

Bot Got Your Tongue?

Social Learning with Endogenous Action Visibility

John Cremin
Aix-Marseille School of Economics
Bristol Economic Theory Seminar

27th October 2026



Handout

Introduction

- In many situations where we learn by observing others' actions, they first choose whether to make said action public.
 - Having made a purchase do you brag about it?
 - After investing in a particular stock do you let others know?
 - Having formed an opinion in a political debate do you post it online?
- And yet the literature on sequential social learning largely neglects this...
- Whenever agents might face some penalty for being caught adopting the 'wrong' position, and are rewarded for being seen to be in the 'right', this will create endogenous observation networks



Handout

Preview I

This project:

- Studies sequential social learning where *social* agents can choose the visibility of their action, and *noise* agents act randomly.
- Finds that the outcome depends on whether the observation network is *sparse* or *dense*.

Sparse:

- $\exists M \in \mathbb{N}$ max # neighbours
- Unravelling, exaggerated noise and reduced learning
- Mild reputation concerns can help learning, strong ones harm it

Dense:

- No such M
- No unravelling
- Can facilitate learning in removing *cascade beliefs*.

Outline

Model

Dense Networks

Sparse: Unravelling

Sparse: Countervailing Effects

Self-Locating Information

Literature

- SSL models have an infinite sequence of agents $n \in \mathbb{N}$ trying to match action x_n to the state θ , with social and private information.

Background Literature for SSL in general:

- Bikhchandani et al. 1992, Banerjee 1992, Smith & Sorensen 2000
 - Unbounded vs Bounded Private Signals (e.g. Normal vs finite)
- Acemoglu et al. 2011, Lobel & Sadler 2015 & 2016
 - Complete Learning with unbounded beliefs iff *expanding observations*
- Smith and Sorensen 2008, Monzon & Rapp 2014

This project specifically builds on a small ‘visibility’ literature.

- Guarino et al. 2011, Herrera & Horner 2013

Model

Model

The basic elements of the model are:

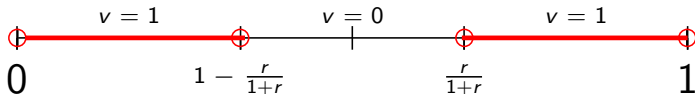
- $\theta \in \{0, 1\}$: $\mathbb{P}(\theta = 0) = \frac{1}{2}$
- Agents have (private) type $\tau_n \in \{S, N\}$
 - Choose action $x_n \in \{0, 1\}$ & visibility $v_n \in \{0, 1\}$
 - $\mathbb{P}(\tau = N) = \eta \in (0, 1) \Rightarrow x_n$ random and $v_n = 1$.
 - $\tau = S$: Reputation Concern: $r_n \sim^{i.i.d} \Delta([1, \infty))$
 - Candour $c_n = \frac{1}{r_n}$

$$u_n(x_n, v_n, \theta) = \begin{cases} 1 + v_n & \text{if } x_n = \theta, \\ -(1 + r_n \cdot v_n) & \text{if } x_n \neq \theta, \end{cases}$$

Impact of Reputation Concerns

$$u_n(x_n, v_n, \theta) = \begin{cases} 1 + v_n & \text{if } x_n = \theta, \\ -(1 + r_n \cdot v_n) & \text{if } x_n \neq \theta, \end{cases}$$

The **Decision Rules** that follow imply they will choose x_n as per usual, and $v_n = 1$ if very confident:



Model II

They have two distinct sources of information:

- A Private Signal: $s_n \in \mathcal{S}_n \sim (\mathbb{F}_0, \mathbb{F}_1) \Rightarrow p_n \sim (\mathbb{G}_0, \mathbb{G}_1)$
(mutually a.c., distinct)
- A Social Signal $\{x_b : b \in B(n)\}$ with $B(n) \sim \mathbb{Q}(n)$
(independent across n)

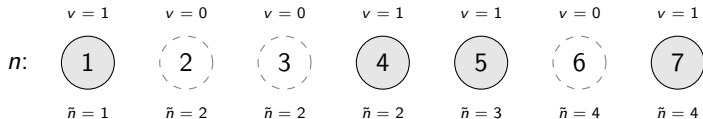
The Social Signal

There are numerous assumptions about the social signal ('Sample'/'Neighborhood') in the literature:

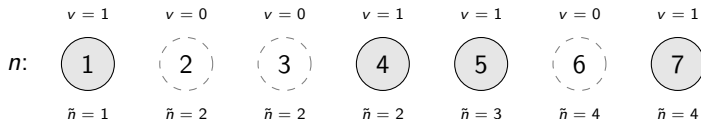
1. ADLO2011: Non-anonymous & $\mathbb{Q}$ common knowledge
2. SS2008: Anonymous & Unordered Sample & Σ common knowledge
3. MR2014: SS2008 & Uncertainty over own position ' $P(i)$ '
 - This article follows the first approach here, though the unravelling mechanic holds for any of these cases.

Indices and Adjusted Indices

- Agents arrive in sequence $n \in \mathbb{N}$, but only *visible* actions can ever be observed.
- Adjusted index: $\tilde{n} = \sum_{i=1}^{n-1} v_i + 1$ (one plus the number of visible actions taken before n arrives): if you act visibly, you will be the $\tilde{n}$ th agent to do so.
- Agents observe their own $\tilde{n}$ (and those of their neighbours), but not n .



Indices and Adjusted Indices II



- The network is defined on adjusted indices: $B(\tilde{n}) \sim \mathbb{Q}_{\tilde{n}}$, independent across $\tilde{n}$.
- Many agents share an adjusted index, but only one *visibly-acting* agent has each: $\tilde{x}_{\tilde{n}}, \tilde{\tau}_{\tilde{n}}$.
- Agents observe $\tilde{n}$, not $n \Rightarrow$ self-locating uncertainty

►► Self-Locating Information

Dense Networks

Dense Networks

- In dense networks, there is no M , and therefore no unravelling.
- Taking the complete network as the canonical dense network, we can show how learning can benefit from reputation concerns and noise:

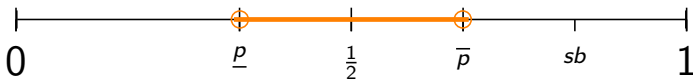
Learning in the Complete Network

The following two statements characterise learning in the complete network:

1. With unbounded beliefs, there is complete learning. The social belief converges to certainty on the true state almost surely.
2. With bounded beliefs supported on $[\underline{p}, \bar{p}] \subset (0, 1)$, and a candour distribution $\Delta(c)$ such that for some $\epsilon > 0$, $(0, c^* + \epsilon) \cup (1 - \epsilon, 1) \subseteq \text{supp}(\Delta(c))$, where $c^* := \frac{p(1-\bar{p})}{\bar{p}(1-p)}$, the social belief converges to certainty on the true state almost surely.

Reputation Concerns and Noise as a Cure for Cascade Beliefs

To see this, consider a complete network. The social belief at a given visible history is sb , and we are interested in the probability with which the next visible action is a 1, $\phi(sb, \theta)$:



Noise but no Reputation Concerns:

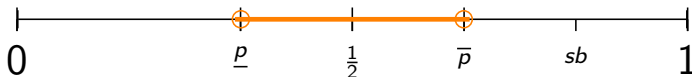
- $\phi(sb, 1) = 1 - \eta/2$
- $\phi(sb, 0) = 1 - \eta/2$
- Learning stops

Reputation Concerns but no Noise:

- $\phi(sb, 1) = 1$
- $\phi(sb, 0) = 1$
- Learning stops

Reputation Concerns and Noise as a Cure for Cascade Beliefs II

To see this, consider a complete network. The social belief at a given visible history is sb , and we are interested in the probability with which the next visible action is a 1, $\phi(sb, \theta)$:



But with *both* noise and reputation concerns satisfying Part 2 of our theorem:

- $\phi(sb, 1) = 1 - \mathbb{P}(\tilde{\tau}_n = N | \theta = 1)/2 < 1 - \eta/2$
- $\phi(sb, 0) = 1 - \mathbb{P}(\tilde{\tau}_n = N | \theta = 0)/2 < \phi(sb, 1)$
- Learning continues...

Learning in the Complete Network: Proof

Suppose $\theta = 1$, let $\ell_{\tilde{n}} = \frac{1-sb_{\tilde{n}}}{sb_{\tilde{n}}}$, and let $\phi(sb, \theta)$ be the probability the next visible action is a 1.

1. Martingale convergence: ℓ a.s. does not converge to ∞ .
2. Smith & Sorensen 2000 (Theorem B.1): ℓ a.s. converges to a *stationary* belief, $\phi(sb, 0) = \phi(sb, 1)$.
3. It suffices to show $\phi(sb, 1) > \phi(sb, 0)$ at every interior sb .

$$\phi(sb, \theta) = \frac{\eta/2 + (1-\eta)R_\theta}{\eta + (1-\eta)(R_\theta + L_\theta)}$$

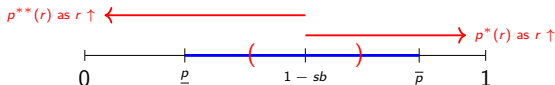
$$R_\theta = \int 1 - \mathbb{G}_\theta(p^*(r)) d\Delta(r), \quad L_\theta = \int \mathbb{G}_\theta(p^{**}(r)) d\Delta(r)$$

- ϕ is strictly increasing in R and decreasing in L .
- $\mathbb{G}_1$ FOSD $\mathbb{G}_0 \Rightarrow R_1 \geq R_0, L_1 \leq L_0$: we need one of these to be strict.

Learning in the Complete Network: The Bounds $\underline{p}, \bar{p}$

If $\tau_n = S$, $v_n = 0$ iff $p_n \in (p^{**}(r), p^*(r))$, where $p^*(1) = p^{**}(1) = 1 - sb$ and:

$$p^{**}(r) = \frac{1 - sb}{1 + (r - 1)sb} \searrow 0, \quad p^*(r) = \frac{r(1 - sb)}{sb + r(1 - sb)} \nearrow 1$$

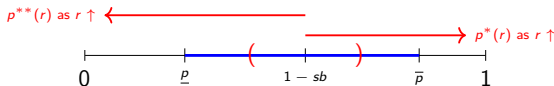


- FOSD is only strict on $(\underline{p}, \bar{p})$: need $p^*(r)$ or $p^{**}(r)$ in there for a positive measure of r .
- $sb \in (1 - \bar{p}, 1 - \underline{p})$: easy, as $p^*(1) = p^{**}(1) = 1 - sb \in (\underline{p}, \bar{p})$ already, so any mass on r near 1 will do.

Learning in the Complete Network: The Bounds $\underline{p}, \bar{p}$

If $\tau_n = S$, $v_n = 0$ iff $p_n \in (p^{**}(r), p^*(r))$, where $p^*(1) = p^{**}(1) = 1 - sb$ and:

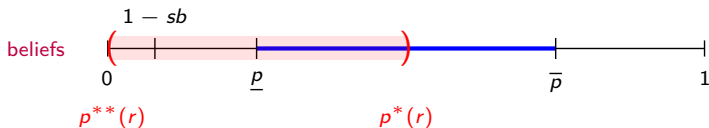
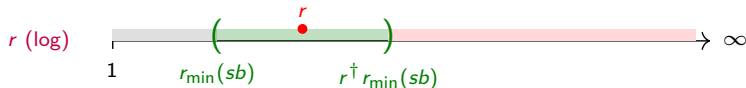
$$p^{**}(r) = \frac{1 - sb}{1 + (r - 1)sb} \searrow 0, \quad p^*(r) = \frac{r(1 - sb)}{sb + r(1 - sb)} \nearrow 1$$



- $sb \geq 1 - \underline{p}$: $p^*(r) \in (\underline{p}, \bar{p})$ iff $r \in (r_{\min}(sb), r^\dagger r_{\min}(sb))$, where $r^\dagger := 1/c^*$ and $r_{\min}(sb) = \underline{p} sb / ((1 - sb)(1 - \underline{p}))$ ($= 1$ at $sb = 1 - \underline{p}$, $\rightarrow \infty$ as $sb \rightarrow 1$) $\Rightarrow (r^\dagger - \epsilon, \infty) \subseteq \text{supp}(\Delta(r))$.
- $sb \leq 1 - \bar{p}$: symmetric, with $p^{**}(r)$.

Learning in the Complete Network: The Bounds $\underline{p}, \bar{p}$

$$sb \geq 1 - \underline{p}: p^*(r) \in (\underline{p}, \bar{p}) \text{ iff } r \in (r_{\min}(sb), r^\dagger r_{\min}(sb))$$



- $r_{\min}(sb) = \underline{p} sb / ((1 - sb)(1 - \underline{p}))$ ($= 1$ at $sb = 1 - \underline{p}$, $\rightarrow \infty$ as $sb \rightarrow 1$), and $r^\dagger r_{\min}(sb) \geq r^\dagger \Rightarrow (r^\dagger - \epsilon, \infty) \subseteq \text{supp}(\Delta(r))$.

► Interactive

Sparse: Unravelling

Theorem

Suppose the network is sparse, with commonly-known bound $M < \infty$ on neighbourhood sizes. Then $\underline{\eta}_\infty = \frac{\eta}{\eta + (1 - \lambda_\infty)(1 - \eta)}$ is a lower bound on the probability that the $\tilde{n}$ th visibly-acting agent is a noise type for any $\tilde{n} \in \mathbb{N}$, and the asymptotic fraction of agents that are noise types is almost surely greater than $\underline{\eta}_\infty$. λ_∞ is the smallest solution in $[0, 1]$ of:

$$\lambda_\infty = \min_{\theta \in \{0, 1\}} \int_{1/\omega_\infty}^{\infty} \mathbb{G}_\theta \left(\left(\frac{r\omega_\infty}{1 + r\omega_\infty} \right)^- \right) - \mathbb{G}_\theta \left(\frac{1}{1 + r\omega_\infty} \right) d\Delta(r)$$

where $\omega_\infty := (\eta / (2 - \eta - 2\lambda_\infty(1 - \eta)))^M$.

Unravelling

- Suppose we have a sparse network topology with common bound M . Any neighbourhood could be entirely noise agents $\Rightarrow [1 - \overline{SB}, \overline{SB}]$
- For a given reputation concern r , we can have $v_n = 1$ only if:

$$\mathbb{P}(\theta = 1 | s_n) > \frac{r[1 - \overline{SB}]}{r[1 - \overline{SB}] + \overline{SB}}$$

$$\text{or } \mathbb{P}(\theta = 1 | s_n) < 1 - \frac{r[1 - \overline{SB}]}{r[1 - \overline{SB}] + \overline{SB}}$$

Thus, with $r > r^* := \frac{(1 - \frac{\eta}{2})^M}{(\frac{\eta}{2})^M}$, we have $v_n = 0$ for all:

$$\mathbb{P}(\theta = 1 | s_n) \in \underbrace{\left[1 - \frac{r(\frac{\eta}{2})^M}{r(\frac{\eta}{2})^M + (1 - \frac{\eta}{2})^M}\right]}_{L_0}, \underbrace{\left[\frac{r(\frac{\eta}{2})^M}{r(\frac{\eta}{2})^M + (1 - \frac{\eta}{2})^M}\right]}_{U_0}$$

Reputation Concerns & Visibility II

Therefore any social agent chooses $v_n = 0$ with at least probability:

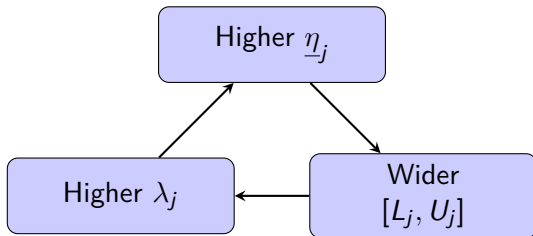
$$\lambda_0 := \min\{\mathbb{G}_0(U_0^-) - \mathbb{G}_0(L_0), \mathbb{G}_1(U_0^-) - \mathbb{G}_1(L_0)\}$$

Neighbors are drawn only from the $v_n = 1$ pool, thus the relevant probabilities are not $\eta, 1 - \eta$, since at most $(1 - \lambda_0)$ of the social agents choose visible actions.

- $\underline{\eta}_1 := \frac{\eta}{(\eta + (1 - \lambda_0)(1 - \eta))}$ is a lower bound for the noise probability.

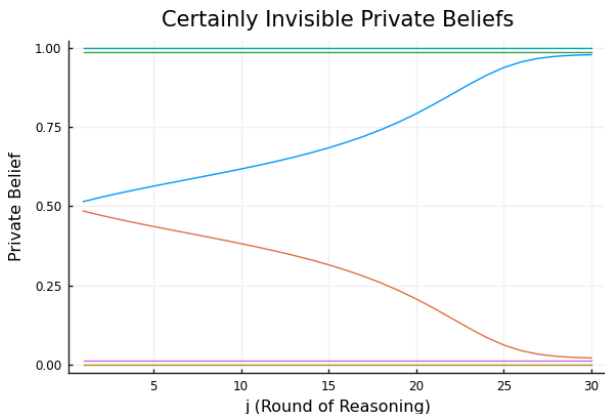
Reputation Concerns & Visibility III

We can then iterate this reasoning ad infinitum:



- $\underline{\eta}_\infty = \lim_{j \rightarrow \infty} \underline{\eta}_j$ (inc. & bounded above)

Reputation Concerns & Visibility IV



- $f_{\theta}(\cdot) = g_{\theta}(\cdot)$, $M = 2$, $\eta = 0.2$, $r = 86$, $\underline{\eta}_{\infty} = 0.864...$

Unravelling with an r-Distribution

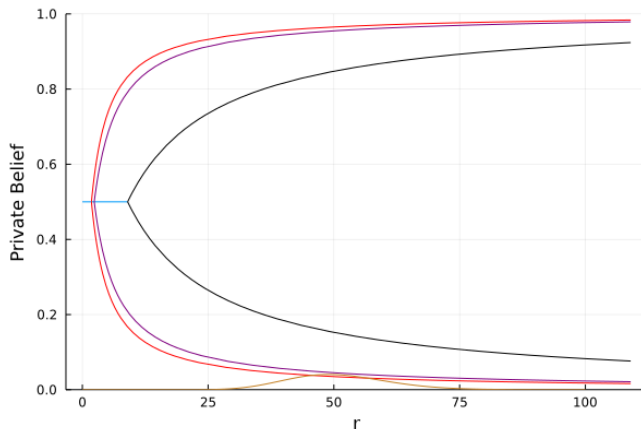


Figure: The parameters for this example are $f_1(x) = 2x$, $f_0(x) = 2 - 2x$, $M = 1$, $\eta = 0.2$; the distribution of $r_n - 1$ is a Chi-squared with 50 degrees of freedom.

Characterising λ_∞

- Everything in round j depends on $\underline{\eta}_j$ only through $\omega_j := (\underline{\eta}_j / (2 - \underline{\eta}_j))^M$:

$$L_j(r) = \frac{1}{1 + r\omega_j}, \quad U_j(r) = \frac{r\omega_j}{1 + r\omega_j}, \quad r_j^* = \frac{1}{\omega_j}$$

- Since $\underline{\eta}_j = \eta / (\eta + (1 - \lambda_{j-1})(1 - \eta))$, we can write ω_j directly in terms of λ_{j-1} :

$$\omega_j = \left(\frac{\eta}{2 - \eta - 2\lambda_{j-1}(1 - \eta)} \right)^M =: \Omega(\lambda_{j-1})$$

- The iteration is therefore a recursion in a single number, λ_j .

Characterising λ_∞ II

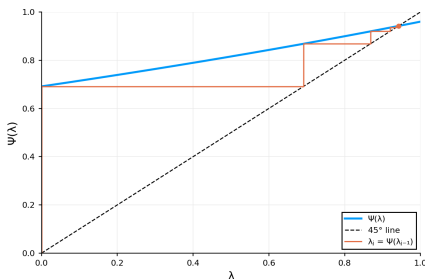
- This gives $\lambda_j = \Psi(\lambda_{j-1})$, where (with $\omega = \Omega(\lambda)$):

$$\Psi(\lambda) := \min_{\theta \in \{0,1\}} \int_{1/\omega}^{\infty} \mathbb{G}_\theta \left(\left(\frac{r\omega}{1+r\omega} \right)^- \right) - \mathbb{G}_\theta \left(\frac{1}{1+r\omega} \right) d\Delta(r)$$

- Ψ is non-decreasing: Ω is increasing in λ , and a higher ω increases $U_j(r)$, decreases $L_j(r)$ and decreases r_j^* .
- Ψ maps $[0, 1]$ into itself and $\lambda_0 = \Psi(0) \geq 0$, so $\{\lambda_j\}_{j \in \mathbb{N}}$ is non-decreasing and bounded above by 1, and converges.
- Monotone convergence applies to each integral, so the limit solves $\lambda_\infty = \Psi(\lambda_\infty)$.

Characterising λ_∞ III

- λ_∞ is the *smallest* solution: if $\lambda^* = \Psi(\lambda^*)$, then $\lambda_j \leq \lambda^*$ implies $\lambda_{j+1} = \Psi(\lambda_j) \leq \Psi(\lambda^*) = \lambda^*$, so an iteration starting from 0 can never pass a fixed point.

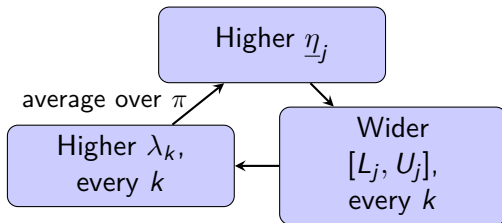


- The example with an r -distribution above:
 $\{2(1-s), 2s\}$, $M = 1$,
 $\eta = 0.2$, $r_n - 1 \sim \chi_{50}^2$.
- $\lambda_0 \approx 0.69$, $\lambda_1 \approx 0.87$,
 $\lambda_2 \approx 0.92$, $\dots$,
 $\lambda_\infty \approx 0.94$, so $\underline{\eta}_\infty \approx 0.81$.

Unravelling without M

- The theorem asks very little of what agents know about the network: only the common bound M .
- With more structure, we can drop M altogether:
 - Each $\tilde{n}$ draws a degree $d_{\tilde{n}} \sim \pi$ (π commonly known), and $|B(\tilde{n})| = \min\{d_{\tilde{n}}, \tilde{n} - 1\}$.
 - Agents observe their own degree, not their neighbours'.
 - π can have unbounded support, even an infinite mean (e.g. scale-free networks, Barabási & Albert 1999).

Same iteration as before, but degree by degree:



Proposition (Unravelling with a Degree Distribution)

The smallest solution $\underline{\eta}_{\infty}^{\pi}$ of $\eta' = \Phi(\eta')$ in $[\eta, 1]$ is a lower bound on the probability that the $\tilde{n}$ th visibly-acting agent is a noise type for any $\tilde{n} \in \mathbb{N}$, and the asymptotic fraction of visibly-acting agents that are noise types is almost surely at least $\underline{\eta}_{\infty}^{\pi}$, where:

$$\Phi(\eta') := \sum_{k=0}^{\infty} \pi(k) \frac{\eta}{\eta + (1 - \lambda_k(\eta'))(1 - \eta)}$$

and $\lambda_k(\eta')$ is the right-hand side of the theorem's equation with ω_{∞} replaced by $(\eta'/(2 - \eta'))^k$.

- $\underline{\eta}_{\infty}^{\pi} > \eta$ iff $\lambda_{\underline{k}}(\eta) > 0$, where $\underline{k}$ is the smallest degree in the support of π .
- $\underline{\eta}_{\infty}^{\pi}$ is decreasing with first order stochastic shifts in π .
- If π puts all its mass on M , $\underline{\eta}_{\infty}^{\pi} = \underline{\eta}_{\infty}$.

Unravelling without M II

- Fewer neighbours $\Rightarrow$ less extreme social beliefs $[1 - \overline{SB}_k, \overline{SB}_k]$
 $\Rightarrow v_n = 0$ more often: λ_k is decreasing in k .
- So unravelling is driven by the agents with the fewest neighbours:
 - Whether it occurs at all is decided by the lowest degree $\underline{k}$.
 - Well-connected agents, however many of them there are, cannot prevent it.
- Only agents' own degrees enter: the short paths of small-world networks (Watts & Strogatz 1998) don't help either.

Unbounded degree:

- No M , but mass at low k
- Unravelling

Diverging degree:

- e.g. the complete network
- No unravelling

Unravelling with a Degree Distribution

The example with an r -distribution above: $\{2(1-s), 2s\}$, $\eta = 0.2$,
 $r_n - 1 \sim \chi_{50}^2$.

π	η_{∞}^{π}
$d = 1$	0.81
$d = 1$ or 2 , each w.p. $\frac{1}{2}$	0.74
$d = 1$ w.p. 0.3 , $d = 100$ otherwise	0.31
$\pi(k) \propto k^{-2}$ ($\mathbb{E}[d] = \infty$)	0.58

- Infinite mean degree, and still almost three times η noise.

Sparse: Countervailing Effects

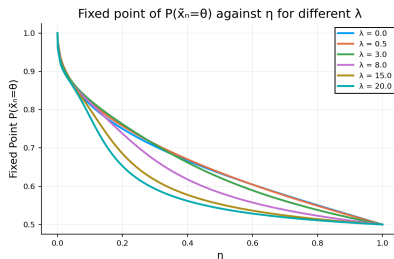
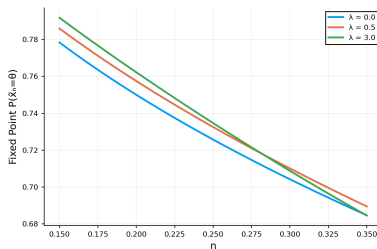
Sparse Networks

Reputation concerns and noise agents have two countervailing effects in general:

- It increases the probability any neighbour is a noise agent (bad)
- It means those neighbours who are not noise agents are acting on the basis of stronger beliefs (good)

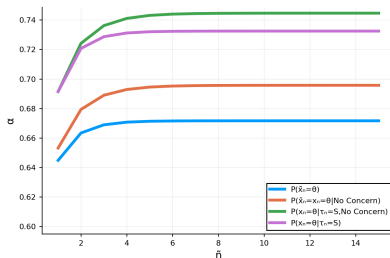
Which of these effects dominates asymptotically depends on the network topology, the exact shape of the reputation-concern distribution and the private signal distributions.

Poisson Beta Example

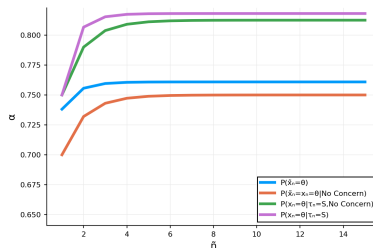


- Agents form an immediate predecessor network, and have private signal and belief density functions $\{2(1-s), 2s\}$.
- $r-1$ is Poisson with mean λ : moving from $\lambda = 0$ to $\lambda = 0.5$ increases accuracy for a large range of η , and moving on to $\lambda = 3$ increases it further for a smaller range (right).

Normal vs Beta Signals Example



(a) $N(\theta, 1)$



(b) $\{2(1-s), 2s\}$

- Agents form an immediate predecessor network, all have $r = 5$, and $\eta = 0.2$.
- Reputation concerns lower asymptotic accuracy in (a), and raise it in (b): nothing has changed but the signal distributions.

Countervailing Effects in the Line Network

Proposition (Countervailing Effects)

Suppose that agents form an immediate predecessor network, all have the same $r \geq 1$, and that the private belief distributions are antisymmetric, with a density that is bounded above, continuous and strictly positive on $(0, 1)$ (so that private beliefs are unbounded). Then low-level reputation concerns improve learning, and strong ones harm it:

$$\begin{aligned}\alpha^*(r) &> \alpha^*(1) \text{ for all } r > 1 \text{ close enough to } 1, \\ \alpha^*(r) &< \alpha^*(1) \text{ for all } r \text{ large enough.}\end{aligned}$$

For these r , and for $r = 1$, both limits exist, in every equilibrium, and the same inequalities hold for $\tilde{\alpha}^*$.

- $\tilde{\alpha}^*(r)$, $\alpha^*(r)$: limits of visible accuracy $\mathbb{P}(\tilde{x}_{\tilde{n}} = \theta)$, and of the accuracy of the first agent to arrive after the $\tilde{n}$ th visible action if he is social (blue and purple lines in the example above).

Countervailing Effects: Proof Sketch

1. If the last visible action is right w.p. α in each state, observers have $sb \in \{\alpha, 1 - \alpha\}$, and:

$$\tilde{\alpha}_{\tilde{n}} = \Gamma(\tilde{\alpha}_{\tilde{n}-1}; r), \quad \Gamma(\alpha; r) := \alpha\phi(\alpha, 1) + (1 - \alpha)\phi(1 - \alpha, 1)$$

with ϕ as in the complete network, and

$\alpha^* = (\Gamma(\tilde{\alpha}^*; 1) - \frac{\eta}{2}) / (1 - \eta)$, increasing in $\tilde{\alpha}^*$.

2. $\Gamma(\cdot; r)$ maps $[\frac{1}{2}, 1 - \frac{\eta}{2}]$ into itself. If $\sup|\partial_\alpha \Gamma| < 1$ it is a contraction: unique fixed point $\tilde{\alpha}^*(r)$, and $\tilde{\alpha}_{\tilde{n}} \rightarrow \tilde{\alpha}^*(r)$.
3. Implicit function theorem:

$$\frac{d\tilde{\alpha}^*}{dr} = \frac{\partial_r \Gamma}{1 - \partial_\alpha \Gamma}$$

so the fixed point moves in the direction of $\partial_r \Gamma$.

Countervailing Effects: Proof Sketch II

4. r close to 1:

$\partial_\alpha \Gamma(\alpha; 1) = (1 - \eta)(\mathbb{G}_1(\alpha) - \mathbb{G}_1(1 - \alpha)) \in [0, 1 - \eta]$, so contraction, and by antisymmetry:

$$\partial_r \Gamma(\alpha; 1) = 2(1 - \eta)^2 \alpha^2 (1 - \alpha) \mathbb{G}'_1(1 - \alpha) [\mathbb{G}_0(\alpha) - \mathbb{G}_1(\alpha)] > 0$$

Raising r first silences agents with $p_n \approx 1 - sb$: nearly indifferent, so right about half the time.

5. r large: $R_1, L_1 \rightarrow 0$ uniformly, so $\phi(\cdot, 1) \rightarrow \frac{1}{2}$, $\Gamma(\cdot; r) \rightarrow \frac{1}{2}$ and $\partial_\alpha \Gamma \rightarrow 0$:

$$\tilde{\alpha}^*(r) \rightarrow \frac{1}{2}, \quad \alpha^*(r) \rightarrow 1 - \mathbb{G}_1(0.5) < \alpha^*(1)$$

Almost every visible agent is noise: the accuracy of an agent with no social information.

Strong Enough Reputation Concerns Kill Learning

Proposition

Suppose we have a learning game with $M < \infty$, and consider a sequence of reputation-concern distributions $\Delta^k(r)$ for $k \in \mathbb{N}$, in which each successive distribution is shifted to the right by some fixed value $\delta > 0$. $\underline{\eta}_{\infty}^k \rightarrow 1$, and from this the limiting visible accuracy $\tilde{\alpha}^* \rightarrow 0.5$ and accuracy $\alpha^* \rightarrow \frac{1}{2}\mathbb{G}_0(0.5) + \frac{1}{2}(1 - \mathbb{G}_1(0.5))$.

- The positive effect of reputation concerns can dominate if they are mild, but the bad must eventually dominate.

Strong Enough Reputation Concerns Kill Learning II

- Also, the lower is M , the sooner this first effect must dominate in terms of the lower bound $\underline{\eta}_\infty$.
- This is true for the two remaining points in the $\underline{\eta}_\infty$ comparative statics lemma below:

Lemma (Comparative Statics)

$\underline{\eta}_\infty$ is:

- Increasing in η .
- Increasing with First order stochastic shifts in $\Delta(r)$.
- Decreasing in M .
- Decreasing with any 'Birnbbaum-spread' about 0.5 in $\mathbb{G}_0$ and $\mathbb{G}_1$.

Why Birnbaum and not Mean-Preserving Spreads?

Definition (Birnbaum 1948)

F' is *less peaked about* $\frac{1}{2}$ than F if

$F'(\{p : |p - \frac{1}{2}| \geq z\}) \geq F(\{p : |p - \frac{1}{2}| \geq z\})$ for all $z \geq 0$: F' assigns weakly less mass than F to every interval centred on $\frac{1}{2}$.

- λ_∞ is the mass the private belief distributions assign to the invisibility regions $(L_\infty(r), U_\infty(r))$, and $L_\infty(r) = 1 - U_\infty(r)$: these are intervals centred on $\frac{1}{2}$.
- Less mass on every such interval $\Rightarrow$ lower $\lambda_\infty \Rightarrow$ lower $\underline{\eta}_\infty$.
- A mean-preserving spread can *raise* the central mass:

$$\mathbb{G}' = \frac{1}{2}U[0.4, 0.6] + \frac{1}{4}U[0, 0.1] + \frac{1}{4}U[0.9, 1]$$

is a mean-preserving spread of $U[0, 1]$, but the mass on $[0.4, 0.6]$ rises from 0.2 to 0.5.

Tail Thickness

Proposition

Consider a learning game with some $M < \infty$ and some set of private belief distributions $\{\mathbb{G}_0, \mathbb{G}_1\}$ such that $U_\infty^1(\underline{r}) > L_\infty^1(\underline{r})$. If we take a sequence of new unbounded belief distributions $\{\mathbb{G}_0^k, \mathbb{G}_1^k\}_{k \in \mathbb{N}}$ such that each is less peaked about $\frac{1}{2}$ than its successor, and the mass assigned to $[L_\infty^1(\underline{r}), U_\infty^1(\underline{r})]$ converges to 1, then $\underline{\eta}_\infty^k \rightarrow 1$.

- $\underline{r}$ is the lowest supported r .
- As with reputation concerns, this convergence will happen faster when M is lower.
- Without the constraint on $\underline{r}$, $\underline{\eta}_\infty^k$ still climbs, but to some limit strictly below 1.

Self-Locating Information

Self-Locating Information

- Recall: agents observe $\tilde{n}$ rather than n .
- One play of the game visits the same information set many times: *self-locating uncertainty*.
- Inspired by the Absent-minded Driver Paradox, the SBP begins on Sunday.

	$\mathbb{P}(H) = 0.5$	
	Heads	Tails
Monday	●	●
Tuesday		●

Self-Locating Information II

- Such games can feature a dynamic inconsistency, a *Newcomb tension* (Cremin 2026, *Sleeping Newcomb*):
 - Were agents able to commit to a strategy at a planning stage, they would act *as if* they held beliefs other than the Bayesian beliefs they form at the information set.
- *Sleeping Newcomb*, Theorem 2.2: a committed thirder acts as an uncommitted agent with *one-boxer beliefs*.
- *Sleeping Newcomb*, Proposition 2.1(i): payoffs here depend only on (x_n, v_n, θ) , so one-boxer beliefs on θ are simply the prior.
- By assumption: $\mathbb{P}(\theta = 1 | \tilde{n}) = \frac{1}{2}$
- This is tantamount to assuming agents have commitment power.

Ill-defined thirder beliefs

- Without commitment: $\mathbb{P}(\theta = 1|\tilde{n}) = \frac{\mathbb{P}(\tilde{n}|\theta=1)}{\mathbb{P}(\tilde{n}|\theta=1)+\mathbb{P}(\tilde{n}|\theta=0)} = \frac{0}{0+0}$

Could alternatively give agents an improper prior $\epsilon > 0$ on each $n \in \mathbb{N}$. What happens if we do?

- My benchmark theorem with no noise agents no longer holds...
- But all other results hold anyway, though we may need higher r values for unravelling.

Improper prior beliefs

- Improper prior: $\epsilon > 0$ on each $n \in \mathbb{N}$.
- Observing $\tilde{n} = 1$: I may be agent 1, or agent k with $k - 1$ invisible predecessors. A Bayesian considers every such hypothesis, in both states.

$$\begin{aligned}\mathbb{P}(\tilde{n} = 1 | \theta = 0) &= \epsilon + \epsilon \mathbb{P}(v_1 = 0 | \theta = 0) + \dots \\ &\quad + \epsilon \mathbb{P}(v_1 = 0 | \theta = 0) \times \dots \times \mathbb{P}(v_{k-1} = 0 | \theta = 0) + \dots\end{aligned}$$

- Each term is less likely the more invisible predecessors it requires, but none is ruled out.
- Suppose $\mathbb{P}(v_k = 0 | \theta) \in [\underline{p}, \bar{p}]$ for all k and θ . Then:

$$\frac{\epsilon}{1 - \underline{p}} \leq \mathbb{P}(\tilde{n} = 1 | \theta) \leq \frac{\epsilon}{1 - \bar{p}}$$

- Noise $\Rightarrow$ infinitely many visible agents a.s., so the same bounds hold for any $\tilde{n}$.

Bounding indexical beliefs

$$\mathbb{P}(\theta = 1|\tilde{n}) = \frac{1}{1 + \frac{\mathbb{P}(\tilde{n}|\theta=0)}{\mathbb{P}(\tilde{n}|\theta=1)}} \leq \frac{1}{1 + \frac{1-\bar{p}}{1-\underline{p}}}$$

- Noise agents always act visibly: $\bar{p} = 1 - \eta$. Taking $\underline{p} = 0$ (and symmetry):

$$\mathbb{P}(\theta = 1|\tilde{n}) \in \left[\frac{\eta}{1 + \eta}, \frac{1}{1 + \eta} \right]$$

- Sparse networks: social agents act invisibly w.p. at least λ_∞ , so $\underline{p} = (1 - \eta)\lambda_\infty$, and $1 - \underline{p} = \eta/\underline{\eta}_\infty$:

$$\mathbb{P}(\theta = 1|\tilde{n}) \leq \frac{1 - (1 - \eta)\lambda_\infty}{1 - (1 - \eta)\lambda_\infty + \eta} = \frac{1}{1 + \underline{\eta}_\infty}$$

- E.g. $\eta = 0.2$: $[0.17, 0.83]$, but $[0.46, 0.54]$ with $\underline{\eta}_\infty = 0.864$ (earlier example).

Improper priors: which results survive?

- Indexical beliefs are bounded away from 0 and 1, which is all the sparse results need:
 - Social beliefs are still bounded away from 0 and 1 $\Rightarrow$ a central interval of private beliefs acts invisibly for high enough r .
 - Unravelling results hold with unchanged expressions, though higher r may be needed.
 - Limit accuracy now depends on $\tilde{n}$, through indexical beliefs.
- Dense networks: the social belief is still a martingale converging to 1 on the true state; indexical beliefs, like private beliefs, are eventually swamped.
- Fails: the no-noise benchmark ($\eta = 0$), where not every information set is reached with probability 1 and $\bar{p} = 1 - \eta$ gives no bound.

Summary

- I introduce choice over action-visibility into sequential social learning.
- The combined effects of introducing reputation concerns and noise depend on whether the network is sparse or dense.
 - In dense networks, noise and reputation concerns can remove cascade beliefs and facilitate learning, unravelling is also no longer a threat.
 - In sparse networks they produce a form of unravelling, leading to an overrepresentation of noise and damaging learning (with strong enough reputation concerns), though mild reputation concerns can help the latter.

►► One-Sided Noise

Decision Rules

Lemma

After any history, h_n , agent n will choose $x_n = 1$ if

$$\mathbb{P}(\theta = 1|s_n) + \mathbb{P}(\theta = 1|B(\tilde{n})) > 1$$

and $x_n = 0$ if this sum is strictly less than 1. If they exactly equal 1, the agent is indifferent and assumed to randomise uniformly.

Lemma

After any history h_n , the agent chooses $v_n = 0$ if:

$$1 - (r - 1)\mathbb{P}(\theta = 1|B(\tilde{n}))\mathbb{P}(\theta = 1|s_n) < S_n < 1 + \frac{r - 1}{r}\mathbb{P}(\theta = 1|B(\tilde{n}))\mathbb{P}(\theta = 1|s_n)$$

for $S_n = \mathbb{P}(\theta = 1|B(\tilde{n})) + \mathbb{P}(\theta = 1|s_n)$.

One-Sided Noise

- So far noise agents randomise uniformly between the two actions. What if they all choose $x = 1$?
- E.g. a targeted disinformation campaign by a hostile state during an election.
- Sparse networks: bots still prevent learning, but have an exaggeratedly negative impact on the very state they are promoting ($\theta = 1$).
- Dense networks: noise and reputation concerns help learning in a smaller set of circumstances than in the main model.

One-Sided Noise: Unravelling

- A neighbourhood in which all choose $x = 1$ could be entirely noise agents, so social beliefs are no stronger than $\frac{(1-\eta)^M}{(1-\eta)^M + \eta^M}$ (main model: $\eta/2$ in place of η).
- No such bound for social beliefs in favour of $\theta = 0$: anyone choosing $x = 0$ is a social type.
- Thus agents will not choose $\{x_n = 1, v_n = 1\}$ for:

$$\mathbb{P}(\theta = 1 | s_n) \in [0, \underbrace{\frac{r_n \eta^M}{r_n \eta^M + (1-\eta)^M}}_{U_0^1}]$$

- $\lambda_0^1 := \min\{\mathbb{G}_0(U_0^1), \mathbb{G}_1(U_0^1)\}$ and $\underline{\eta}_1^1 := \frac{\eta}{(\eta + (1-\lambda_0^1)(1-\eta))}$
- Iterate exactly as before $\Rightarrow \underline{\eta}_\infty^1, \lambda_\infty^1$, but $\underline{\eta}_\infty^1$ is only relevant for neighbours who chose $x = 1$.

One-Sided Noise: Sparse Results

Proposition

We can find equivalents of the unravelling results:

1. $\underline{\eta}_\infty^1$ is a lower bound on the probability that the $\tilde{n}$ th visibly-acting agent is a noise type for any $\tilde{n} \in \mathbb{N}$, and the asymptotic fraction of noise types is almost surely greater than $\underline{\eta}_\infty^1$.
2. Conditional on θ , accuracy can be bounded by

$$\begin{aligned}\mathbb{P}(x_n = \theta | \theta = 1) &\leq 1 - \mathbb{G}_1 \left(1 - \frac{(1 - \eta)^M}{(1 - \eta)^M + \eta^M} \right) \\ \mathbb{P}(x_n = \theta | \theta = 0) &\leq (1 - \eta^M) + \eta^M \mathbb{G}_0(0.5)\end{aligned}$$

and we can tighten these by replacing η with $\underline{\eta}_\infty^1$.

3. *Strong Enough Reputation Concerns Kill Learning and Tail Thickness* hold without adjustment, with $\underline{\eta}_\infty^1$ in place of $\underline{\eta}_\infty$.

One-Sided Noise: Bots Hurt Their Own Cause?

- $\theta = 1$: social beliefs in favour of $\theta = 1$ are capped by η rather than $\eta/2$, a tighter bound on accuracy than in the main model.
- $\theta = 0$: no cap, but with probability η^M an agent's neighbourhood is entirely noise, and they can do no better than $\mathbb{G}_0(0.5)$.
- Without noise, unbounded beliefs give complete learning. With one-sided noise this is lost, and strong enough reputation concerns again completely smother learning.
- Why do bots do more damage to the state they are promoting?
 - Agents know η , and know bots always choose $x = 1$, so they heavily discount neighbourhoods of $x = 1$.
 - Real disinformation works because voters fail to account for bots: this would require misspecification, as in Bohren and Hauser (2021).

One-Sided Noise: Dense Networks

Proposition

Suppose noise agents always choose $x_n = 1$. If private beliefs have support $(0, \bar{b})$, i.e. they are bounded on the $\theta = 1$ side only, and the candour distribution $\Delta(c)$ satisfies $[0, \epsilon) \cup (1 - \epsilon, 1) \subseteq \text{supp}(\Delta(c))$ for some $\epsilon > 0$, the social belief converges to certainty on the true state almost surely.

- No noise agents choose $x = 0$, so there may be social beliefs $sb > 0.5$ at which no agent chooses $\{x_n = 0, v_n = 1\}$: the main model's bounded beliefs result is lost.
- Unbounded beliefs in favour of $\theta = 0 \Rightarrow$ no cascade beliefs above 0.5.
- Below $1 - \bar{b}$: $\{x_n = 1, v_n = 1\}$ w.p. at least η , and $\mathbb{G}_0(sb) > \mathbb{G}_1(sb) \Rightarrow \phi(sb, 0) < \phi(sb, 1)$, so learning continues.