Aggregative Efficiency of Bayesian Learning in Networks

Krishna Dasaratha Kevin He

September 18, 2020

Social-Learning Dynamics in Different Networks

- **Social learning**: info about unknown state dispersed among society of agents, agents act based on private signals and observations of social neighbors
- How does social network affect efficiency of info aggregation?
- Esp. relevant today as communication technology reshapes networks: Facebook, Twitter, ...
- Existing work focuses on complete network
- Open question: impact of network on how well signals are aggregated — and hence how quickly rational agents learn

Golub and Sadler (2016): "A significant gap in our knowledge concerns short-run dynamics and rates of learning in these models. [...] The complexity of Bayesian updating in a network makes this difficult, but even limited results would offer a valuable contribution to the literature."

Environment and Key Results

Introduce **tractable model of rational sequential learning** that lets us compare learning dynamics across different networks

- rich signals, rich actions: Gaussian private signal, infer neighbors' beliefs perfectly from their actions
- strips away other sources of learning-rate inefficiency
- unique equilibrium of social-learning game has log-linear form

Highlight **network-based informational confounds**

- suppose 2 and 3 see 1, but 4 sees only 2 and 3
- 1's action confounds the info content of 2 and 3's behavior
- show how rational agents solve this signal-extraction problem

Generations network – observe subset of agents in previous gen

- express learning rate as simple function of network parameters
- extent of info loss: under a symmetry condition, learning aggregates no more than 2 signals per gen asymptotically
- applications to org structure: (1) value of mentorship in organizations; (2) benefits and costs of information silos

Related Literature

Sequential social learning

- Banerjee (1992), Bikhchandani, Hirshleifer, Welch (1992)
- Correct learning under mild conditions: Acemoglu, Dahleh, Lobel, Ozdaglar (2011), Lobel and Sadler (2015). This paper: speed.

Obstructions to the efficient learning rate in sequential social learning

- Coarse action space: Harel, Mossel, Strack, Tamuz (2020), Rosenberg and Vieille (2019), Hann-Caruthers, Martynov, Tamuz (2018)
  - HMST's "rational groupthink": trapped in wrong consensus for a long time as small belief changes are not reflected in actions
  - Rate of learning efficient if actions were rich
- Endogenous info: Burguet and Vives (2000), Mueller-Frank and Pai (2016), Ali (2018), Lomys (2019), Liang and Mu (2020).
- This paper: network-based obstructions to fast learning.

Lobel, Acemoglu, Dahleh, Ozdaglar (2009): compare two specific network structures with nbhd size 1. This paper: arbitrary fixed networks. Info confounding only appears in networks with nbhd size > 1.

Speed of learning under non-rational heuristics: Ellison and Fudenberg (1993), Golub and Jackson (2012), Molavi, Tahbaz-Salehi, Jadbabaie (2018). This paper: rational learning.

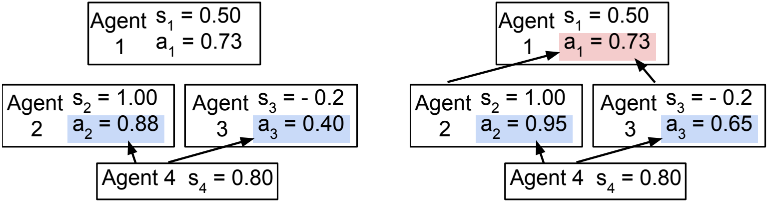
Outline

1. Setup and example of informational confound
2. Characterization results
   2.1 Log-linearity of the equilibrium
   2.2 Signal-counting interpretation of equilibrium accuracy
   2.3 Condition for long-run learning
3. The generations network
   3.1 Learning rate when agents fully observe the previous generation
   3.2 Applications: mentorship, information silos
   3.3 Main theorem: learning rate in any symm generations network
4. Efficiency of learning and welfare comparisons
5. Simulations on the robustness of results

Model and Notations

- Two equally likely states $\omega \in \{0, 1\}$
- Agents $i = 1, 2, 3, ...$ move in order, each acting once
  - i observes **private signal** $s_i \in \mathbb{R}$ and actions of **neighbors**, $N(i) \subseteq \{1, ...i-1\}$
  - picks **action** $a_i \in [0, 1]$ to maximize expectation of $-(a_i - \omega)^2$
- Signals are Gaussian and conditionally i.i.d. given state, $s_i \sim \mathcal{N}(1, \sigma^2)$ when $\omega = 1$ and $s_i \sim \mathcal{N}(-1, \sigma^2)$ when $\omega = 0$
- Neighborhoods define an **observation network** M, with $M_{i,j} = 1$ if $j \in N(i)$, $M_{i,j} = 0$ else. M is common knowledge.
- A **strategy** for i specifies i's play as a function of:
  1. observed actions from neighbors $N(i)$, and
  2. private signal s_i.
- Sequential nature of game $\Rightarrow$ there is a unique perfect-Bayesian **equilibrium** strategy profile

An Example of Informational Confound



- 4 perfectly infers 2 and 3's signals from their actions
- 4's accuracy $=$ 3 signals, fully incorporates info in s_2, s_3, and s_4

- a_1 influences both a_2 and a_3, but is unobserved by 4
- 4 cannot fully incorporate s_2 and s_3 without over-counting s_1
- optimal signal extraction: 4 puts "**2/3 as much weight**" on a_2 and a_3 as in other network
- 4's accuracy $=$ "**3.67 signals**"
  - (to be formalized soon)

Log-Linearity of the Equilibrium

WLOG apply log-transformations and work with log-variables

- **log-signal**, $\tilde{s}_i := \ln\left(\frac{\mathbb{P}[\omega=1|s_i]}{\mathbb{P}[\omega=0|s_i]}\right)$, **log-actions**, $\tilde{a}_i := \ln\left(\frac{a_i}{1-a_i}\right)$
- these changes are 1-to-1, so there is a (unique) map from i's neighbors' log-actions and i's log-signal to i's eqm log-action
- next proposition says this map is linear

Proposition 1

For each agent i with $N(i) = \{j(1), ..., j(d)\}$, there exist constants $(\beta_{i,j(k)})_{k=1}^{d}$ s.t.

$$\tilde{a}_i^* = \tilde{s}_i + \sum_{k=1}^{d} \beta_{i,j(k)} \tilde{a}_{j(k)}^*.$$

The vector of coefficients $\vec{\beta}_{i,\cdot}$ is given by

$$\vec{\beta}_{i,\cdot} = 2\mathbb{E}[(\tilde{a}_{j(1)}^*, ..., \tilde{a}_{j(d)}^*) \mid \omega = 1] \cdot \mathrm{Cov}[\tilde{a}_{j(1)}^*, ..., \tilde{a}_{j(d)}^* \mid \omega = 1]^{-1}.$$

Discussion of Proposition 1

Proposition 1

For each agent i with $N(i) = \{j(1), ..., j(d)\}$, there exist constants $(\beta_{i,j(k)})_{k=1}^{d}$ s.t. $\tilde{a}_i^* = \tilde{s}_i + \sum_{k=1}^{d} \beta_{i,j(k)} \tilde{a}_{j(k)}^*$. The vector of coefficients $\vec{\beta}_{i,\cdot}$ is given by

$$\vec{\beta}_{i,\cdot} = 2\mathbb{E}[(\tilde{a}_{j(1)}^*, ..., \tilde{a}_{j(d)}^*) \mid \omega = 1] \cdot \mathrm{Cov}[\tilde{a}_{j(1)}^*, ..., \tilde{a}_{j(d)}^* \mid \omega = 1]^{-1}.$$

- For general private signal distributions, Bayesian updating in networks intractable as Golub and Sadler (2016) point out
- Gaussian info structure leads to **log-linear eqm** and **closed-form expression of linear weights** that solve signal-extraction problem: downweight neighbors' log-actions if they have higher equilibrium correlation conditional on ω
- $\vec{\beta}_{i,\cdot}$ depends on network M, but not on signal precision $1/\sigma^2$

Signal-Counting Interpretation of Eqm Accuracy

If i's only info is $n \in \mathbb{N}_+$ indep signals, $\tilde{a}_i \sim \mathcal{N}\left(\pm n \cdot \frac{2}{\sigma^2}, n \cdot \frac{4}{\sigma^2}\right)$.

Definition

Social learning **aggregates** $r \in \mathbb{R}_+$ **signals by agent** i if the equilibrium log-action $\tilde{a}_i^* \sim \mathcal{N}\left(\pm r \cdot \frac{2}{\sigma^2}, r \cdot \frac{4}{\sigma^2}\right)$ in the two states.

- When agents use arbitrary strategy profile (even if log-linear), need not have $\tilde{a}_i \sim \mathcal{N}\left(\pm r \cdot \frac{2}{\sigma^2}, r \cdot \frac{4}{\sigma^2}\right)$ for **any** $r \in \mathbb{R}$
- But, **equilibrium** log-actions always admit this kind of signal-counting interpretation, suff. stat for rational accuracy

Proposition 2

There exist $(r_i)_{i\geq 1}$ so that social learning aggregates r_i signals by agent i. These $(r_i)_{i\geq 1}$ depend on the network M, but not on σ^2.

- Can help solve for eqm strategy profile in some cases
- $\lim_{i\to\infty}(r_i/i) \in [0,1]$ called aggregative efficiency of M

Condition for Long-Run Learning

Say society **learns completely in the long run** if equilibrium actions (a_i^*) converge to ω in probability.

Proposition 3

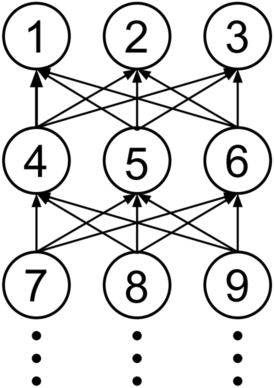
Society learns completely in the long run if and only if

$$\lim_{i\to\infty}\left[\max_{j\in N(i)} j\right]=\infty.$$

- If we consider the most recent neighbor of each agent, then this sequence of most-recent-neighbors tends to ∞
- Analog of Acemoglu, Dahleh, Lobel, and Ozdaglar (2011)'s **expanding observations** property for deterministic network
- Mild and clearly necessary: else for some $C<\infty$, infinitely many i cannot access the signal of any $j>C$ except their own
- Long-run learning not a useful way to compare networks
  - Instead, compare $(r_i)_{i\geq 1}$ and aggregative efficiency
  - May have $r_i\to\infty$ yet aggregative efficiency far below 1

The Maximal Generations Network

- $K \geq 1$ agents per generation
- Agents in gen t observe all agents in gen $t-1$



Proposition 4

In the maximal generations network:

- *Society learns completely in the long run with any K.*
- $\lim_{i \to \infty}(r_i/i) = \frac{(2K-1)}{K^2}$.
- *In the long run, social learning aggregates...*
  - *fewer signals per agent with larger K*
  - *fewer than 2 signals per generation with any K*
- *For any K and any i, i' in generations t and $t-1$ with $t \geq 3$, $r_i \leq r_{i'} + 3$.*

Bounds on Signals Aggregated Per Generation

- Agents in generation t have observation paths of length $t-1$
- Can show in any network, this implies $r_i \geq t$
- Social learning must aggregate at least 1 signal per gen
- This lower-bound not too far from the actual learning rate:

$$r_i \,/\underbrace{\lceil i/K \rceil}_{\text{gen of } i} = \underbrace{\frac{(2K-1)}{K}}_{<2} + o(1)$$

(No more than **2 signals** per gen in long-run, for any K)

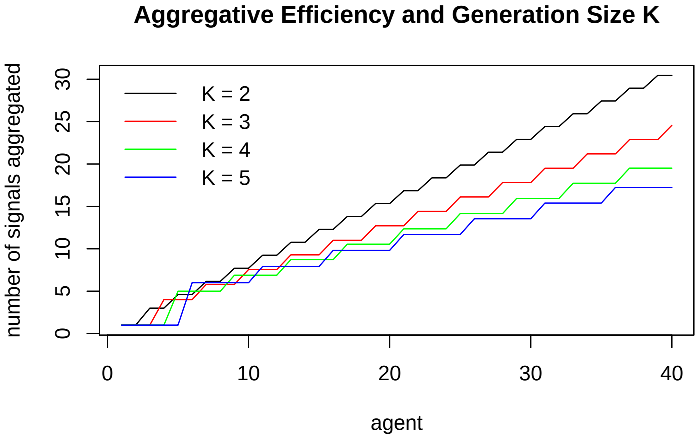
$$r_i - r_{i'} \leq 3, \qquad \text{for } i, i' \text{ in gen } t, t-1 \text{ where } t \geq 3$$

(No more than **3 signals** per gen **starting with gen 3**, for any K)

- For K large, individuals only manage to aggregate an unboundedly small fraction of their private signals in eqm

Slower Per-Agent Rate of Learning with Larger Gens

- If $K = 1$, every agent perfectly incorporates all past private signals $\Rightarrow$ fastest possible speed of social learning
- Prop 4 says aggregative efficiency strictly decreases in K
- Worse learning with larger K holds numerically starting from agent $i = 16$ when comparing among $K \in \{2, 3, 4, 5\}$



Application 1: Value of Mentorship in Organizations

- Many organizations with overlapping cohorts (e.g., colleges, professional firms, etc.) have mentorship programs, pairing each newcomer with someone from the previous cohort

Corollary 1

In the maximal generations network, if each agent additionally observes the ***private*** *signal of one agent from the previous generation (their "mentor"), then* $r_i \geq i - K$ *for every* i *and aggregative efficiency is 1.*

- Incumbents behave based on individual private info and shared org knowledge (e.g., key internal events in company's recent past)
- Newcomer is unaware of org knowledge, so becomes confused about incumbents' behavior
- De-confounding role of mentors: personal details of just one individual's experience can help interpret everyone's behavior

Application 1: Value of Mentorship in Organizations

Management literature discusses a related "socializing" benefit of mentors.

Chao (2007) in The Handbook of Mentoring at Work: Theory, Research, and Practice:

> *"Mentors can be powerful **socializing agents** as an individual adjusts to a new job or organization. As protégés learn about their roles within the organization, mentors can help them **correctly interpret their experiences** within the organization's expectations and culture."*

In our setting, it is this "interpretive" value of mentorship that helps build a more effective learning organization.

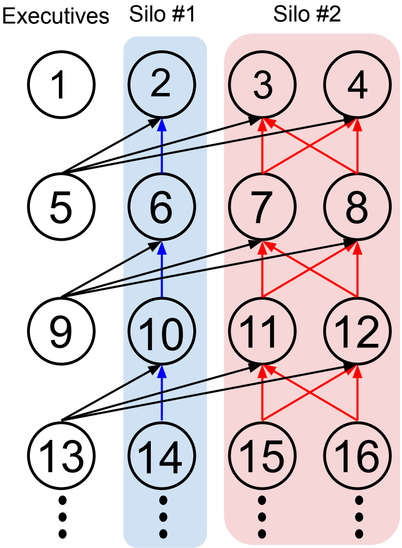
- If mentors generate new signals instead of sharing past signal realizations, social learning does not speed up very much

Application 2: Information Silos in Organizations

Information silos: In management, describes info fragmented among subgroups that do not communicate with each other

- Gillian Tett's 2015 book The Silo Effect documents prevalence of silos in government bureaucracies, technology firms, banks
- E.g. departments in the same municipal government, product divisions in a company, ...
- Causes: pay structure discourages collaboration across silos, technical barrier prevents flow of ideas across specialties, ...
- Silos persist for decades, as cohorts of new workers join the organization and bring in new info
- Tett (2015) joins a consensus in management consulting today in advocating breaking down silos
- We use a generations network to argue org can actually benefit from silos compared with fully transparent data sharing

Application 2: Information Silos in Organizations



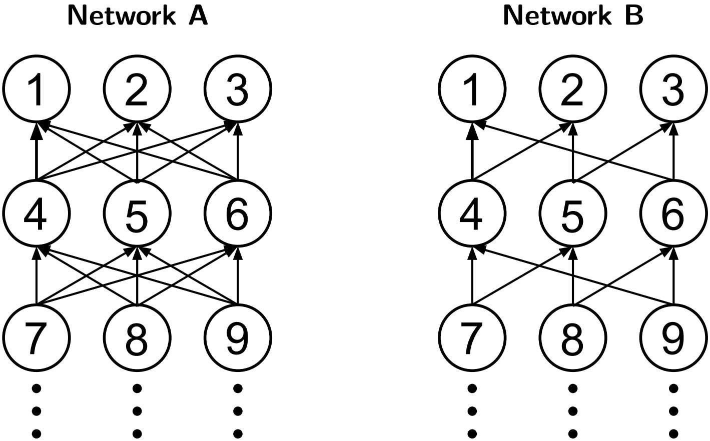
Corollary 2

Suppose org consists of N silos with $s_1, \ldots, s_N$ agents per generation, plus 1 executive per generation. In the long run, silo n aggregates $\frac{2s_n - 1}{s_n} < \frac{2K-1}{K}$ signals per gen, while the executives aggregate $\sum_{n=1}^{N} \frac{2s_n - 1}{s_n}$ signals per gen.

Application 2: Information Silos in Organizations

- Sacrifice rate of learning within silos to provide less confounded info to executives
- With full data sharing, workers in silos would learn better
  - Newcomers learn from predecessors across the org, instead of only predecessors from the same department
- But full data sharing slows down executives' learning
  - Actions from different silos conditionally independent
- Does breaking down silos help the org? It depends:
  - **NO** if org success closely identified with executives' actions
  - **YES** if everyone's action contributes to org's welfare
  - Negative case studies cited by Tett (2015) and management consultants involve workers in silos who take actions that severely harm the company

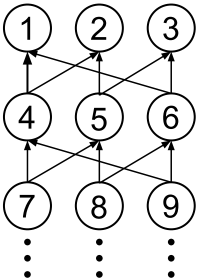
Which Network Leads to Faster Learning?



- **Network A** is the maximal generations network with $K = 3$
- **Network B** puts agents in each gen into 3 slots, $k \in \{1, 2, 3\}$. $k = 1$ sees 1 and 2, $k = 2$ sees 2 and 3, $k = 3$ sees 3 and 1. Less info confounding, but also fewer social observations.
- Need: aggregative efficiency on more general networks.

Generations Network with Partial Observations

- Generations network with K agents per gen
- $\Psi_k \subseteq \{1, \ldots, K\}$, **observation set**, define which gen $t-1$ slots are observed by a gen t agent in slot k
- Maximal generations network is the case of $\Psi_k = \{1, \ldots, K\}$



$$\Psi_1 = \{1, 2\},$$
$$\Psi_2 = \{2, 3\},$$
$$\Psi_3 = \{1, 3\}.$$

Generations Network with Partial Observations

Definition

The observation sets are **symmetric** if all agents observe $d \geq 1$ neighbors and all pairs of distinct agents in the same generation share c common neighbors. That is, for all $i_1 \neq i_2$ in same generation $t \geq 2$, $|N(i_1)| = d$ and $|N(i_1) \cap N(i_2)| = c$.

For example, "**Network B**" is symmetric with $d = 2$, $c = 1$.

More generally, for every $c \geq 1$ and $d = mc + 1$ where m is a positive integer, there exists a symmetric $(\Psi_k)_k$ with parameters d, c.

Speed of Learning with Partial Observations

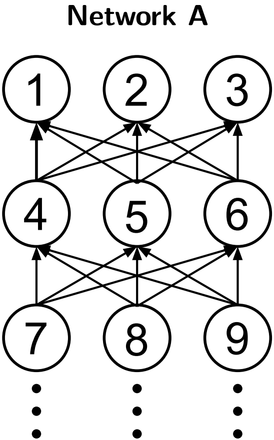
Theorem 1

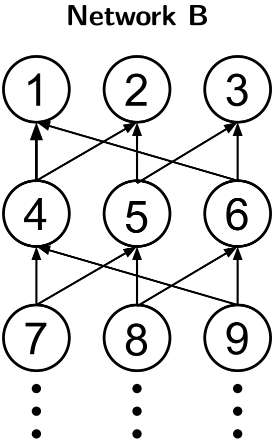
Suppose $(\Psi_k)_k$ *are symmetric. Then*

$$\lim_{i\to\infty}(r_i/i) = \left(1 + \frac{d^2 - d}{d^2 - d + c}\right)\frac{1}{K}.$$

- Exact expression of aggregative efficiency for a broader class of generations networks
- Term in parenthesis increases in d and decreases in c — more obs speeds up rate of learning per gen but more confounding slows it down, all else equal
- Maximal gen network has the worst rate of learning, among all symmetric gen networks with same d
  - Because actions very confounded in maximal gen network
- But Theorem 1 shows asymptotic bound of 2 signals per gen applies to **all** such networks, strengthening Proposition 4

Which Network Leads to Faster Learning?





- Applying Theorem 1, aggregative efficiency is the same in **Network A** ($d = 3, c = 3$) and **Network B** ($d = 2, c = 1$)!
- Extra social obs exactly cancel out reduced info content of each obs

Social Planner's Benchmark

Definition

$(\Psi_k)_k$ are **strongly connected** if for every $1 \leq k_1 \leq k_2 \leq K$, there exist t_1, t_2 so that $t_1 K + k_1$ is connected to $t_2 K + k_2$ in M.

Proposition 5

Suppose $(\Psi_k)_k$ are strongly connected and symmetric with $c \geq 1$. There is a log-linear strategy profile such that, for every $K_0 < K$, eventually agents' actions are are more accurate[1] than aggregating K_0 signals per generation.

- A social planner can aggregate close to all signals
- Slow learning of Thm 1 not intrinsic limitation of gen networks

▶ Conclusion

[1] i's action **more accurate than r signals** if it is more likely to lean towards the correct state than the action of someone who observes r indep signals.

Aggregative Efficiency and Welfare Comparisons

Aggregative efficiency leads to two kinds of welfare comparisons

- Let v_i be expected eqm welfare of i (depends on M and $1/\sigma^2$)
- We always have $-0.25 < v_i < 0$ for every i
- Social learning **strongly attains** $\underline{v}$ by agent I if I is the smallest integer s.t. $v_i \geq \underline{v}$ for all $i \geq I$
- Social learning **weakly attains** $\underline{v}$ by agent i if i is the smallest integer s.t. $v_i \geq \underline{v}$ (but later agents may do worse)

Proposition 6

Suppose aggregative efficiency is strictly positive in M and M', and strictly higher in M. For every $\underline{v} \in (-0.25, 0)$, there exists $\pi > 0$ so that if $0 < 1/\sigma^2 \leq \pi$, then social learning strongly attains $\underline{v}$ in M by agent I and weakly attains $\underline{v}$ by agent i in M', with $I < i$.

Aggregative Efficiency and Welfare Comparisons

Now fix $1/\sigma^2$. Social planner could evaluate utility profiles $v = (v_i)_{i \geq 1}$ using a social welfare function

$$\Lambda(v) = \sum_{i=1}^{\infty} \lambda_i v_i + \lambda_\infty \left(\lim_{i \to \infty} v_i \right)$$

- $\lambda_1, \lambda_2, ..., \lambda_\infty \geq 0$ summable sequence of welfare weights
- λ_∞ weight on "end of time"

"Infinitely patient" planner: Λ_∞ with $\lambda_i = 0$ for $i \in \mathbb{N}_+$, $\lambda_\infty = 1$

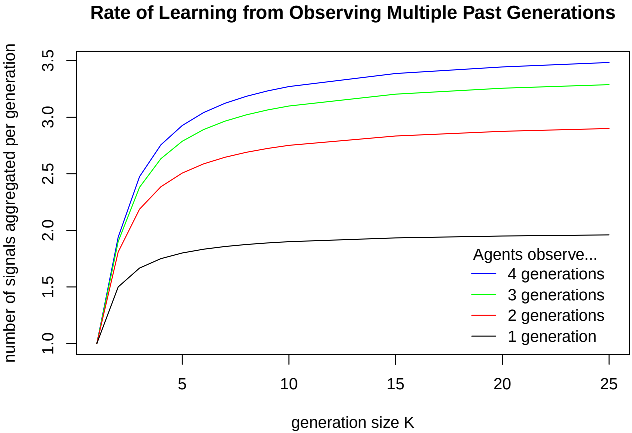
"Very patient" planner: Λ_T with $\lambda_i = 0$ for $i < T$, $\lambda_i > 0$ for $i \geq T$, where $T \in \mathbb{N}_+$ is large

Proposition 7

Suppose society learns completely in the long run in both M and M', but aggregative efficiency is strictly higher in M. There exists $\underline{T}$ so that if $T \geq \underline{T}$, then Λ_T is strictly higher on M than on M', though Λ_∞ is indifferent between M and M'.

Simulation: Observing Multiple Past Generations

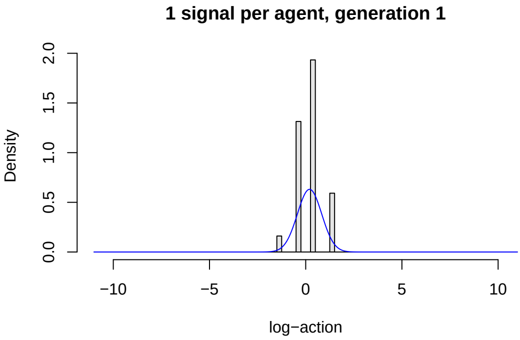
Each agent observes all predecessors from past $\tau \geq 1$ generations



- Limited improvement in aggregative efficiency: removes some confounds but creates new ones

Simulation: General Signal Structures

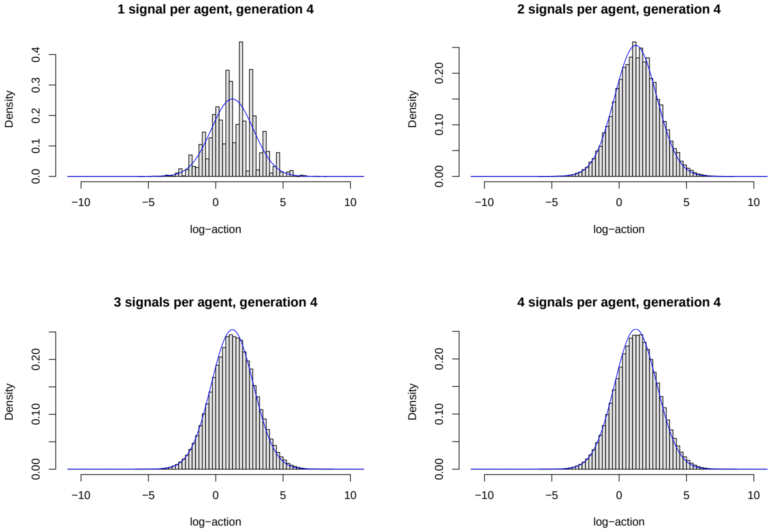
- Each signal is finitely supported



- Each agent has not 1, but n conditionally i.i.d. signals
- Think of agents who gather info over a period of time
- Increase n and scale down informativeness of each signal, fixing mean and SD of private log-belief (based on all n private signals) to match the Gaussian case

Simulation: General Signal Structures

- 2 agents per generation, maximal generations network
- Behavior very close to normal even with small $n > 1$



Simulation: General Signal Structures

- As n grows, mean and SD of gen 4 log-action approach their analogs when each agent has 1 Gaussian signal

each agent has...	mean	SD
1 signal	1.315	1.629
10 signals	1.201	1.543
100 signals	1.207	1.550
1000 signals	1.224	1.575
1 Gaussian signal	1.232	1.570

- Even if $n = 1$, log-actions in later gens resemble Gaussian
  - Sample of 1000 signals $\rightsquigarrow$ Shapiro-Wilk normality test rejects at $p < 0.05$ level with prob close to 1
  - Sample of 1000 gen 10 log-actions $\rightsquigarrow$ Shapiro-Wilk rejects at $p < 0.05$ level with prob 6%
  - Social learning aggregates i.i.d. signals from different agents
- Results and technique for Gaussian case may approximate behavior under general signal structures, at least for later gens

Summary

- A tractable model of rational sequential learning that focuses on how the social network affects aggregative efficiency
- Exact aggregative efficiency in all generations networks with symmetric observation sets
- Significant info loss due to confounding: in any such network, each generation eventually aggregates no more than 2 signals
- Analytic expression for aggregative efficiency permits comparative statics and applications about org structure: mentorship, information silos

Thank you!