Lemma 3.12. *$\mathcal{B}_d^r$ is compact.*

Proof. Lemma 3.9 states that $\mathcal{G}_d$, the set of strongly connected rooted graphs of degree at most d, is compact. Since $\mathcal{B}_d^r$ is a subset of $\mathcal{G}_d$, it remains to show that $\mathcal{B}_d^r$ is closed in $\mathcal{G}_d$.

The complement of $\mathcal{B}_d^r$ in $\mathcal{G}_d$ is the set of graphs in $\mathcal{G}_d$ that are either finite or directed. These are both local properties: if (G, u) is finite (or directed), then there exists a radius r such that examining $B_r(G, u)$ is enough to determine that it is finite (or directed). Hence the sets of finite graphs and directed graphs in $\mathcal{G}_d$ are open in $\mathcal{G}_d$, their intersection is open in $\mathcal{G}_d$, and their complement, $\mathcal{B}_d^r$, is closed in $\mathcal{G}_d$. □

We now state and prove the main lemma of this subsection. Note that the set of graphs $\mathcal{B}_d$ satisfies the conditions of this lemma.

Lemma 3.13. *Let $\mathcal{A}$ be a set of infinite, strongly connected graphs, let $\mathcal{A}^r$ be the set of $\mathcal{A}$-rooted graphs*

$$\mathcal{A}^r = \{(G, u) \,:\, G \in \mathcal{A}, u \in G\},$$

and assume that $\mathcal{A}$ is such that $\mathcal{A}^r$ is compact.

Let P be a local property such that for each $G \in \mathcal{A}$ there exists a vertex $w \in G$ such that $(G, w) \in P$. Then for each $G \in \mathcal{A}$ there exist an r_0 and infinitely many distinct vertices $\{w_n\}_{n=1}^{\infty}$ such that $(G, w_n) \in P^{(r_0)}$ for all n.

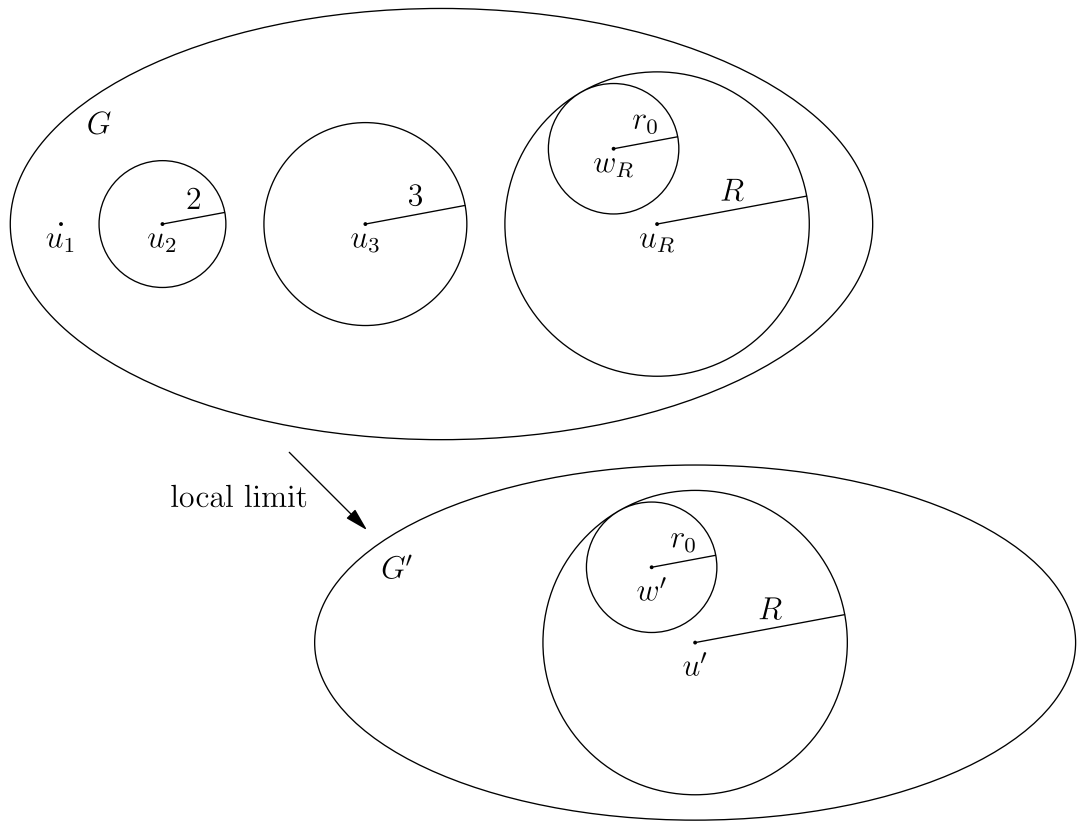

Figure 2: Schematic diagram of the proof of lemma 3.13. The rooted graph (G', u') is a local limit of (G, u_r). For $r \geq R$, the ball $B_R(G', u')$ is isomorphic to the ball $B_R(G, u_r)$, with $w' \in G'$ corresponding to $w_r \in G$.

Proof. Let G be an arbitrary graph in $\mathcal{A}$. Consider a sequence $\{v_r\}_{r=1}^{\infty}$ of vertices in G such that for all $r, s \in \mathbb{N}$ the balls $B_r(G, v_r)$ and $B_s(G, v_s)$ are disjoint.

Since $\mathcal{A}^r$ is compact, the sequence $\{(G, v_r)\}_{r=1}^{\infty}$ has a converging subsequence $\{(G, v_{r_i})\}_{r=1}^{\infty}$ with $r_{i+1} > r_i$. Write $u_r = v_{r_i}$, and let

$$(G', u') = \lim_{r \to \infty} (G, u_r).$$

Note that since $\mathcal{A}^r$ is compact, $(G', u') \in \mathcal{A}^r$ and in particular $G' \in \mathcal{A}$ is an infinite, strongly connected graph. Note also that since $r_{i+1} > r_i$, it also holds that the balls $B_r(G, u_r)$ and $B_s(G, u_s)$ are disjoint for all $r, s \in \mathbb{N}$.

Since $G' \in \mathcal{A}$, there exists a vertex $w' \in G'$ such that $(G', w') \in P$. Since P is a local property, $(G', w') \in P^{(r_0)}$ for some r_0, so that if $B_{r_0}(G', w') \cong B_{r_0}(G, w)$ then $(G, w) \in P$.

Let $R = d(u', w') + r_0$, so that $B_{r_0}(G', w') \subseteq B_R(G', u')$. Then, since the sequence (G, u_r) converges to (G', u'), for all $r \geq R$ it holds that $B_R(G, u_r) \cong B_R(G', u')$. Therefore, for all $r > R$ there exists a vertex $w_r \in B_R(G, u_r)$ such that $B_{r_0}(G, w_r) \cong B_{r_0}(G', w')$. Hence $(G, w_r) \in P^{(r_0)}$ for all $r > R$ (see Fig 2). Furthermore, for $r, s > R$, the balls $B_R(G, u_r)$ and $B_R(G, u_s)$ are disjoint, and so $w_r \neq w_s$.

We have therefore shown that the vertices $\{w_r\}_{r>R}$ are an infinite set of distinct vertices such that $(G, w_r) \in P^{(r_0)}$, as required.

$\square$

3.3 Coupling isomorphic balls

This section includes three claims that we will use repeatedly later. Their spirit is that everything that happens to an agent up to time t depends only on the state of the world and a ball of radius t around it.

Recall that $\mathcal{F}_u(t)$, the information available to agent u at time t, is the algebra generated by W_u and $A_w(t')$ for all w neighbors of u and $t' < t$. Recall that $I_u(t)$ denotes this exact set of actions:

$$I_u(t) = \left\{\bar{A}_w(t) : \ w \in N(u)\right\} = \left\{A_w(t') : \ w \in N(u), t' < t\right\}.$$

Claim 3.14. *For all agents u and times t, $I_u(t)$ a deterministic function of $W(B_t(G,u))$.*

Recall (Definition 3.4) that $W(B_t(G,u))$ are the private signals of the agents in $B_t(G,u)$, the ball of radius t around u (Definition 3.7).

Proof. We prove by induction on t. $I_u(1)$ is empty, and so the claim holds for $t = 1$.

Assume the claim holds up to time t. By definition, $A_u(t+1)$ is a function of W_u and of $I_u(t+1)$, which includes $\{A_w(t') : w \in N(u), t' \le t\}$. $A_w(t')$ is a function of W_w and $I_w(t')$, and hence by the inductive assumption it is a function of $W(B_{t'}(G,w))$. Since $t' < t+1$ and the distance between u and w is one, $W(B_{t'}(G,w)) \subseteq W(B_{t+1}(G,u))$, for all $w \in N(u)$ and $t' \le t$. Hence $I_u(t+1)$ is a function of $W(B_{t+1}(G,u))$, the private signals in $B_{t+1}(G,u)$. $\square$

The following lemma follows from Claim 3.14 above:

Lemma 3.15. *Consider two processes with identical private signal distributions (μ_0, μ_1), on different graphs $G = (V,E)$ and $G' = (V',E')$.*

Let $t \ge 1$, $u \in V$ and $u' \in V'$ be such that there exists a rooted graph isomorphism $h : B_t(G,u) \to B_t(G',w')$.

Let M be a random variable that is measurable in $\mathcal{F}_u(t)$. Then there exists an M' that is measurable in $\mathcal{F}_{u'}(t)$ such that the distribution of (M,S) is identical to the distribution of (M',S').

Recall that a graph isomorphism between $G = (V,E)$ and $G' = (V',E')$ is a bijective function $h : V \to V$ such that $(u,v) \in E$ iff $(h(u),h(v)) \in E'$.

Proof. Couple the two processes by setting $S = S'$, and letting $W_w = W_{w'}$ when $h(w) = w'$. Note that it follows that $W_u = W_{u'}$. By Claim 3.14 we have that $I_u(t) = I_{u'}(t)$, when using h to identify vertices in V with vertices in V'.

Since M is measurable in $\mathcal{F}_u(t)$, it must, by the definition of $\mathcal{F}_u(t)$, be a function of $I_u(t)$ and W_u. Denote then $M = f(I_u(t), W_u)$. Since we showed that $I_u(t) = I_{u'}(t)$, if we let $M' = f(I_{u'}(t), W_{u'})$ then the distribution of (M,S) and (M',S') will be identical. $\square$

In particular, we use this lemma in the case where M is an estimator of S. Then this lemma implies that the probability that $M = S$ is equal to the probability that $M' = S'$.

Recall that $p_u(t) = \mathbb{P}[A_u(t) = S] = \max_{A \in \mathcal{F}_u(t)} \mathbb{P}[A = S]$. Hence we can apply this lemma (3.15) above to $A_u(t)$ and $A_{u'}(t)$:

Corollary 3.16. *If $B_t(G,u)$ and $B_t(G',u')$ are isomorphic then $p_u(t) = p_{u'}(t)$.*

3.4 δ-independence

To prove that agents learn S we will show that the agents must, over the duration of this process, gain access to a large number of measurements of S that are *almost* independent. To formalize the notion of almost-independence we define δ-independence and prove some easy results about it. The proofs in this subsection are relatively straightforward.

Let μ and ν be two measures defined on the same space. We denote the total variation distance between them by $d_{TV}(\mu, \nu)$. Let A and B be two random variables with joint distribution $\mu_{(A,B)}$. Then we denote by μ_A the marginal distribution of A, μ_B the marginal distribution of B, and $\mu_A \times \mu_B$ the product distribution of the marginal distributions.

Definition 3.17. *Let $(X_1, X_2, \ldots, X_k)$ be random variables. We refer to them as δ-**independent** if their joint distribution $\mu_{(X_1,\ldots,X_k)}$ has total variation distance of at most δ from the product of their marginal distributions $\mu_{X_1} \times \cdots \times \mu_{X_k}$:*

$$d_{TV}(\mu_{(X_1,\ldots,X_k)}, \mu_{X_1} \times \cdots \times \mu_{X_k}) \leq \delta.$$

Likewise, $(X_1, \ldots, X_l)$ are δ-**dependent** if the distance between the distributions is more than δ.

We remind the reader that a coupling ν, between two random variables A_1 and A_2 distributed ν_1 and ν_2, is a distribution on the product of the spaces ν_1, ν_2 such that the marginal of A_i is ν_i. The total variation distance between A_1 and A_2 is equal to the minimum, over all such couplings ν, of $\nu(A_1 \neq A_2)$.

Hence to prove that X, Y are δ-independent it is sufficient to show that there exists a coupling ν between ν_1, the joint distribution of (X, Y) and ν_2, the products of the marginal distributions of X and Y, such that $\nu((X_1, Y_1) \neq (X_2, Y_2)) \leq \delta$.

Alternatively, to prove that (A, B) are δ-independent, one could directly bound the total variation distance between $\mu_{(A,B)}$ and $\mu_A \times \mu_B$ by δ. This is often done below using the fact that the total variation distance satisfies the triangle inequality $d_{TV}(\mu, \nu) \leq d_{TV}(\mu, \gamma) + d_{TV}(\gamma, \nu)$.

We state and prove some straightforward claims regarding δ-independence.

Claim 3.18. *Let A, B and C be random variables such that $\mathbb{P}[A \neq B] \leq \delta$ and (B, C) are δ'-independent. Then (A, C) are $2\delta + \delta'$-independent.*

Proof. Let $\mu_{(A,B,C)}$ be a joint distribution of A, B and C such that $\mathbb{P}[A \neq B] \leq \delta$.

Since $\mathbb{P}[A \neq B] \leq \delta$, $\mathbb{P}[(A, C) \neq (B, C)] \leq \delta$, in both cases that A, B, C are picked from either $\mu_{(A,B,C)}$ or $\mu_{(A,B)} \times \mu_C$. Hence

$$d_{TV}(\mu_{(A,C)}, \mu_{(B,C)}) \leq \delta$$

and

$$d_{TV}(\mu_A \times \mu_C, \mu_B \times \mu_C) \leq \delta.$$

Since (B, C) are δ'-independent,

$$d_{TV}(\mu_B \times \mu_C, \mu_{(B,C)}) \leq \delta'.$$

The claim follows from the triangle inequality

$$\begin{aligned} d_{TV}(\mu_{(A,C)}, \mu_A \times \mu_C) &\leq d_{TV}(\mu_{(A,C)}, \mu_{(B,C)}) + d_{TV}(\mu_{(B,C)}, \mu_B \times \mu_C) + d_{TV}(\mu_B \times \mu_C, \mu_A \times \mu_C) \\ &\leq 2\delta + \delta'. \end{aligned}$$

□

Claim 3.19. *Let (X, Y) be δ-independent, and let $Z = f(Y, B)$ for some function f and B that is independent of both X and Y. Then (X, Z) are also δ-independent.*

Proof. Let $\mu_{(X,Y)}$ be a joint distribution of X and Y satisfying the conditions of the claim. Then since (X, Y) are δ-independent,

$$d_{TV}(\mu_{(X,Y)}, \mu_X \times \mu_Y) \leq \delta.$$

Since B is independent of both X and Y,

$$d_{TV}(\mu_{(X,Y)} \times \mu_B, \mu_X \times \mu_Y \times \mu_B) \leq \delta$$

and (X, Y, B) are δ-independent. Therefore there exists a coupling between $(X_1, Y_1, B_1) \sim \mu_{(X,Y)} \times \mu_B$ and $(X_2, Y_2, B_2) \sim \mu_X \times \mu_Y \times \mu_B$ such that $\mathbb{P}[(X_1, Y_1, B_1) \neq (X_2, Y_2, B_2)] \leq \delta$. Then

$$\mathbb{P}[(X_1, f(Y_1, B_1)) \neq (X_2, f(Y_2, B_2))] \leq \delta$$

and the proof follows. □

Claim 3.20. *Let $A = (A_1, \ldots, A_k)$, and X be random variables. Let $(A_1, \ldots, A_k)$ be δ_1-independent and let (A, X) be δ_2-independent. Then $(A_1, \ldots, A_k, X)$ are $(\delta_1 + \delta_2)$-independent.*

Proof. Let $\mu_{(A_1,\ldots,A_k,X)}$ be the joint distribution of $A = (A_1, \ldots, A_k)$ and X. Then since $(A_1, \ldots, A_k)$ are δ_1-independent,

$$d_{TV}(\mu_A, \mu_{A_1} \times \cdots \times \mu_{A_k}) \leq \delta_1.$$

Hence

$$d_{TV}(\mu_A \times \mu_X, \mu_{A_1} \times \cdots \times \mu_{A_k} \times \mu_X) \leq \delta_1.$$

Since (A, X) are δ_2-independent,

$$d_{TV}(\mu_{(A,X)}, \mu_A \times \mu_X) \leq \delta_2.$$

The claim then follows from the triangle inequality

$$d_{TV}(\mu_{(A,X)}, \mu_{A_1} \times \cdots \times \mu_{A_k} \times \mu_X) \leq d_{TV}(\mu_{(A,X)}, \mu_A \times \mu_X) + d_{TV}(\mu_A \times \mu_X, \mu_{A_1} \times \cdots \times \mu_{A_k} \times \mu_X).$$

□

Lemma 3.21. *For every $1/2 < p < 1$ there exist $\delta = \delta(p) > 0$ and $\eta = \eta(p) > 0$ such that if S and (X_1, X_2, X_3) are binary random variables with $\mathbb{P}[S = 1] = 1/2$, $1/2 < p - \eta \leq \mathbb{P}[X_i = S] < 1$, and (X_1, X_2, X_3) are δ-independent conditioned on S then $\mathbb{P}[a(X_1, X_2, X_3) = S] > p$, where a is the MAP estimator of S given (X_1, X_2, X_3).*

In other words, one's odds of guessing S using three conditionally almost-independent bits are greater than using a single bit.

Proof. We apply Lemma 3.22 below to three conditionally independent bits which are each equal to S w.p. at least $p-\eta$. Then

$$\mathbb{P}\left[a(X_1,X_2,X_3)=S\right]\geq p-\eta+\epsilon_{p-\eta}$$

where $\epsilon_q=\frac{1}{100}(2q-1)(3q^2-2q^3-q)$.

Since ϵ_q is continuous in q and positive for $1/2<q<1$, it follows that for η small enough $p-\eta+\epsilon_{p-\eta}>p$. Now, take $\delta<\epsilon_{p-\eta}-\eta$. Then, since we can couple δ-independent bits to independent bits so that they differ with probability at most δ, the claim follows. □

Lemma 3.22. *Let S and (X_1,X_2,X_3) be binary random variables such that $\mathbb{P}[S=1]=1/2$. Let $1/2<p\leq\mathbb{P}[X_i=S]<1$. Let $a(X_1,X_2,X_3)$ be the MAP estimator of S given (X_1,X_2,X_3). Then there exists an $\epsilon_p>0$ that depends only on p such that if (X_1,X_2,X_3) are independent conditioned on S then $\mathbb{P}[a(X_1,X_2,X_3)=S]\geq p+\epsilon_p$.*

In particular the statement holds with

$$\epsilon_p=\frac{1}{100}(2p-1)(3p^2-2p^3-p).$$

Proof. Denote $X=(X_1,X_2,X_3)$.

Assume first that $\mathbb{P}[X_i=S]=p$ for all i. Let $\delta_1,\delta_2,\delta_3$ be such that $p+\delta_i=\mathbb{P}[X_i=1|S=1]$ and $p-\delta_i=\mathbb{P}[X_i=0|S=0]$.

To show that $\mathbb{P}[a(X)=S]\geq p+\epsilon_p$ it is enough to show that $\mathbb{P}[b(X)=S]\geq p+\epsilon_p$ for some estimator b, by the definition of a MAP estimator. We separate into three cases.

1. If $\delta_1=\delta_2=\delta_3=0$ then the events $X_i=S$ are independent and the majority of the X_i's is equal to S with probability $p'=p^3+3p^2(1-p)$, which is greater than p for $\frac{1}{2}<p<1$. Denote $\eta_p=p'-p$. Then $\mathbb{P}[a(X)=S]\geq p+\eta_p$.

2. Otherwise if $|\delta_i|\leq\eta_p/6$ for all i then we can couple X to three bits $Y=(Y_1,Y_2,Y_3)$ which satisfy the conditions of case 1 above, and so that $\mathbb{P}[X\neq Y]\leq\eta_p/2$. Then $\mathbb{P}[a(X)=S]\geq p+\eta_p/2$.

3. Otherwise we claim that there exist i and j such that $|\delta_i+\delta_j|>\eta_p/12$.

   Indeed assume w.l.o.g. that $\delta_1\geq\eta_p/6$. Then if it doesn't hold that $\delta_1+\delta_2\geq\eta_p/12$ and it doesn't hold that $\delta_1+\delta_3\geq\eta_p/12$ then $\delta_2\leq-\eta_p/12$ and $\delta_3\leq-\eta_p/12$ and therefore $\delta_2+\delta_3\leq-\eta_p/12$.

   Now that this claim is proved, assume w.l.o.g. that $\delta_1+\delta_2\geq\eta_p/12$. Recall that $X_i\in\{0,1\}$, and so the product X_1X_2 is also an element of $\{0,1\}$. Then

   $$\begin{aligned}\mathbb{P}[X_1X_2=S]&=\tfrac{1}{2}\mathbb{P}[X_1X_2=1|S=1]+\tfrac{1}{2}\mathbb{P}[X_1X_2=0|S=0]\\&=\tfrac{1}{2}\left((p+\delta_1)(p+\delta_2)+(p-\delta_1)(p-\delta_2)+(p-\delta_1)(1-p+\delta_2)+(1-p+\delta_1)(p-\delta_2)\right)\\&=p+\tfrac{1}{2}(2p-1)(\delta_1+\delta_2)\\&\geq p+(2p-1)\eta_p/12,\end{aligned}$$

   and so $\mathbb{P}[a(X)=S]\geq p+(2p-1)\eta_p/12$.

Finally, we need to consider the case that $\mathbb{P}[X_i = S] = p_i > p$ for some i. We again consider two cases. Denote $\epsilon_p = (2p-1)\eta_p/100$. If there exists an i such that $p_i > \epsilon_p$ then this bit is by itself an estimator that equals S with probability at least $p + \epsilon_p$, and therefore the MAP estimator equals S with probability at least $p + \epsilon_p$.

Otherwise $p \le p_i \le p_i + \epsilon_p$ for all i. We will construct a coupling between the distributions of $X = (X_1, X_2, X_3)$ and $Y = (Y_1, Y_2, Y_3)$ such that the Y_i's are conditionally independent given S and $\mathbb{P}[Y_i = S] = p$ for all i, and furthermore $\mathbb{P}[Y \ne X] \le 3\epsilon_p$. By what we've proved so far the MAP estimator of S given Y equals S with probability at least $p + (2p-1)\eta_p/12 \ge p + 8\epsilon_p$. Hence by the coupling, the same estimator applied to X is equal to S with probability at least $p + 8\epsilon_p - 3\epsilon_p > p + \epsilon_p$.

To couple X and Y let Z_i be a real i.i.d. random variables uniform on $[0,1]$. When $S = 1$ let $X_i = Y_i = S$ if $Z_i > p_i + \delta_i$, let $X_i = S$ and $Y_i = 1 - S$ if $Z_i \in [p + \delta_i, p_i + \delta_i]$, and otherwise $X_i = Y_i = 1 - S$. The construction for $S = 0$ is similar. It is clear that X and Y have the required distribution, and that furthermore $\mathbb{P}[X_i \ne Y_i] = p_i - p \le \epsilon_p$. Hence $\mathbb{P}[X \ne Y] \le 3\epsilon_p$, as needed.

□

3.5 Asymptotic learning

In this section we prove Theorem 2.

Theorem (2). *Let μ_0, μ_1 be such that for every connected, undirected graph G there exists a random variable A such that almost surely $A_u = A$ for all $u \in V$. Then there exists a sequence $q(n) = q(n, \mu_0, \mu_1)$ such that $q(n) \to 1$ as $n \to \infty$, and $\mathbb{P}[A = \{S\}] \ge q(n)$, for any choice of undirected, connected graph G with n agents.*

To prove this theorem we will need a number of intermediate results, which are given over the next few subsections.

3.5.1 Estimating the limiting optimal action set A

We would like to show that although the agents have a common optimal action set A only at the limit $t \to \infty$, they can estimate this set well at a large enough time t.

The action $A_u(t)$ is agent u's MAP estimator of S at time t (see Remark 2.3). We likewise define $K_u(t)$ to be agent u's MAP estimator of A, at time t:

$$K_u(t) = \operatorname*{argmax}_{K \in \{\{0\},\{1\},\{0,1\}\}} \mathbb{P}[A = K | \mathcal{F}_u(t)]. \tag{1}$$

We show that the sequence of random variables $K_u(t)$ converges to A for every u, or that alternatively $K_u(t) = A$ for each agent u and t large enough:

Lemma 3.23. $\mathbb{P}[\lim_{t\to\infty} K_u(t) = A] = 1$ *for all $u \in V$.*

This lemma (3.23) follows by direct application of the more general Lemma 3.24 which we prove below. Note that a consequence is that $\lim_{t\to\infty} \mathbb{P}[K_u(t) = A] = 1$.

Lemma 3.24. *Let $\mathcal{K}_1 \subseteq \mathcal{K}_2, \ldots$ be a filtration of σ-algebras, and let $\mathcal{K}_\infty = \cup_t \mathcal{K}_t$. Let K be a random variable that takes a finite number of values and is measurable in $\mathcal{K}_\infty$. Let $M(t) = \operatorname{argmax}_k \mathbb{P}[K = k | \mathcal{K}(t)]$ be the MAP estimator of K given $\mathcal{K}_t$. Then*

$$\mathbb{P}\left[\lim_{t\to\infty} M(t) = K\right] = 1.$$

Proof. For each k in the support of K, $\mathbb{P}[K = k|\mathcal{K}_t]$ is a bounded martingale which converges almost surely to $\mathbb{P}[K = k|\mathcal{K}_\infty]$, which is equal to $\mathbf{1}(K = k)$, since K is measurable in G_∞. Therefore $M(t) = \operatorname{argmax}_k \mathbb{P}[K = k|\mathcal{K}_t]$ converges almost surely to $\operatorname{argmax}_k \mathbb{P}[K = k|\mathcal{K}_\infty] = K$. □

We would like at this point to provide the reader with some more intuition on $A_u(t)$, $K_u(t)$ and the difference between them. Assuming that $A = \{1\}$ then by definition, from some time t_0 on, $A_u(t) = 1$, and from Lemma 3.23, $K_u(t) = \{1\}$. The same applies when $A = \{0\}$. However, when $A = \{0, 1\}$ then $A_u(t)$ may take both values 0 and 1 infinitely often, but $K_u(t)$ will eventually equal $\{0, 1\}$. That is, agent u will realize at some point that, although it thinks at the moment that 1 is preferable to 0 (for example), it is in fact the most likely outcome that its belief will converge to 1/2. In this case, although it is not optimal, a *uniformly random* guess of which is the best action may not be so bad. Our next definition is based on this observation.

Based on $K_u(t)$, we define a second "action" $C_u(t)$.

Definition 3.25. *Let $C_u(t)$ be picked uniformly from $K_u(t)$: if $K_u(t) = \{1\}$ then $C_u(t) = 1$, if $K_u(t) = \{0\}$ then $C_u(t) = 0$, and if $K_u(t) = \{0, 1\}$ then $C_u(t)$ is picked independently from the uniform distribution over $\{0, 1\}$.*

Note that we here extend our probability space by including in $I_u(t)$ (the observations of agent u up to time t) an extra uniform bit that is independent of all else and S in particular. Hence this does not increase u's ability to estimate S, and if we can show that in this setting u learns S then u can also learn S without this bit. In fact, we show that asymptotically it is as good an estimate for S as the best estimate $A_u(t)$:

Claim 3.26. $\lim_{t\to\infty} \mathbb{P}[C_u(t) = S] = \lim_{t\to\infty} \mathbb{P}[A_u(t) = S] = p$ *for all* u.

Proof. We prove the claim by showing that it holds both when conditioning on the event $A = \{0, 1\}$ and when conditioning on its complement.

When $A \neq \{0, 1\}$ then for t large enough $A = \{A_u(t)\}$. Since (by Lemma 3.23) $\lim K_u(t) = A$ with probability 1, in this case $C_u(t) = A_u(t)$ for t large enough, and

$$\lim_{t\to\infty} \mathbb{P}[C_u(t) = S|A \neq \{0, 1\}] \ = \ \mathbb{P}[A = \{S\}|A \neq \{0, 1\}] \ = \ \lim_{t\to\infty} \mathbb{P}[A_u(t) = S|A \neq \{0, 1\}].$$

When $A = \{0, 1\}$ then $\lim X_u(t) = \lim \mathbb{P}[A_u(t) = S|\mathcal{F}_u(t)] = 1/2$ and so $\lim \mathbb{P}[A_u(t) = S] = 1/2$. This is again also true for $C_u(t)$, since in this case it is picked at random for t large enough, and so

$$\lim_{t\to\infty} \mathbb{P}[C_u(t) = S|A = \{0, 1\}] \ = \ \frac{1}{2} \ = \ \lim_{t\to\infty} \mathbb{P}[A_u(t) = S|A = \{0, 1\}].$$

□

3.5.2 The probability of getting it right

Recall Definition 3.3: $p_u(t) = \mathbb{P}[A_u(t) = S]$ and $p_u = \lim_{t\to\infty} p_u(t)$ (i.e., $p_u(t)$ is the probability that agent u takes the right action at time t). We prove here a few easy related claims that will later be useful to us.

Claim 3.27. $p_u(t + 1) \geq p_u(t)$.

Proof. Condition on $\mathcal{F}_u(t+1)$, the information available to agent u at time $t+1$. Hence the probability that $A_u(t+1) = S$ is at least as high as the probability $A_u(t) = S$, since

$$A_u(t+1) = \operatorname*{argmax}_s \mathbb{P}\left[S = s | \mathcal{F}(t+1)\right]$$

and $A_u(t)$ is measurable in $\mathcal{F}(t+1)$. The claim is proved by integrating over all possible values of $\mathcal{F}_u(t+1)$. □

Since $p_u(t)$ is bounded by one, Claim 3.27 means that the limit p_u exists. We show that this value is the same for all vertices.

Claim 3.28. *There exists a $p \in [0,1]$ such that $p_u = p$ for all u.*

Proof. Let u and w be neighbors. As in the proof above, we can argue that $\mathbb{P}\left[A_u(t+1) = S | \mathcal{F}_u(t+1)\right] \geq \mathbb{P}\left[A_w(t) = S | \mathcal{F}_u(t+1)\right]$, since $A_w(t)$ is measurable in $\mathcal{F}_u(t+1)$. Hence the same holds unconditioned, and so we have that $p_u \geq p_w$, by taking the limit $t \to \infty$. Since the same argument can be used with the roles of u and w reversed, we have that $p_u = p_w$, and the claim follows from the connectedness of the graph, by induction. □

We make the following definition in the spirit of these claims:

Definition 3.29. $p = \lim_{t\to\infty} \mathbb{P}\left[A_u(t) = S\right]$.

In the context of a specific social network graph G we may denote this quantity as $p(G)$.

For time $t = 1$ the next standard claim follows from the fact that the agents' signals are informative.

Claim 3.30. $p_u(t) > 1/2$ *for all u and t.*

Proof. Note that

$$\mathbb{P}\left[A_u(1) = S | W_u\right] = \max\{X_u(1), 1 - X_u(1)\} = \max\{\mathbb{P}\left[S = 0 | W_u\right], \mathbb{P}\left[S = 1 | W_u\right]\}.$$

Recall that $p_u(1) = \mathbb{P}\left[A_u(1) = S\right]$. Hence

$$\begin{aligned} p_u(1) &= \mathbb{E}\left[\mathbb{P}\left[A_u(1) = S | W_u\right]\right] \\ &= \mathbb{E}\left[\max\{\mathbb{P}\left[S = 0 | W_u\right], \mathbb{P}\left[S = 1 | W_u\right]\}\right] \end{aligned}$$

Since $\max\{a, b\} = \frac{1}{2}(a+b) + \frac{1}{2}|a-b|$, and since $\mathbb{P}\left[S = 0 | W_u\right] + \mathbb{P}\left[S = 1 | W_u\right] = 1$, it follows that

$$\begin{aligned} p_u(1) &= \tfrac{1}{2} + \tfrac{1}{2}\mathbb{E}\left[\left|\mathbb{P}\left[S = 0 | W_u\right] - \mathbb{P}\left[S = 1 | W_u\right]\right|\right] \\ &= \tfrac{1}{2} + \tfrac{1}{2} D_{TV}(\mu_0, \mu_1), \end{aligned}$$

where the last equality follows by Bayes' rule. Since $\mu_0 \neq \mu_1$, the total variation distance $D_{TV}(\mu_0, \mu_1) > 0$ and $p_u(1) > \frac{1}{2}$. For $t > 1$ the claim follows from Claim 3.27 above. □

Recall that $|N(u)|$ is the out-degree of u, or the number of neighbors that u observes. The next lemma states that an agent with many neighbors will have a good estimate of S already at the second round, after observing the first action of its neighbors. This lemma is adapted from Mossel and Tamuz [13], and provided here for completeness.

Lemma 3.31. *There exist constants $C_1 = C_1(\mu_0, \mu_1)$ and $C_2 = C_2(\mu_0, \mu_1)$ such that for any agent u it holds that*

$$p_u(2) \geq 1 - C_1 e^{-C_2 \cdot N(u)}.$$

Proof. Conditioned on S, private signals are independent and identically distributed. Since $A_w(1)$ is a deterministic function of W_w, the initial actions $A_w(1)$ are also identically distributed, conditioned on S. Hence there exists a q such that $p_w(1) = \mathbb{P}[A_w(t) = S] = q$ for all agents w. By Lemma 3.30 above, $q > 1/2$. Therefore

$$\mathbb{P}[A_w(1) = 1 | S = 1] \neq \mathbb{P}[A_w(1) = 1 | S = 0],$$

and the distribution of $A_w(1)$ is different when conditioned on $S = 0$ or $S = 1$.

Fix an agent u, and let $n = |N(u)|$ be the out-degree of u, or the number of neighbors that it observes. Let $\{w_1, \ldots, w_{|N(u)|}\}$ be the set of u's neighbors. Recall that $A_u(2)$ is the MAP estimator of S given $(A_{w_1}(1), \ldots, A_{w_n}(1))$, and given u's private signal.

By standard asymptotic statistics of hypothesis testing (cf. [7]), testing an hypothesis (in our case, say, $S = 1$ vs. $S = 0$) given n informative, conditionally i.i.d. signals, succeeds except with probability that is exponentially low in n. It follows that $\mathbb{P}[A_u(2) \neq S]$ is exponentially small in n, so that there exist C_1 and C_2 such that

$$p_u(2) = \mathbb{P}[A_u(2) = S] \geq 1 - C_1 e^{-C_2 \cdot N(u)}.$$

□

The following claim is a direct consequence of the previous lemmas of this section.

Claim 3.32. *Let $d(G) = \sup_u \{N(u)\}$ be the out-degree of the graph G; note that for infinite graphs it may be that $d = \infty$. Then there exist constants $C_1 = C_1(\mu_0, \mu_1)$ and $C_2 = C_2(\mu_0, \mu_1)$ such that*

$$p(G) \geq 1 - C_1 e^{-C_2 \cdot d(G)}$$

for all agents u.

Proof. Let u be an arbitrary vertex in G. Then by Lemma 3.31 it holds that

$$p_u(2) \geq 1 - C_1 e^{-C_2 \cdot N(u)},$$

for some constants C_1 and C_2. By Lemma 3.27 we have that $p_u(t+1) \geq p_u(t)$, and therefore

$$p_u = \lim_{n \to \infty} p_u(t) \geq 1 - C_1 e^{-C_2 \cdot N(u)}.$$

Finally, $p(G) = p_u$ by Lemma 3.28, and so

$$p_u \geq 1 - C_1 e^{-C_2 \cdot N(u)}.$$

Since this holds for an arbitrary vertex u, the claim follows. □

3.5.3 Local limits and pessimal graphs

We now turn to apply local limits to our process. We consider here and henceforth the same model of Definitions 2.1 and 2.2, as applied, with the same private signals, to different graphs. We write $p(G)$ for the value of p on the process on G, $A(G)$ for the value of A on G, etc.

Lemma 3.33. *Let* $(G, u) = \lim_{r\to\infty}(G_r, u_r)$*. Then* $p(G) \leq \liminf_r p(G_r)$*.*

Proof. Since $B_r(G_r, u_r) \cong B_r(G, u)$, by Lemma 3.16 we have that $p_u(r) = p_{u_r}(r)$. By Claim 3.27 $p_{u_r}(r) \leq p(G_r)$, and therefore $p_u(r) \leq p(G_r)$. The claim follows by taking the lim inf of both sides. □

A particularly interesting case in the one the different G_r's are all the same graph:

Corollary 3.34. *Let G be a (perhaps infinite) graph, and let $\{u_r\}$ be a sequence of vertices. Then if the local limit* $(H, u) = \lim_{r\to\infty}(G, u_r)$ *exists then* $p(H) \leq p(G)$*.*

Recall that $\mathcal{B}_d$ denotes the set of infinite, connected, undirected graphs of degree at most d. Let

$$\mathcal{B} = \bigcup_d \mathcal{B}_d.$$

Definition 3.35. *Let*

$$p^* = p^*(\mu_0, \mu_1) = \inf_{G\in\mathcal{B}} p(G)$$

be the probability of learning in the pessimal graph.

Note that by Claim 3.30 we have that $p^* > 1/2$. We show that this infimum is in fact attained by some graph:

Lemma 3.36. *There exists a graph* $H \in \mathcal{B}$ *such that* $p(H) = p^*$*.*

Proof. Let $\{G_r = (V_r, E_r)\}_{r=1}^\infty$ be a series of graphs in $\mathcal{B}$ such that $\lim_{r\to\infty} p(G_r) = p^*$. Note that $\{G_r\}$ must all be in $\mathcal{B}_d$ for some d (i.e., have uniformly bounded degrees), since otherwise the sequence $p(G_r)$ would have values arbitrarily close to 1 and its limit could not be p^* (unless indeed $p^* = 1$, in which case our main Theorem 2 is proved). This follows from Lemma 3.31.

We now arbitrarily mark a vertex u_r in each graph, so that $u_r \in V_r$, and let (H, u) be the limit of some subsequence of $\{G_r, u_r\}_{r=1}^\infty$. Since $\mathcal{B}_d$ is compact (Lemma 3.12), (H, u) is guaranteed to exist, and $H \in \mathcal{B}_d$.

By Lemma 3.33 we have that $p(H) \leq \liminf_r p(G_r) = p^*$. But since $H \in \mathcal{B}$, $p(H)$ cannot be less than p^*, and the claim is proved. □

3.5.4 Independent bits

We now show that on infinite graphs, the private signals in the neighborhood of agents that are "far enough away" are (conditioned on S) almost independent of A (the final consensus estimate of S).

Lemma 3.37. *Let G be an infinite graph. Fix a vertex u_0 in G. Then for every $\delta > 0$ there exists an r_δ such that for every $r \geq r_\delta$ and every vertex u with $d(u_0, u) > 2r$ it holds that $W(B_r(G, u))$, the private signals in $B_r(G, u)$, are δ-independent of A, conditioned on S.*

Here we denote graph distance by $d(\cdot,\cdot)$.

Proof. Fix u_0, and let u be such that $d(u_0,u) > 2r$. Then $B_r(G,u_0)$ and $B_r(G,u)$ are disjoint, and hence independent conditioned on S. Hence $K_{u_0}(r)$ is independent of $W(B_r(G,u))$, conditioned on S.

Lemma 3.23 states that $\mathbb{P}[\lim_{r\to\infty} K_{u_0}(r) = A] = 1$, and so there exists an r_δ such that for every $r \geq r_\delta$ it holds that $\mathbb{P}[K_{u_0}(r) = A] > 1 - \frac{1}{2}\delta$.

Recall Claim 3.18: for any A, B, C, if $\mathbb{P}[A = B] = 1 - \frac{1}{2}\delta$ and B is independent of C, then (A, C) are δ-independent.

Applying Claim 3.18 to A, $K_{u_0}(r)$ and $W(B_r(G,u))$ we get that for any r greater than r_δ it holds that $W(B_r(G,u))$ is δ-independent of A, conditioned on S. □

We will now show, in the lemmas below, that in infinite graphs each agent has access to any number of "good estimators": δ-independent measurements of S that are each almost as likely to equal S as p^*, the minimal probability of estimating S on any infinite graph.

Definition 3.38. *We say that agent $u \in G$ has k (δ,ϵ)-**good estimators** if there exists a time t and estimators $M_1,\ldots,M_k$ such that $(M_1,\ldots,M_k) \in \mathcal{F}_u(t)$ and*

1. *$\mathbb{P}[M_i = S] > p^* - \epsilon$ for $1 \leq i \leq k$.*
2. *$(M_1,\ldots,M_k)$ are δ-independent, conditioned on S.*

Claim 3.39. *Let P denote the property of having k (δ,ϵ)-good estimators. Then P is a* local property *(Definition 3.10) of the rooted graph (G,u). Furthermore, if $u \in G$ has k (δ,ϵ)-good estimators measurable in $\mathcal{F}_u(t)$ then $(G,u) \in P^{(t)}$, i.e., (G,u) has property P with radius t.*

Proof. If $(G,u) \in P$ then by definition there exists a time t such that $(M_1,\ldots,M_k) \in \mathcal{F}_u(t)$. Hence by Lemma 3.15, if $B_t(G,u) \cong B_t(G',u')$ then $u' \in G'$ also has k (δ,ϵ)-good estimators $(M'_1,\ldots,M'_k) \in \mathcal{F}_{u'}(t)$ and $(G',u') \in P$. In particular, $(G,u) \in P^{(t)}$, i.e., (G,u) has property P with radius t. □

We are now ready to prove the main lemma of this subsection:

Lemma 3.40. *For every $d \geq 2$, $G \in \mathcal{B}_d$, $\epsilon,\delta > 0$ and $k \geq 0$ there exists a vertex u, such that u has k (δ,ϵ)-good estimators.*

Informally, this lemma states that if G is an infinite graph with bounded degrees, then there exists an agent that eventually has k almost-independent estimates of S with quality close to p^*, the minimal probability of learning.

Proof. In this proof we use the term "independent" to mean "independent conditioned on S".

We choose an arbitrary d and prove by induction on k. The basis $k = 0$ is trivial. Assume the claim holds for k, any $G \in \mathcal{B}_d$ and all $\epsilon,\delta > 0$. We shall show that it holds for $k+1$, any $G \in \mathcal{B}_d$ and any $\delta,\epsilon > 0$.

By the inductive hypothesis for every $G \in \mathcal{B}_d$ there exists a vertex in G that has k $(\delta/100,\epsilon)$-good estimators $(M_1,\ldots,M_k)$.

Now, having k $(\delta/100,\epsilon)$-good estimators is a *local property* (Claim 3.39). We now therefore apply Lemma 3.13: since every graph $G \in \mathcal{B}_d$ has a vertex with k $(\delta/100,\epsilon)$-good estimators, any graph $G \in \mathcal{B}_d$ has a time t_k for which infinitely many distinct vertices $\{w_r\}$ have k $(\delta/100, \epsilon)$-good estimators measurable at time t_k.

In particular, if we fix an arbitrary $u_0 \in G$ then for every r there exists a vertex $w \in G$ that has k $(\delta/100, \epsilon)$-good estimators and whose distance $d(u_0, w)$ from u_0 is larger than r.

We shall prove the lemma by showing that for a vertex w that is far enough from u_0 which has $(\delta/100, \epsilon)$-good estimators $(M_1, \ldots, M_k)$, it holds that for a time t_{k+1} large enough $(M_1, \ldots, M_k, C_w(t_{k+1}))$ are (δ, ϵ)-good estimators.

By Lemma 3.37 there exists an r_δ such that if $r > r_\delta$ and $d(u_0, w) > 2r$ then $W(B_r(G, w))$ is $\delta/100$-independent of A. Let $r^* = \max\{r_\delta, t_k\}$, where t_k is such that there are infinitely many vertices in G with k good estimators measurable at time t_k.

Let w be a vertex with k $(\delta/100, \epsilon)$-good estimators $(M_1, \ldots, M_k)$ at time t_k, such that $d(u_0, w) > 2r^*$. Denote

$$\bar{M} = (M_1, \ldots, M_k).$$

Since $d(u_0, w) > 2r_\delta$, $W(B_{r^*}(G, w))$ is $\delta/100$-independent of A, and since $B_{t_k}(G, w) \subseteq B_{r^*}(G, w)$, $W(B_{t_k}(G, w))$ is $\delta/100$-independent of A. Finally, since $\bar{M} \in \mathcal{F}_w(t_k)$, $\bar{M}$ is a function of $W(B_{t_k}(G, w))$, and so by Claim 3.19 we have that $\bar{M}$ is also $\delta/100$-independent of A.

For t_{k+1} large enough it holds that

- $K_w(t_{k+1})$ is equal to L with probability at least $1 - \delta/100$, since
$$\lim_{t \to \infty} \mathbb{P}[K_w(t) = A] = 1,$$
by Claim 3.23.
- Additionally, $\mathbb{P}[C_w(t_{k+1}) = S] > p^* - \epsilon$, since
$$\lim_{t \to \infty} \mathbb{P}[C_w(t) = S] = p \geq p^*,$$
by Claim 3.26.

We have then that $(\bar{M}, A)$ are $\delta/100$-independent and $\mathbb{P}[K_w(t_{k+1}) \neq A] \leq \delta/100$. Claim 3.18 states that if (A, B) are δ-independent $\mathbb{P}[B \neq C] \leq \delta'$ then (A, C) are $\delta + 2\delta'$-independent. Applying this here we get that $(\bar{M}, K_w(t_{k+1}))$ are $\delta/25$-independent.

It follows by application of Claim 3.20 that $(M_1, \ldots, M_k, K_w(t_{k+1}))$ are δ-independent. Since $C_w(t_{k+1})$ is a function of $K_w(t_{k+1})$ and an independent bit, it follows by another application of Claim 3.19 that $(M_1, \ldots, M_k, C_w(t_{k+1}))$ are also δ-independent.

Finally, since $\mathbb{P}[C_w(t_{k+1}) = S] > p^* - \epsilon$, w has the $k+1$ (δ, ϵ)-good estimators $(M_1, \ldots, C_w(t_{k+1}))$ and the proof is concluded.

□

3.5.5 Asymptotic learning

As a tool in the analysis of finite graphs, we would like to prove that in infinite graphs the agents learn the correct state of the world almost surely.

Theorem 3.41. *Let $G = (V, E)$ be an infinite, connected undirected graph with bounded degrees (i.e., G is a general graph in $\mathcal{B}$). Then $p(G) = 1$.*

Note that an alternative phrasing of this theorem is that $p^* = 1$.

Proof. Assume the contrary, i.e. $p^* < 1$. Let H be an infinite, connected graph with bounded degrees such that $p(H) = p^*$, such as we've shown exists in Lemma 3.36.

By Lemma 3.40 there exists for arbitrarily small $\epsilon, \delta > 0$ a vertex $w \in H$ that has access at some time T to three δ-independent estimators (conditioned on S), each of which is equal to S with probability at least $p^* - \epsilon$. By Claims 3.21 and 3.30, the MAP estimator of S using these estimators equals S with probability higher than p^*, for the appropriate choice of low enough ϵ, δ. Therefore, since w's action $A_w(T)$ is the MAP estimator of S, its probability of equaling S is $\mathbb{P}[A_w(T) = S] > p^*$ as well, and so $p(H) > p^*$ - contradiction. □

Using Theorem 3.41 we prove Theorem 2, which is the corresponding theorem for finite graphs:

Theorem (2)**.** *Let μ_0, μ_1 be such that for every connected, undirected graph G there exists a random variable A such that almost surely $A_u = A$ for all $u \in V$. Then there exists a sequence $q(n) = q(n, \mu_0, \mu_1)$ such that $q(n) \to 1$ as $n \to \infty$, and $\mathbb{P}[A = \{S\}] \geq q(n)$, for any choice of undirected, connected graph G with n agents.*

Proof. Assume the contrary. Then there exists a series of graphs $\{G_r\}$ with r agents such that $\lim_{r\to\infty} \mathbb{P}[A(G_r) = \{S\}] < 1$, and so also $\lim_{r\to\infty} p(G_r) < 1$.

By the same argument of Theorem 3.41 these graphs must all be in $\mathcal{B}_d$ for some d, since otherwise, by Lemma 3.32, there would exist a subsequence of graphs $\{G_{r_d}\}$ with degree at least d and $\lim_{d\to\infty} p(G_{r_d}) = 1$. Since $\mathcal{B}_d$ is compact (Lemma 3.12), there exists a graph $(G, u) \in \mathcal{B}_d$ that is the limit of a subsequence of $\{(G_r, u_r)\}_{r=1}^{\infty}$.

Since G is infinite and of bounded degree, it follows by Theorem 3.41 that $p(G) = 1$, and in particular $\lim_{r\to\infty} p_u(r) = 1$. As before, $p_{u_r}(r) = p_u(r)$, and therefore $\lim_{r\to\infty} p_{u_r}(r) = 1$. Since $p(G_r) \geq p_{u_r}(r)$, $\lim_{r\to\infty} p(G_r) = 1$, which is a contradiction. □

3.6 Convergence to identical optimal action sets

In this section we prove Theorem 1.

Theorem (1)**.** *Let (μ_0, μ_1) induce non-atomic beliefs. Then there exists a random variable A such that almost surely $A_u = A$ for all u.*

In this section we shall assume henceforth that the distribution of initial private beliefs is non-atomic.

3.6.1 Previous work

The following theorem is due to Gale and Kariv [9]. Given two agents u and w, let E_u^0 denote the event that $A_u(t)$ equals 0 infinitely often E_w^1 and the event that $A_w(t)$ equals 1 infinitely often.

Theorem 3.42 (Gale and Kariv)**.** *If agent u observes agent w's actions then*

$$\mathbb{P}\left[E_u^0, E_w^1\right] = \mathbb{P}\left[X_u = 1/2, E_u^0, E_w^1\right].$$

I.e., if agent u takes action 0 infinitely often, agent w takes action 1 infinitely, and u observes w then u's belief is 1/2 at the limit, almost surely.

Corollary 3.43. *If agent u observes agent w's actions, and w takes both actions infinitely often then $X_u = 1/2$.*

Proof. Assume by contradiction that $X_u < 1/2$. Then u takes action 0 infinitely often. Therefore Theorem 3.42 implies that $X_u = 1/2$ - contradiction.

The case where $X_u > 1/2$ is treated similarly. □

3.6.2 Limit log-likelihood ratios

Denote

$$Y_u(t) = \log \frac{\mathbb{P}\left[I_u(t)\middle|S = 1, \bar{A}_u(t)\right]}{\mathbb{P}\left[I_u(t)\middle|S = 0, \bar{A}_u(t)\right]}.$$

In the next claim we show that $Z_u(t)$, the log-likelihood ratio inspired by u's observations up to time t, can be written as the sum of two terms: $Z_u(1) = \frac{d\mu_1}{d\mu_0}(W_u)$, which is the log-likelihood ratio inspired by u's private signal W_u, and $Y_u(t)$, which depends only on the actions of u and its neighbors, and does not depend directly on W_u.

Claim 3.44.

$$Z_u(t) = Z_u(1) + Y_u(t).$$

Proof. By definition we have that

$$Z_u(t) = \log \frac{\mathbb{P}\left[S = 1|\mathcal{F}_u(t)\right]}{\mathbb{P}\left[S = 0|\mathcal{F}_u(t)\right]} = \log \frac{\mathbb{P}\left[S = 1|I_u(t), W_u\right]}{\mathbb{P}\left[S = 0|I_u(t), W_u\right]}.$$

and by the law of conditional probabilities

$$\begin{aligned}
Z_u(t) &= \log \frac{\mathbb{P}\left[I_u(t)|S = 1, W_u\right]\mathbb{P}\left[W_u|S = 1\right]}{\mathbb{P}\left[I_u(t)|S = 0, W_u\right]\mathbb{P}\left[W_u|S = 0\right]} \\
&= \log \frac{\mathbb{P}\left[I_u(t)|S = 1, W_u\right]}{\mathbb{P}\left[I_u(t)|S = 0, W_u\right]} + Z_u(1).
\end{aligned}$$

Now $I_u(t)$, the actions of the neighbors of u up to time t, are a deterministic function of $W(B_t(G, u))$, the private signals in the ball of radius t around u, by Claim 3.14. Conditioned on S these are all independent, and so, from the definition of actions, these actions depend on u's private signal W_u only in as much as it affects the actions of u. Hence

$$\mathbb{P}\left[I_u(t)|S = s, W_u\right] = \mathbb{P}\left[I_u(t)\middle|S = s, \bar{A}_u(t)\right],$$

and therefore

$$\begin{aligned}
Z_u(t) &= \log \frac{\mathbb{P}\left[I_u(t)\middle|S = 1, \bar{A}_u(t)\right]}{\mathbb{P}\left[I_u(t)\middle|S = 0, \bar{A}_u(t)\right]} + Z_u(1) \\
&= Z_u(1) + Y_u(t).
\end{aligned}$$

□

Note that $Y_u(t)$ is a deterministic function of $I_u(t)$ and $\bar{A}_u(t)$.

Following our notation convention, we define $Y_u = \lim_{t\to\infty} Y_u(t)$. Note that this limit exists almost surely since the limit of $Z_u(t)$ exists almost surely. The following claim follows directly from the definitions:

Claim 3.45. *Y_u is measurable in $(\bar{A}_u, I_u)$, the actions of u and its neighbors.*

3.6.3 Convergence of actions

The event that an agent takes both actions infinitely often is (almost surely) a sufficient condition for convergence to belief 1/2. This follows from the fact that these actions imply that its belief takes values both above and below 1/2 infinitely many times. We show that it is also (almost surely) a necessary condition. Denote by E_u^a the event that u takes action a infinitely often.

Theorem 3.46.

$$\mathbb{P}\left[E_u^0 \cap E_u^1, X_u = 1/2\right] = \mathbb{P}\left[X_u = 1/2\right].$$

I.e., it a.s. holds that $X_u = 1/2$ iff u takes both actions infinitely often.

Proof. We'll prove the claim by showing that $\mathbb{P}\left[\neg(E_u^0 \cap E_u^1), X_u = 1/2\right] = 0$, or equivalently that $\mathbb{P}\left[\neg(E_u^0 \cap E_u^1), Z_u = 0\right] = 0$ (recall that $Z_u = \log X_u/(1 - X_u)$ and so $X_u = 1/2 \Leftrightarrow Z_u = 0$).

Let $\bar{a} = (a(1), a(2), \ldots)$ be a sequence of actions, and denote by W_{-u} the private signals of all agents except u. Conditioning on W_{-u} and S we can write:

$$\begin{aligned}\mathbb{P}\left[\bar{A}_u = \bar{a}, Z_u = 0\right] &= \mathbb{E}\left[\mathbb{P}\left[\bar{A}_u = \bar{a}, Z_u = 0 \middle| W_{-u}, S\right]\right] \\ &= \mathbb{E}\left[\mathbb{P}\left[\bar{A}_u = \bar{a}, Z_u(1) = -Y_u \middle| W_{-u}, S\right]\right]\end{aligned}$$

where the second equality follows from Claim 3.44. Note that by Claim 3.45 Y_u is fully determined by $\bar{A}_u$ and W_{-u}. We can therefore write

$$\begin{aligned}\mathbb{P}\left[\bar{A}_u = \bar{a}, Z_u = 0\right] &= \mathbb{E}\left[\mathbb{P}\left[\bar{A}_u = \bar{a}, Z_u(1) = -Y_u(W_{-u}, \bar{a}) \middle| W_{-u}, S\right]\right] \\ &\leq \mathbb{E}\left[\mathbb{P}\left[Z_u(1) = -Y_u(W_{-u}, \bar{a}) | W_{-u}, S\right]\right]\end{aligned}$$

Now, conditioned on S, the private signal W_u is distributed μ_S and is independent of W_{-u}. Hence its distribution when further conditioned on W_{-u} is still μ_S. Since $Z_u(1) = \log \frac{d\mu_1}{d\mu_0}(W_u)$, its distribution is also unaffected, and in particular is still non-atomic. It therefore equals $-Y_u(W_{-u}, \bar{a})$ with probability zero, and so

$$\mathbb{P}\left[\bar{A}_u = \bar{a}, Z_u = 0\right] = 0.$$

Since this holds for all sequences of actions $\bar{a}$, it holds in particular for all sequences which converge. Since there are only countably many such sequences, the probability that the action converges (i.e., $\neg(E_u^0 \cap E_u^1)$) and $Z_u = 0$ is zero, or

$$\mathbb{P}\left[\neg(E_u^0 \cap E_u^1), Z_u = 0\right] = 0.$$

□

Hence it impossible for an agent's belief to converge to 1/2 and for the agent to only take one action infinitely often. A direct consequence of this, together with Thm. 3.42, is the following corollary:

Corollary 3.47. *The union of the following three events occurs with probability one:*

1. $\forall u \in V : \lim_{t\to\infty} A_u(t) = S$. *Equivalently, all agents converge to the correct action.*
2. $\forall u \in V : \lim_{t\to\infty} A_u(t) = 1 - S$. *Equivalently, all agents converge to the wrong action.*

3. $\forall u \in V : X_u = 1/2$, *and in this case all agents take both actions infinitely often and hence don't converge at all.*

Proof. Consider first the case that there exists a vertex u such that u takes both actions infinitely often. Let w be a vertex that observes u. Then by Corollary 3.43 we have that $X_w = 1/2$, and by Theorem 3.46 w also takes both actions infinitely often. Continuing by induction and using the fact that the graph is strongly connected we obtain the third case that none of the agents converge and $X_u = 1/2$ for all u.

It remains to consider the case that all agents' actions converge to either 0 or 1. Using strong connectivity, to prove the theorem it suffices to show that it cannot be the case that w observes u and they converge to different actions. In this case, by Corollary 3.43 we have that $X_w = 1/2$, and then by Theorem 3.46 agent w's actions do not converge - contradiction. □

Theorem 1 is an easy consequence of this theorem. Recall that $A_u = \{1\}$ when $X_u > 1/2$, $A_u = \{0\}$ when $X_u < 1/2$ and $A_u = \{0, 1\}$ when $X_u = 1/2$.

Theorem (1)**.** *Let* (μ_0, μ_1) *induce non-atomic beliefs. Then there exists a random variable* A *such that almost surely* $A_u = A$ *for all* u*.*

Proof. Fix an agent v. When $X_v < 1/2$ (resp. $X_v > 1/2$) then the first (resp. second) case of corollary 3.47 occurs and $A = \{0\}$ (resp. $A = \{1\}$). Likewise when $X_v = 1/2$ then the third case occurs, $X_u = 1/2$ for all $u \in V$ and $A_u = \{0, 1\}$ for all $u \in V$. □

3.7 Extension to L-locally connected graphs

The main result of this article, Theorem 2, is a statement about undirected graphs. We can extend the proof to a larger family of graphs, namely, L-locally connected graphs.

Definition 3.48. *Let* $G = (V, E)$ *be a directed graph.* G *is* L*-locally strongly connected if, for each* $(u, w) \in E$*, there exists a path in* G *of length at most* L *from* w *to* u*.*

Theorem 2 can be extended as follows.

Theorem 3.49. *Fix* L*, a positive integer. Let* μ_0, μ_1 *be such that for every strongly connected, directed graph* G *there exists a random variable* A *such that almost surely* $A_u = A$ *for all* $u \in V$*. Then there exists a sequence* $q(n) = q(n, \mu_0, \mu_1)$ *such that* $q(n) \to 1$ *as* $n \to \infty$*, and* $\mathbb{P}[A = \{S\}] \geq q(n)$*, for any choice of* L*-locally strongly connected graph* G *with* n *agents.*

The proof of Theorem 3.49 is essentially identical to the proof of Theorem 2. The latter is a consequence of Theorem 3.41, which shows learning in bounded degree infinite graphs, and of Lemma 3.32, which implies asymptotic learning for sequences of graphs with diverging maximal degree.

Note first that the set of L-locally strongly connected rooted graphs with degrees bounded by d is compact. Hence the proof of Theorem 3.41 can be used as is in the L-locally strongly connected setup.

In order to apply Lemma 3.32 in this setup, we need to show that when in-degrees diverge then so do out-degrees. For this note that if (u, v) is a directed edge then u is in the (directed) ball of radius L around v. Hence, if there exists a vertex v with in-degree D then in the ball of radius L around it there are at least D vertices. On the other hand, if the out-degree is bounded by d, then the number of vertices in this ball is at most $L \cdot d^L$. Therefore, $d \to \infty$ as $D \to \infty$.

A Example of Non-atomic private beliefs leading to non-learning

We sketch an example in which private beliefs are atomic and asymptotic learning does not occur.

Example A.1. *Let the graph G be the undirected chain of length n, so that $V = \{1, \dots, n\}$ and (u, v) is an edge if $|u - v| = 1$. Let the private signals be bits that are each independently equal to S with probability $2/3$. We choose here the tie breaking rule under which agents defer to their original signals*[2].

We leave the following claim as an exercise to the reader.

Claim A.2. *If an agent u has at least one neighbor with the same private signal (i.e., $W_u = W_v$ for v a neighbor of u) then u will always take the same action $A_u(t) = W_u$.*

Since this happens with probability that is independent of n, with probability bounded away from zero an agent will always take the wrong action, and so asymptotic learning does not occur. It is also clear that optimal action sets do not become common knowledge, and these fact are indeed related.

References

[1] Aldous, D., Steele, J.: The objective method: Probabilistic combinatorial optimization and local weak convergence. Probability on Discrete Structures (Volume 110 of Encyclopaedia of Mathematical Sciences), ed. H. Kesten **110**, 1–72 (2003)

[2] Aumann, R.: Agreeing to disagree. The Annals of Statistics **4**(6), 1236–1239 (1976)

[3] Bala, V., Goyal, S.: Learning from neighbours. Review of Economic Studies **65**(3), 595–621 (1998). URL `http://ideas.repec.org/a/bla/restud/v65y1998i3p595-621.html`

[4] Banerjee, A.V.: A simple model of herd behavior. The Quarterly Journal of Economics **107**(3), 797–817 (1992). DOI 10.2307/2118364. URL `http://dx.doi.org/10.2307/2118364`

[5] Benjamini, I., Schramm, O.: Recurrence of distributional limits of finite planar graphs. Selected Works of Oded Schramm pp. 533–545 (2011)

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

[7] DasGupta, A.: Asymptotic theory of statistics and probability. Springer Verlag (2008)

[8] Ellison, G., Fudenberg, D.: Rules of thumb for social learning. Journal of Political Economy **110**(1), 93–126 (1995)

[9] Gale, D., Kariv, S.: Bayesian learning in social networks. Games and Economic Behavior **45**(2), 329–346 (2003). URL `http://ideas.repec.org/a/eee/gamebe/v45y2003i2p329-346.html`

[2] We conjecture that changing the tie-breaking rule does not produce asymptotic learning, even for randomized tie-breaking.

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

[11] McKelvey, R., Page, T.: Common knowledge, consensus, and aggregate information. Econometrica: Journal of the Econometric Society pp. 109–127 (1986)

[12] Ménager, L.: Consensus, communication and knowledge: an extension with bayesian agents. Mathematical Social Sciences **51**(3), 274–279 (2006)

[13] Mossel, E., Tamuz, O.: Making consensus tractable (2010). Preprint at http://arxiv.org/abs/1007.0959v2

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

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

[16] Smith, L., Sørensen, P.: Pathological outcomes of observational learning. Econometrica **68**(2), 371–398 (2000)