Why Social Media Platforms Shape What You Think

Sarah Cen Asu Ozdaglar James Siderius

Massachusetts Institute of Technology

INFORMS 2022

October 16, 2022

The Influence of Social Media

Social media influences our thoughts & behaviors.

Platforms shape our beliefs by **curating** what we see:

source: iphonelife.com

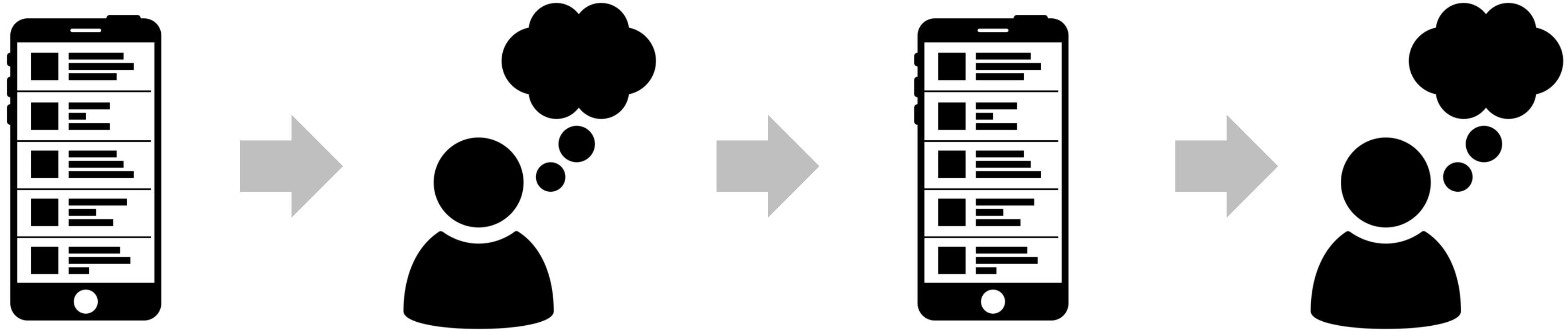
source: gizbot.com

source: later.com

How are platforms incentivized to shape our beliefs?

Belief shaping

Common approach: Users have **fixed preferences**.



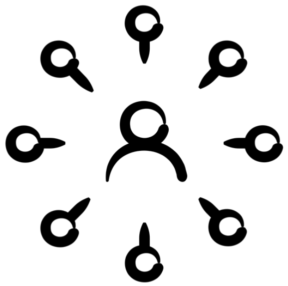
In this work, we study …

How injecting content shapes users' beliefs over time.

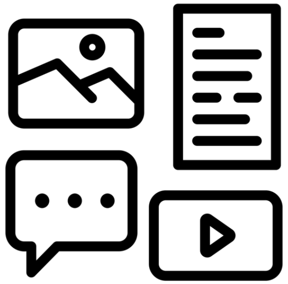
What platforms are incentivized to show users.

Building intuition

Two types of content:

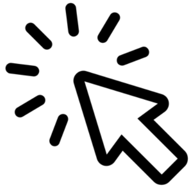


Natural content from social network
(e.g., posts from friends & pages user follows)



Injected content from other sources
(e.g., suggested posts & ads)

Platform wants to maximize engagement + revenue.

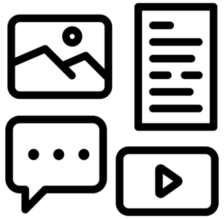


How much user engages depends on their current beliefs.

What platform shows to user influences their future beliefs.

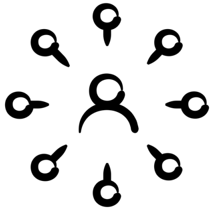
Building intuition

Platform wants to maximize engagement + revenue.



Platform's mechanism: choose what injected content each user sees.

But there are few things out of the platform's control:



Social network

Content sources

Profitability

Uncertainty

Today

I. **Model**
   A. Belief dynamics
   B. Platform objectives
II. **Platform's optimal behavior**
   A. Closed-form solution (with assumptions)
   B. Gradient descent algorithm (assumption-free)
III. **Implications**
   A. Network: How does network affect platform behavior?
   B. Sources: How does presence of extreme sources affect users?
   C. Personalization: How does platform target in different situations?

Related Work

Belief dynamics and strategic influence

Allcott, Braghieri, Eichmeyer, and Gentzkow (2020); Mostagir, Ozdaglar, and Siderius (2022); Candogan, Immorlica, Light, and Anunrojwong (2022); Golub and Jackson (2010); Jadbabaie, Molavi, Sandroni, and Tahbaz-Salehi (2012); DeMarzo, Vayanos, and Zwiebel (2003)

Recommender systems and AI-driven platform algorithms

Cen and Shah (2021); Bozdag and van den Hoven (2015); Helberger, Karppinen, and D'Acunto (2018); DeVito, Gergle, and Birnholtz (2017)

Spread of misinformation and extremism on social media

Bakshy, Messing, and Adamic (2015); Bail, Argyle, Brown, and Volfovsky (2018), Vosoughi, Roy, and Aral (2018); Levy (2021); Acemoglu, Ozdaglar, and Siderius (2022); Mostagir and Siderius (2022)

Model

Model

Belief dynamics: $X_t = \alpha X_{t-1} + (1 - \alpha) C_t$

users' beliefs at time t
(N x d matrix)

continuity

content at time t
(N x d matrix)

Content curation: $C_t = \beta A X_{t-1} + (1 - \beta) B Y$

proportion of
natural content

social network
(N x N matrix)

how platform connects
sources to users
(N x M matrix)

content sources
(M x d matrix)

Platform’s objective

$$X_t = \alpha X_{t-1} + (1-\alpha) C_t$$
$$C_t = \beta A X_{t-1} + (1-\beta) B Y$$

$$\min_{B \in \mathbb{R}^{N \times M}} \sum_{t=1}^{T} \gamma^t \, ||(C_{t+1} - X_t) P||_F^2 + \delta \, ||BQ||_F^2$$

Platform’s objective

$$X_t = \alpha X_{t-1} + (1-\alpha) C_t$$
$$C_t = \beta A X_{t-1} + (1-\beta) BY$$

$$\min_{B \in \mathbb{R}^{N \times M}} \sum_{t=1}^{T} \gamma^t \left|\left|(C_{t+1} - X_t)P\right|\right|_F^2 + \delta \left|\left|BQ\right|\right|_F^2$$

Maximize long-term engagement: the closer content is to user beliefs/preferences, the more engagement it receives (γ = patience).

Platform's objective

$$X_t = \alpha X_{t-1} + (1-\alpha) C_t$$
$$C_t = \beta A X_{t-1} + (1-\beta) BY$$

$$\min_{B \in \mathbb{R}^{N \times M}} \sum_{t=1}^{T} \gamma^t \, ||(C_{t+1} - X_t)P||_F^2 + \delta \, ||BQ||_F^2$$

Maximize long-term engagement: the closer content is to user beliefs/preferences, the more engagement it receives (γ = patience).

Ensure content sources are happy: regularization term that encourages giving each source some viewership (disappears if $\delta = 0$).

Platform's objective

$$X_t = \alpha X_{t-1} + (1-\alpha)C_t$$
$$C_t = \beta A X_{t-1} + (1-\beta)BY$$

$$\min_{B \in \mathbb{R}^{N \times M}} \sum_{t=1}^{T} \gamma^t \,||(C_{t+1} - X_t)P||_F^2 + \delta \,||BQ||_F^2$$

Maximize long-term engagement: the closer content is to user beliefs/preferences, the more engagement it receives (γ = patience).

Ensure content sources are happy: regularization term that encourages giving each source some viewership (disappears if $\delta = 0$).

Prioritize profitable content: use profitable sources more often (default to $\mathrm{I}_{M \times M}$).

Platform's objective

$$X_t = \alpha X_{t-1} + (1-\alpha)C_t$$
$$C_t = \beta A X_{t-1} + (1-\beta)BY$$

$$\min_{B \in \mathbb{R}^{N\times M}} \sum_{t=1}^{T} \gamma^t \, ||(C_{t+1} - X_t)P||_F^2 + \delta \, ||BQ||_F^2$$

Maximize long-term engagement: the closer content is to user beliefs/preferences, the more engagement it receives (γ = patience).

Ensure content sources are happy: regularization term that encourages giving each source some viewership (disappears if $\delta = 0$).

Prioritize profitable content: use profitable sources more often (default to $\mathrm{I}_{M\times M}$).

Prioritize certain types engagement: e.g., due to uncertainty in the platform's knowledge of users' beliefs along certain dimensions (default to $\mathrm{I}_{d\times d}$).

Platform's objective

$$X_t = \alpha X_{t-1} + (1-\alpha) C_t$$
$$C_t = \beta A X_{t-1} + (1-\beta) BY$$

$$\sum_{t=1}^{T} \gamma^t \, ||C_{t+1} - X_t||_F^2$$

Maximize long-term engagement: the closer content is to user beliefs/preferences, the more engagement it receives (γ = patience).

Ensure content sources are happy: regularization term that encourages giving each source some viewership (disappears if $\delta = 0$).

Prioritize profitable content: use profitable sources more often (default to $I_{M \times M}$).

Prioritize certain types engagement: e.g., due to uncertainty in the platform's knowledge of users' beliefs along certain dimensions (default to $I_{d \times d}$).

Optimal platform behavior

Closed-form solution

Suppose we consider the simple engagement objective:

$$\sum_{t=1}^{T} \gamma^t \, ||C_{t+1} - X_t||_F^2$$

Theorem 1. If entries in B can take any real value,

$$B^* = (1-\beta)\left(\sum_{t=0}^{T-1} \gamma^t D_t^+ D_t\right)\left(\sum_{t=0}^{T-1} D_t^\top (\beta A - \mathbb{I})\tilde{A}^t X_0 Y^+\right)$$

$$\tilde{A} = \alpha \mathbb{I} + (1-\alpha)\beta A \qquad D_t = (1-\alpha)(\beta A - \mathbb{I})\sum_{\tau=0}^{t-1} \tilde{A}^\tau + \mathbb{I}$$

Algorithm

$$\min_{B \in \Delta_M^N} \sum_{t=1}^{T} \gamma^t \, ||(C_{t+1} - X_t)P||_F^2 + \delta \, ||BQ||_F^2$$

Can solve this problem using projected gradient descent.

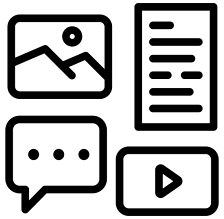
1. Initialize row-stochastic $B^{(0)}$.
2. Take gradient step (nice closed-form expression).
3. Project onto probability simplex.
4. Stop if $B^{(t)}$ and $B^{(t-1)}$ are close; otherwise, repeat 2-4.

Guaranteed to converge to global minimum!
Main proof step is to show that Hessian is PSD.

Simulations

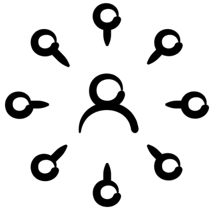
Reminder

Platform wants to maximize engagement + revenue.



Platform's mechanism: choose what injected content each user sees.

Exogeneous factors:



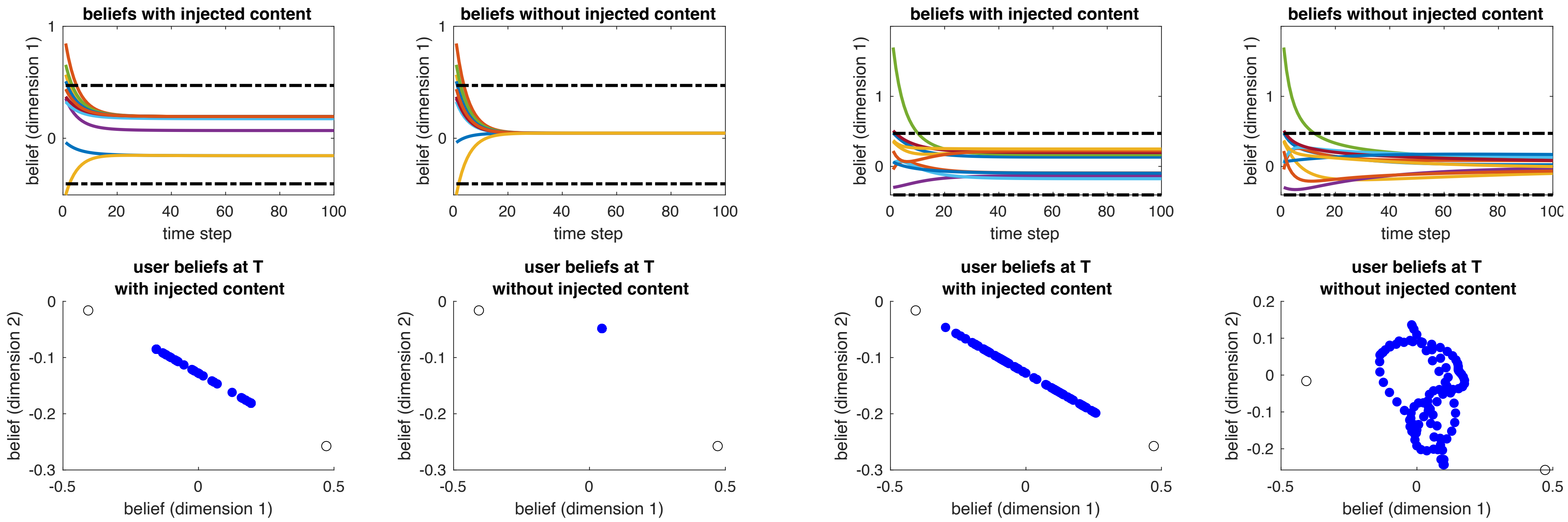
Social network

Content sources

Profitability

Uncertainty

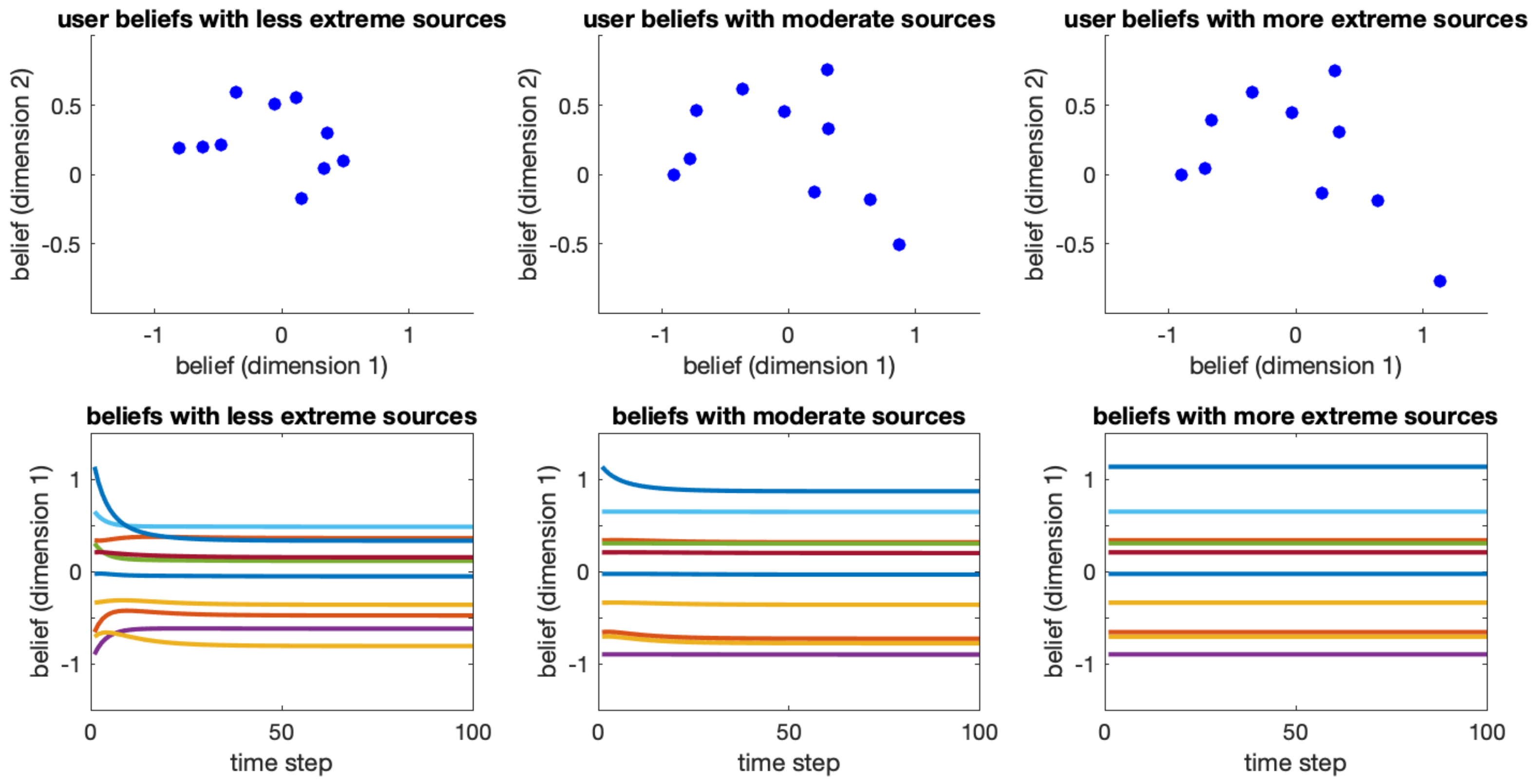
How does network affect user beliefs?



Intuition: Content sources = stubborn agents. Sparse network + stubborn agents → less mixing. Platform incentivized to cater rather than drive consensus.

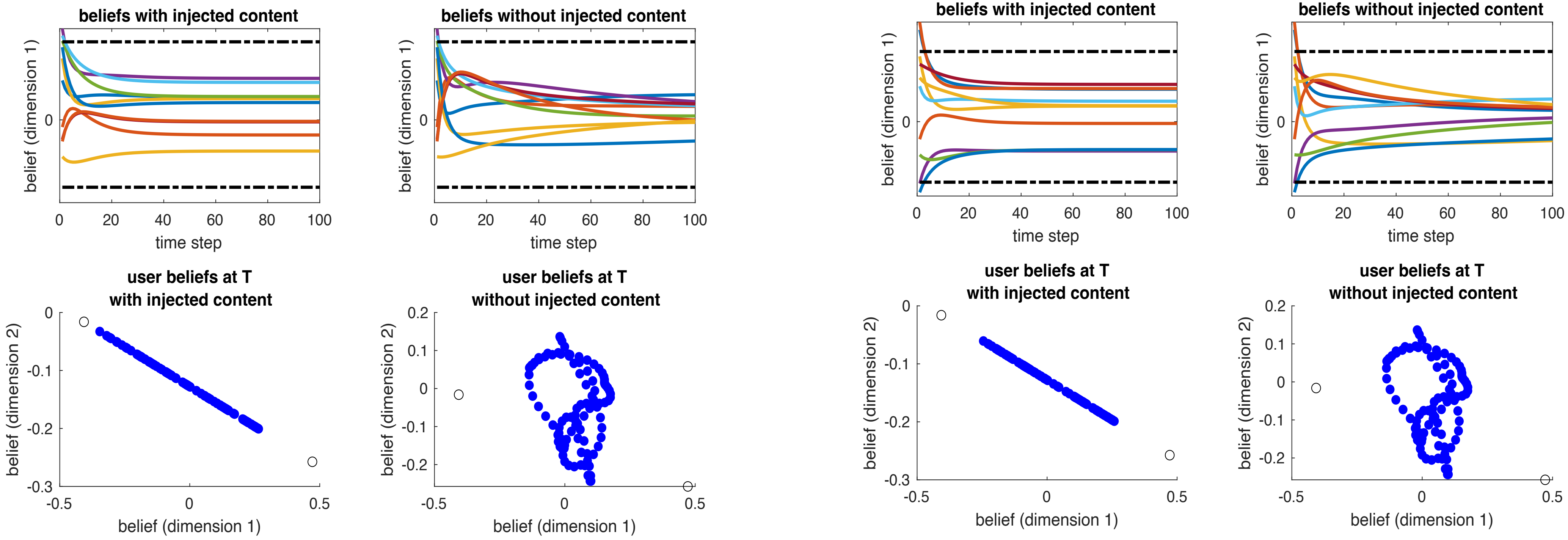
How do extreme sources affect beliefs?

$$Y = \begin{pmatrix} Y_{\text{base}} \\ \zeta Y_{\text{base}} \end{pmatrix}$$



Intuition: Extreme sources → permit extreme beliefs. Not always b/c platform drives to extremes - can cater to extreme users, who pull neighbors extreme.

What happens if the platform is patient?



Intuition: a patient platform induces lower consensus → makes extreme users slightly more moderate, then uses them like stubborn agents

Recommendations

We study how the platform is incentivized to connect sources to users.

This framework can help platforms design their editorial policies.

Findings: interesting interplay between personalization and consensus.

- Content sources = stubborn agents → disincentivize consensus, especially in sparse networks.
  - Suggests serendipitous connections can better drive consensus, when desirable.
- Extreme sources permit extreme beliefs *not* because drive extreme users more extreme, but because can cater to extreme users and then allow them to take over.
  - Suggests that have to do more than not show extreme content to moderate users.
- Polarization cannot be solved with patience. In fact, patience can favor extreme beliefs.
  - Suggests that platforms often rely on extreme platforms a bit too much.

Thank you!

shcen@mit.edu