Breaking Bad News

Moritz Meyer-ter-Vehn ® Simon Board

UCLA

September 8, 2025

Motivation

Decentralized Learning on Networks

- New drugs, Coleman (1957)
- New electronic gadgets, Goolsbee Klenow (2002)
- Microfinance, Banerjee et al (2013)

Social Planner can reveal early outcomes to late movers

- Medical association
- Industry watchdog
- Government regulator

Motivation

Decentralized Learning on Networks

- New drugs, Coleman (1957)
- New electronic gadgets, Goolsbee Klenow (2002)
- Microfinance, Banerjee et al (2013)

Bad News Events

- Drugs have side effects
- New gadgets explode
- Financial products are scams

Transparency Policies for Bad News Events

- FDA requires drug companies report adverse health events
- CPSC requires manufacturers report unsafe products
- SEC requires customers be notified of fraud

Bad News Learning

Modeling Approach

- Network design as realistic instance of information design
- Challenge: How to optimize over networks?

Problems with Transparency

- Agents want neighbors to act as guinea pigs (Sgroi, 02)
- Planner cares about "no news" event (Che Horner, 18)
- What are Pareto-optimal networks?

Bad News Learning

Modeling Approach

- Network design as realistic instance of information design
- Challenge: How to optimize over networks?

Problems with Transparency

- Agents want neighbors to act as guinea pigs (Sgroi, 02)
- Planner cares about "no news" event (Che Horner, 18)
- What are Pareto-optimal networks?

Nevertheless ...

- Binary outcomes: Full transparency is optimal
- General outcomes: Transparency can Pareto-improve

Literature

Designed Social Learning

- BHW: Bikhchandani, Hirshleifer, Welch (1992)
- Sgroi (2002), Che, Horner (2018), Knöpfle, Salmi (2024)

Bayesian Learning on Networks

- Acemoglu et al (2011) - network model
- Board, MtV (2021) - observable outcomes

General Dynamic Information Design

- Ely (2017)
- Bergemann, Morris (2019)

Here: Bad News Learning

Model

Model

Players and Actions

- Single product with quality $q \in \{L, H\}$
- Agents $t \in \{1, ..., T\}$ can consume product, $c_t \in \{0, 1\}$

Binary "bad news" outcomes $\xi_t \in \{\dagger, \checkmark\}$

- $\dagger$ = "harm" if $c_t = 1$ and $q = L$
- $\checkmark$ = "safe" if $c_t = 0$ or $q = H$

Information and Payoffs

- Observe type θ_t and outcomes of neighbors $G_t \subseteq \{1, ..., t-1\}$
- Arbitrary beliefs $\Pr_{\theta_t}(q, \theta_{-t})$
- Private payoffs $c_t \cdot u(\theta_t, \xi_t)$, with $u(\theta_t, \dagger) < u(\theta_t, \checkmark)$

Dominance-solvable (break ties in favor of $c_t = 1$)

Discussion of Model

Special Cases of Model

- Common prior over $\Omega = \{L, H\} \times \Theta$; update to get $\Pr_{\theta_t}$

(1) BHW types: $p_t = \Pr_{\theta_t}(H) \sim U[0,1]$ conditionally independent

(2) Private value types: $\kappa_t \sim U[0,1]$ independent

$$u(\theta_t, \dagger) = -\kappa_t < 1 - \kappa_t = u(\theta_t, \checkmark)$$

Generalizations of Model

- Imperfect harm: $\Pr_{\theta_t}(\xi_t = \dagger | c_t = 1, q = L) < 1$ (independent)
- Random network $\mathcal{G}$ (independent)
- Timing uncertainty: t moves at time $\tau(\theta_t) \in \{1, ..., T\}$

Example

Preliminaries: Updating and Best Responses

Blackwell-measure of θ_t's social information

$$y^G(\theta_t) = \Pr{}_{\theta_t}(\xi_s = \dagger \text{ for some } s \in G_t | L)$$

- G *Pareto-dominates* G' iff $y^G(\theta_t) \geq y^{G'}(\theta_t)$ for all t, θ_t

When observing no harm

- Bayesian updating

$$\Pr{}_{\theta_t}(H | \xi_s = \checkmark \text{ for all } s \in G_t) = \frac{p_t}{1 - (1 - p_t) y^G(\theta_t)} \qquad (*)$$

- Consume if $(*)$ exceeds indifference threshold $p = p^*(\theta_t)$

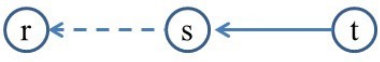
$$p \cdot u(\theta_t, \checkmark) + (1 - p) \cdot u(\theta_t, \dagger) = 0$$

Guinea Pig Problem

Effect of $r \leftarrow s$ on Tara

- Sven gets pessimistic when Ruby is harmed
- Sven gets optimistic when Ruby is safe

Guinea Pig Problem



Parametric effect of $r \leftarrow s$ on Tara

- Assume $\theta_s = \kappa_s \sim U[0,1]$, so Sven consumes iff $P_s \geq \kappa_s$
- Tara's learning from guinea pig Sven

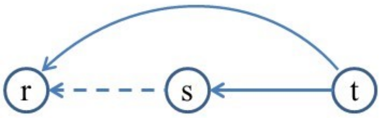
$$y^G(\theta_t) = \Pr{}_{\theta_t}[P_s \geq \kappa_s | L] = E_{\theta_t}[P_s | L]$$

falls in Sven's information

Welfare effect of new link

- Sven benefits, Tara loses
- Utilitarian welfare falls if there are many Taras

Guinea Pig Solution



Effect of $r \leftarrow s$ on Tara's learning from Sven

1. negative if $\xi_r = \dagger$
2. positive if $\xi_r = \checkmark$

- "On average", (1) outweighs (2)
- Given $r \leftarrow t$, Tara only cares about (2)

Resolving the conflict of interest

- Tara wants to maximize learning from guinea pig Sven
- Transparency directs experimentation to pivotal event, $\xi_r = \checkmark$

Optimal Networks

Main Result

Transitive Networks

- Complete network $\bar{G}$, where $\bar{G}_t = \{1, ..., t-1\}$
- More generally, G is *transitive* if $s \in G_t \Rightarrow G_s \subset G_t$

Theorem 1.
Full transparency $\bar{G}$ Pareto-dominates any $G \neq \bar{G}$.
More strongly, G Pareto-dominates all subnetworks $G' \subset G$, iff G is transitive.

Main Result

Transitive Networks

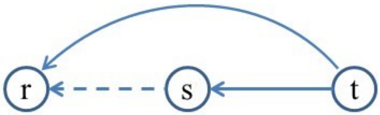
- Complete network $\bar{G}$, where $\bar{G}_t = \{1, ..., t-1\}$
- More generally, G is *transitive* if $s \in G_t \Rightarrow G_s \subset G_t$

Theorem 1.
Full transparency $\bar{G}$ Pareto-dominates any $G \neq \bar{G}$.
More strongly, G Pareto-dominates all subnetworks $G' \subset G$, iff G is transitive.

Corollary: Full transparency remains optimal with info delays

- For instance $\mathcal{G}^1 := \{G : (t-1) \notin G_t\}$
- Maximal $\bar{G}^1$ with $\bar{G}^1_t = \{1, ..., t-2\}$ dominates any $G \in \mathcal{G}^1$

Isn't Transparency Always the Best Policy?



Bad News: $\Pr_{\theta_t}(\xi_s = \dagger | q) > 0$ only if $q = L, c_s = 1$

- Harm $\xi_r = \dagger$ proves $q = L$, obviating further learning
- Absence of harm is good news, inducing Sven to consume
- Transparency for Sven improves Tara's learning

Good News: $\Pr_{\theta_t}(\xi_s = \checkmark | q) > 0$ only if $q = H, c_s = 1$

- Success $\xi_r = \checkmark$ proves $q = H$, obviating further learning
- Absence of success is bad news, inducing Sven to not consume
- Transparency for Sven reduces Tara's learning

Proof that Transitive G Pareto-dominates any $G' \subseteq G$

Probability $y(\theta_t) = \Pr_{\theta_t}(B_t|L)$ of *Harm States*

$$B_t := \{(q,\theta) : \xi_s = \dagger \text{ for some } s \in G_t\} \subseteq \Omega$$

Proof that Transitive G Pareto-dominates any $G' \subseteq G$

Probability $y'(\theta_t) = \Pr_{\theta_t}(B'_t|L)$ of *Harm States*

$$B'_t := \{(q, \theta) : \xi'_s = \dagger \text{ for some } s \in G'_t\} \subseteq \Omega$$

Proof that Transitive G Pareto-dominates any $G' \subseteq G$

Probability $y'(\theta_t) = \Pr_{\theta_t}(B'_t|L)$ of *Harm States*

$$B'_t := \{(q,\theta) : \xi'_s = \dagger \text{ for some } s \in G'_t\} \subseteq \Omega$$

Claim: Inductively $B_t \supseteq B'_t$

- Anchor: $B_1 = B'_1 = \emptyset$
- Step: Assume $B_s \supseteq B'_s$ for all $s < t$
- Fix $\omega \in B'_t$, so $\xi'_s = \dagger$ for some $s \in G'_t \subseteq G_t$
- Since G is transitive, $B_s \subseteq B_t$, so focus on $\omega \notin B_s$
- By induction, Sven is more likely to see harm in G than in G'

$$y(\theta_s) = \Pr_{\theta_s}(B_s|L) \geq \Pr_{\theta_s}(B'_s|L) = y'(\theta_s)$$

so Sven consumes and is harmed in G; thus $\omega \in B_t$

Towards a Characterization of the Pareto-Ranking

G Pareto-dominates $G' \subset G$

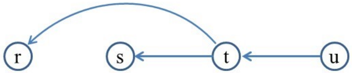
- ... if G is transitive
- ... only if $s \in G_t$ and $r \in G_s \backslash G'_s$ implies $r \in G_t$

Towards a Characterization of the Pareto-Ranking

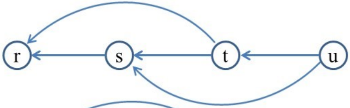
G Pareto-dominates $G' \subset G$

- ... if G is transitive
- ... only if $s \in G_t$ and $r \in G_s \backslash G'_s$ implies $r \in G_t$

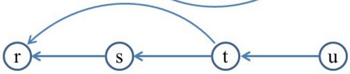
But G' =



is dominated by



is not dominated by



Application: Selling the Product

Corollary. The complete network maximizes high-quality demand

$$E^G\left[\sum_{t=1}^{T} c_t | H\right]$$

Proof. High-quality is harmless, so $\bar{G}$ maximizes c_t for all ω.

Application: Selling the Product

Corollary. The complete network maximizes high-quality demand

$$E^G\left[\sum_{t=1}^{T} c_t | H\right]$$

Proof. High-quality is harmless, so $\bar{G}$ maximizes c_t for all ω.

Low-quality demand tends to go the other way

- Assume common prior $\Pr$, and uniform cost types $\kappa_t \sim U[0,1]$
- Expected demand independent of G

$$\pi^G = E^G\left[\sum c_t\right] = E^G\left[\sum P_t\right] = T \cdot \Pr[H]$$

Application: Selling the Product

Corollary. The complete network maximizes high-quality demand

$$E^G \left[\sum_{t=1}^{T} c_t | H \right]$$

Proof. High-quality is harmless, so $\bar{G}$ maximizes c_t for all ω.

Assume informed seller chooses G

- High-quality seller chooses $\bar{G}$
- Low-quality seller must follow suit

Beyond Perfect Bad News

Key assumption in Theorem 1

- Harm proves $q = L$, obviating further learning

Imperfect bad news: $\Pr_{\theta_t}(\xi_s = \dagger | H) = \epsilon > 0$

- For ϵ small, $\bar{G}$ remains optimal by continuity
- For ϵ larger, Tara may prefer $r, s \leftarrow t$ over $\bar{G}$

Polar opposite: Perfect good news learning

- Optimal policy for Tara is t-star, $G_t = \{1, ..., t-1\}, G_s = \emptyset$

Information Design

Beyond Pure Perfect Bad News Learning

Non-Binary Outcomes

- Outcomes $\Xi \supseteq \{\dagger, \checkmark\}$
- Binary harm indicators $\Phi := \{\dagger, \Xi \backslash \dagger\}$
- Example: Observable actions $\Xi = \{\dagger, \checkmark\} \times \{0, 1\}$

Independent beliefs $\pi_{\theta_t, c_t, q} \in \Delta(\Xi)$

- Non-consumption, $c_t = 0$, keeps Tara safe: $\xi_t = \checkmark$
- High-quality, $q = H$, prevents harm: $\xi_t \neq \dagger$

Payoffs

$$E_{\theta_t,1,L}[u(\theta_t, \xi_t)] < E_{\theta_t,1,H}[u(\theta_t, \xi_t)]$$

Beyond Pure Perfect Bad News Learning

Non-Binary Outcomes

- Outcomes $\Xi \supseteq \{\dagger, \checkmark\}$
- Binary harm indicators $\Phi := \{\dagger, \Xi \backslash \dagger\}$
- Example: Observable actions $\Xi = \{\dagger, \checkmark\} \times \{0, 1\}$

Independent beliefs $\pi_{\theta_t, c_t, q} \in \Delta(\Xi)$

- Non-consumption, $c_t = 0$, keeps Tara safe: $\xi_t = \checkmark$
- High-quality, $q = H$, prevents harm: $\xi_t \neq \dagger$

Payoffs

$$E_{\theta_t,1,L}[u(\theta_t, \xi_t)] < E_{\theta_t,1,H}[u(\theta_t, \xi_t)]$$

$\bar{G}$ need not Pareto-improve G... What does?

Attempt 1: Broadcasting Harm G^+

Tara's information

- Outcomes ξ_s^+ of neighbors $s \in G_t$, and harm ϕ_s^+ for all $s < t$
- "Multi-layer network": $G_t^\xi = G_t$, $G_t^\phi = \bar{G}$

But: G^+ does not Pareto-improve G if G has indirect link

$$r \leftarrow s \leftarrow t \quad \text{with} \quad r > 1$$

- $\phi_1^+ \neq \dagger$ can induce $c_r^+ = 1$ instead of $c_r = 0$
- ξ_r^+ can be worse news than $\xi_r = \checkmark$ for some θ_s
- Can induce $c_s^+ = 0 < 1 = c_s$, lowering learning for some θ_t

Attempt 1: Broadcasting Harm G^+

Tara's information

- Outcomes ξ_s^+ of neighbors $s \in G_t$, and harm ϕ_s^+ for all $s < t$
- "Multi-layer network": $G_t^\xi = G_t$, $G_t^\phi = \bar{G}$

But: G^+ does not Pareto-improve G if G has indirect link

$$r \leftarrow s \leftarrow t \quad \text{with} \quad r > 1$$

- $\phi_1^+ \neq \dagger$ can induce $c_r^+ = 1$ instead of $c_r = 0$
- ξ_r^+ can be worse news than $\xi_r = \checkmark$ for some θ_s
- Can induce $c_s^+ = 0 < 1 = c_s$, lowering learning for some θ_t

Caveat: Pareto-improvement for all (t, θ_t) is a very high bar

Attempt 2: Adjusted Harm Broadcast G^*

Broadcast harm, $\phi_s^* = \dagger$, asap; before that, inductively at t

- Principal reveals outcomes ξ_t from original network
- Hides outcomes ξ_t^* from actual choice c_t^*
- E.g. With observable actions, reveal hypothetical choice c_t.

Is this feasible?

- Survey hypothetical choice c_t (which is based solely on ξ_{G_t})
- G^* raises optimism so $c_t^* \geq c_t$ and principal infers ξ_t from ξ_t^*

Theorem 2.
G^* Pareto-improves G.

Attempt 2: Adjusted Harm Broadcast G^*

Broadcast harm, $\phi_s^* = \dagger$, asap; before that, inductively at t

- Principal reveals outcomes ξ_t from original network
- Hides outcomes ξ_t^* from actual choice c_t^*
- E.g. With observable actions, reveal hypothetical choice c_t.

Is this feasible?

- Survey hypothetical choice c_t (which is based solely on ξ_{G_t})
- G^* raises optimism so $c_t^* \geq c_t$ and principal infers ξ_t from ξ_t^*

Theorem 2.
G^* Pareto-improves G.

- G^* is "robust" (Wilson 87), and "detail-free" (Maskin 00)
- But it is not a network ...

Information "Design"

Mechanism M

- Private deterministic message $m_t = m(\xi^{t-1}, \omega)$ to t

Definition 1.
A class of mechanisms $\mathcal{M}$ satisfies "breaking bad news" BBN if for any $M \in \mathcal{M}$ there exists a Pareto-better $\hat{M} \in \mathcal{M}$ that publicly broadcasts harm, i.e. $\hat{m}_t = \dagger$ iff $\hat{\xi}_s = \dagger$ for some $s < t$.

Summary of Results

- $\{M : m_t = \xi_{G_t}$ for some $G\}$ satisfies BBN (Thm1, $\Xi = \{\dagger, \checkmark\}$)
- $\{M : m_t = \xi_{G_t}$ or $(\xi^+_{G_t}, \phi^+_{\bar{G}_t})$ for some $G\}$ fails BBN
- $\{M : m_t = \xi_{G_t}$ or $(\xi_{G_t}, \phi^*_{\bar{G}_t})$ for some $G\}$ satisfies BBN (Thm2)

Classes of Mechanisms

Mechanisms with unknown types

- Send message $m_t(\xi^{t-1})$
- Special case: Multi-layer networks (G^{ξ}, G^{ϕ}): $m_t = (\xi_{G_t^{\xi}}, \phi_{G_t^{\phi}})$

Classes of Mechanisms

Mechanisms with unknown types

- Send message $m_t(\xi^{t-1})$
- Special case: Multi-layer networks (G^ξ, G^ϕ): $m_t = (\xi_{G_t^\xi}, \phi_{G_t^\phi})$

Mechanisms with known types

- Send message $m_t(\xi^{t-1}, \theta^t)$

Classes of Mechanisms

Mechanisms with unknown types

- Send message $m_t(\xi^{t-1})$
- Special case: Multi-layer networks (G^ξ, G^ϕ): $m_t = (\xi_{G_t^\xi}, \phi_{G_t^\phi})$

Mechanisms with elicitation (menus of messages)

- Elicit $\tilde{\theta}_t$, send message $m_t(\xi^{t-1}, \tilde{\theta}^t)$
- Truth-telling constraint: $\tilde{\theta}_t = \theta_t$

Mechanisms with known types

- Send message $m_t(\xi^{t-1}, \theta^t)$

Classes of Mechanisms

Mechanisms with unknown types

- Send message $m_t(\xi^{t-1})$
- Special case: Multi-layer networks (G^ξ, G^ϕ): $m_t = (\xi_{G^\xi_t}, \phi_{G^\phi_t})$

Mechanisms with surveys

- Survey $\tilde{\theta}_{t-1}$, send message $m_t(\xi^{t-1}, \tilde{\theta}^{t-1})$
- Truth-telling $\tilde{\theta}_t = \theta_t$ automatic since c_t fixed
- Special case: Adjusted harm broadcast $m_t = (\xi_{G_t}, \phi^*_{\bar{G}_t})$

Mechanisms with elicitation (menus of messages)

- Elicit $\tilde{\theta}_t$, send message $m_t(\xi^{t-1}, \tilde{\theta}^t)$
- Truth-telling constraint: $\tilde{\theta}_t = \theta_t$

Mechanisms with known types

- Send message $m_t(\xi^{t-1}, \theta^t)$

Summary of BBN improvement principle

Multi-layer Networks

- $\{M : m_t = (\xi_{G_t^\xi}, \phi_{G_t^\phi})$ for some $(G^\xi, G^\phi)\}$ fails BBN

Mechanisms with Unknown Types

- $\{M : m_t(\xi^{t-1})\}$ fails BBN

Mechanisms with Surveyed Types

- $\{M : m_t(\xi^{t-1}, \tilde{\theta}^{t-1})\}$ satisfies BBN

Mechanisms with Elicited Types

- $\{M : m_t(\xi^{t-1}, \tilde{\theta}^{t})\}$ fails BBN

Mechanisms with Known Types

- $\{M : m_t(\xi^{t-1}, \theta^{t})\}$ satisfies BBN

The Problem with Eliciting Types

Example: Type θ facing two binary signal structures Σ, Σ'

Σ'

H, ✓	1	1		2
L, ✓		1	1	
L, †				6

Σ

- A priori, Σ' uninformative, $p \equiv 1/3$, so choose Σ
- ✓ reverses this: Σ uninformative, $p \equiv 2/3$, so choose Σ'
- For $\theta, \Sigma \cap \Sigma'$, "no harm" lowers consumption, $\hat{c} = 0 < 1 = c$
- Later agents with $\Pr(\theta, \Sigma \cap \Sigma') = 1$ prefer M over $\hat{M}$

Conclusion

Bad news learning

- Guinea-pig relationship raises apparent conflict
- Yet, complete network is best for all agents
- Pareto-ranking over networks

Different information design question

- Trade-offs: KG11's prosecutor; Che, Horner (2018)'s censor
- For which environments does transparency Pareto-improve?

New class of dynamic information mechanisms

- Bergemann, Morris (2019): Unknown, elicited, known types
- Here additionally: Surveyed types