Crowding out the truth? A simple model of misinformation, polarization and meaningful social interactions

Fabrizio Germano
U Pompeu Fabra

Vicenç Gómez
U Pompeu Fabra

Francesco Sobbrio
Tor Vergata U. Rome

Computational Politics e-symposium

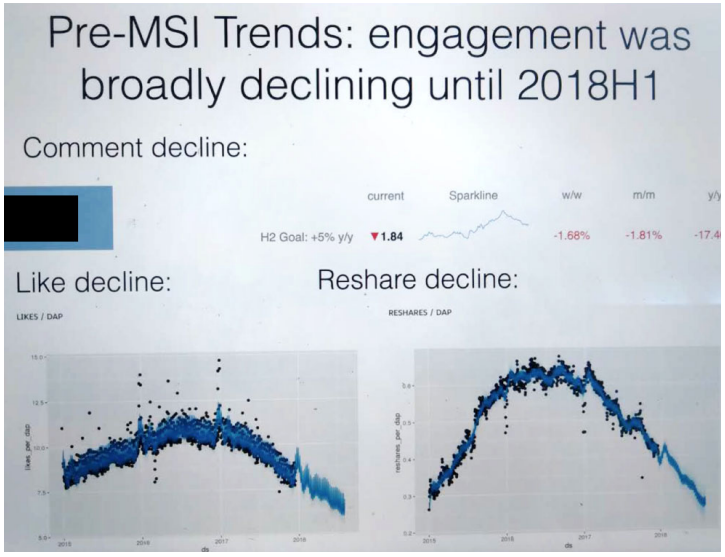
Ranking matters for users!

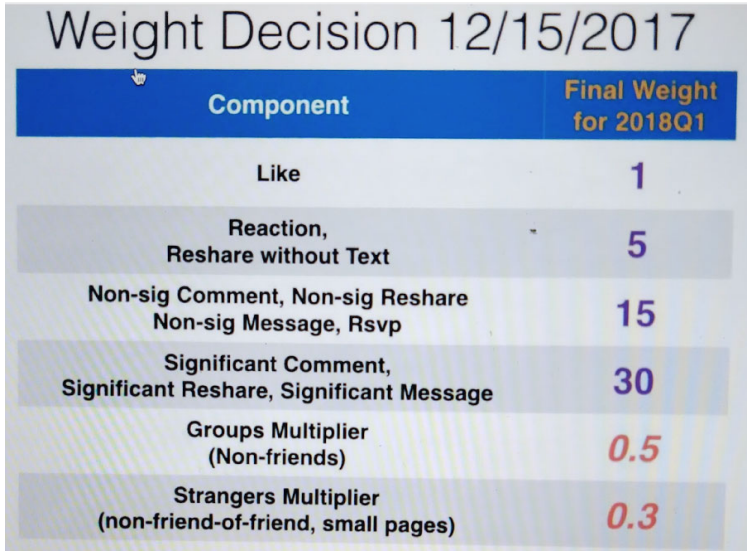
Pan et al. (2007); Novarese & Wilson (2013)

Yom-Tov et al. (2013); Glick et al. (2014)

Epstein & Robertson (2015)

Engagement matters for platforms!





Facebook whistleblowers (*WSJ*, 2021): MSI allegedly led to adverse effects in terms of misinformation and polarization (among others)

This paper

1) Theoretical framework

Interactions of behavioural individuals with algorithmic weights

$\Rightarrow$ Assess impact of an increase in MSI & personalization on:

- Platform Engagement
- Misinformation
- Polarization

Main insights:

- MSI: $\uparrow$ Engagement; $\uparrow$ Misinformation; $\uparrow$ Polarization
- Personalization: $\uparrow$ Engagement; $\uparrow$ Polarization

2) Direct empirical evidence on impact of MSI on polarization

▸ Lit.

Model

State of the world $\theta \in \mathbb{R}$ (e.g., net benefits of vaccines/emission reduction)

- M **news items** (e.g., Facebook's post, Tweet, etc).
  - Each carries an informative signal $y_m \sim N(\theta, \sigma_y^2)$.
- N **individuals**:
  - Each receives a private informative signal $x_n \sim N(\theta, \sigma_x^2)$.
  - Sequentially access (in random order) a social media platform to read and, possibly, "highlight" (e.g., share) a news item m
  - Are able to see whether m is "like-minded" or not. Yet they need to click on the news item in order to see y_m.

Model – Clicking *(absent ranking)*

γ_n = individual n's propensity to click on "like-minded" news, *absent ranking*

Individuals can be of three **clicking types**:

- *confirmatory* (τ_C): more likely to click on "like-minded" news ($\gamma_C > 1/2$)
- *exploratory* (τ_E): less likely to click on "like-minded" news ($\gamma_E < 1/2$)
- *indifferent (ranking-driven)* (τ_I): $\gamma_I = 1/2$

The three types occur with probabilities $p_C \geq 0, p_E \geq 0$, & $p_I = 1 - p_C - p_E$.

Model – Highlighting

- After clicking on m, individual sees the actual signal y_m
- Then *highlight* (e.g., share/comment) m with probability p_a

Assumptions:

- Highlight only if sufficiently close to prior
  ($|x_n - y_m| < \frac{\sigma_x}{2}$, An et al. 2014; Garz et al 2020)

- Individuals with more extreme priors are more likely to highlight.
  (Bakshy, Messing, Adamic, 2015, for "hard" news (i.e., political).

▸ Bakshy et. al (2015)

Model – *Attention Bias*

Individuals have an attention bias calibrated by $\beta \geq 1$.

Interpretation:
If news items m_a and m_b have the same sign and m_a is one position up in the ranking $\Rightarrow$ m_a will be β times more likely to be clicked wrt to m_b

Model – *Attention Bias*

Individuals have an attention bias calibrated by $\beta \geq 1$.

Interpretation:
If news items m_a and m_b have the same sign and m_a is one position up in the ranking $\Rightarrow$ m_a will be β times more likely to be clicked wrt to m_b

- All in all, the higher:
  - a) the ranking of news item m;
  - b) the propensity (*absent ranking*) of individual n to click on m

  $\Rightarrow$ the more likely m is to be clicked ▸ Clicking Prob.

  (Germano, Gómez, Le Mens 2019; Germano and Sobbrio, 2020)

Model – Algorithm: Popularity Ranking

Ranking algorithm updates popularity of each news item such that:

- a click has a weight of 1
- a highlight has a weight of $\eta \in \mathbb{R}_+$.

Popularity of news item m, $\kappa_{n,m}$ updated according to:

$$\kappa_{n,m} = \kappa_{n-1,m} + \begin{cases} 0 & \text{if } m \text{ is not clicked on by } n \\ 1 & \text{if } m \text{ is clicked on and not highlighted by } n \\ 1+\eta & \text{if } m \text{ is clicked on and highlighted by } n \end{cases}$$

Ranking observed by n inversely related to popularity before clicking:

$$r_{n,m} < r_{n,m'} \iff \kappa_{n-1,m} < \kappa_{n-1,m'}.$$

Recap

At time $t = n$, (random) individual n:

- Gets private signal x_n on θ (e.g., net benefits of vaccine)
- Access social media and observes ranking of news items $r_{n,m}$
- Given ranking, attention bias β, and propensity to choose like-minded items γ_n: decides which m to click
- After learning y_m, highlights m with probability p_a and only if sufficiently close to her prior
- Algorithm updates the popularity (and ranking) of items:

$$\kappa_{n,m} = \kappa_{n-1,m} + \begin{cases} 0 & \text{if } m \text{ is not clicked on by } n \\ 1 & \text{if } m \text{ is clicked on and not highlighted by } n \\ 1+\eta & \text{if } m \text{ is clicked on and highlighted by } n. \end{cases}$$

At time $t + 1 = n + 1$....

Model – Algorithm: Popularity & Personalized ranking

Algorithm personalizes the ranking according to whether x_n on the left/right wrt θ (group L/R)

Two rankings based on two separate measures of popularity.

$\lambda \in [0, 1]$ is a parameter calibrating the degree of personalization.

- $\lambda = 0$: clicks and highlights from the other group do not count at all
- $\lambda = 1$: no personalization: popularity for both groups always coincides.

Evaluation indices

Effects of η and λ on engagement and users' welfare?

- *Engagement*: $ENG =$ sum of clicks and highlights
- *Misinformation*: $MIS = \frac{1}{N} \sum_{n \in N} |y(n) - \theta|$;
- *Polarization*: $POL = \frac{1}{N} |\sum_{n \in R} y(n) - \sum_{n' \in L} y(n')|$;

where:

- $y(n) \in M$ denotes the signal of the news item clicked on by individual n.
- L (R) denotes the individuals with signals x_n with $\text{sign}(x_n) = -1$ ($= +1$).

Main results – Crowding out the truth

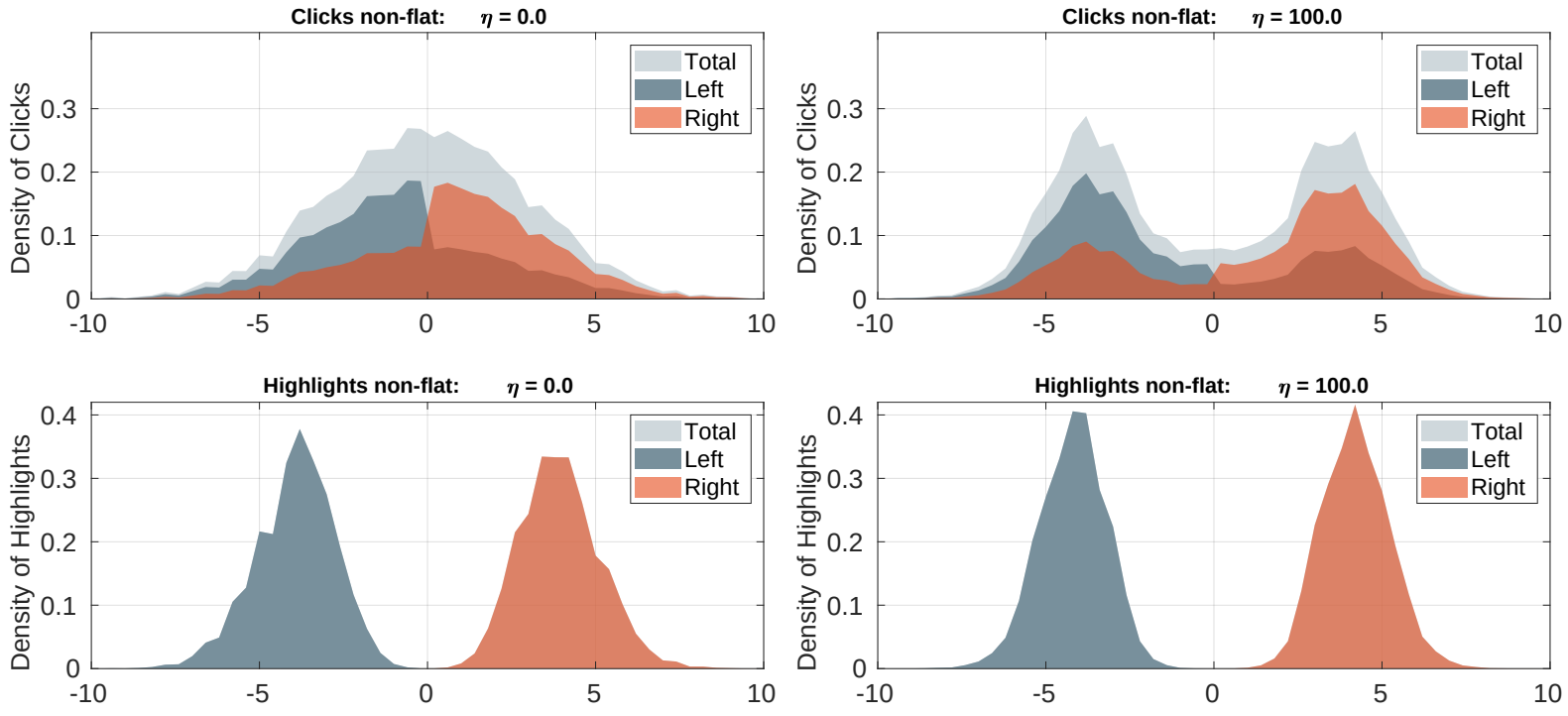
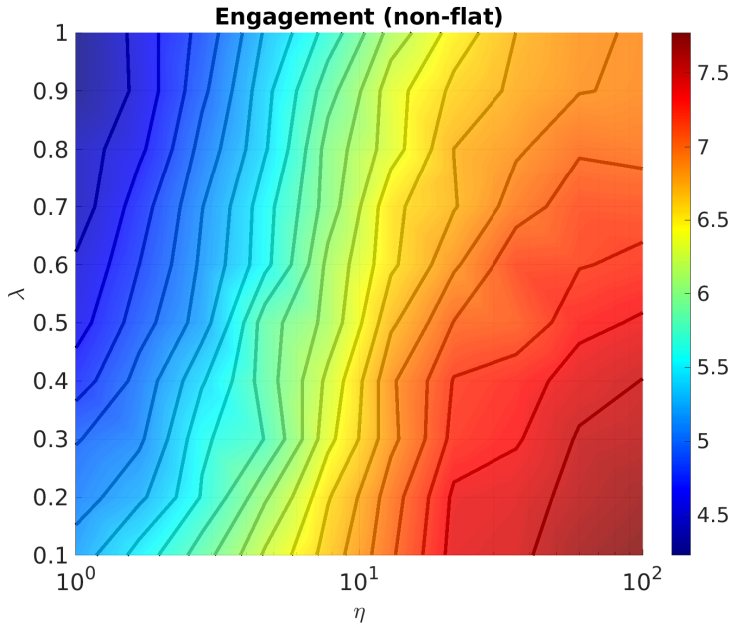
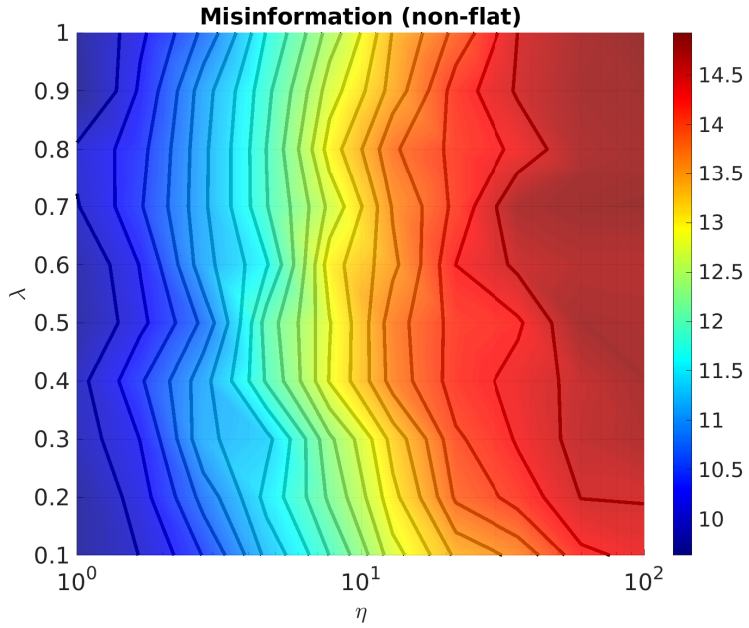
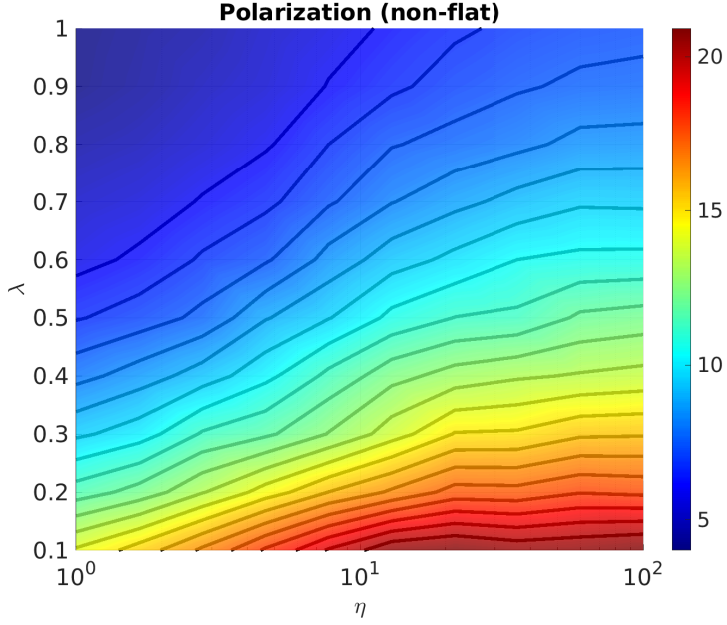


Figure: Users' clicking behavior (top) and highlighting behavior (bottom) for small η (left) and for large η (right) under non-flat highlighting.







▸ Analytical ▸ Idyosincratic ▸ Non-centered ▸ Flat

Main results: Intuition

An increase in η:

- More individuals willing to highlight items (more extremist) will be clicking on items they are actually interested in highlighting $\rightarrow$ Higher engagement

- Individuals less likely to click on news near the truth (y's $\approx \theta$) & more likely to click on items further away from the truth (y's $\approx -x^*, x^*$). $\rightarrow$ More misinformation and polarization

  $\Rightarrow$ *Crowding out the truth*.

MSI and Polarization: Empirical analysis

Theoretical prediction: an increase in weight of "highlights" (η)

$$\Downarrow$$

- Individuals more exposed to extremists contents
- Higher level of political polarization.

Empirical test: exploit Facebook's MSI update

Jan 2018: boost in the weight given to comments and shares

Empirical analysis (2)

Data. Focus on Italy (IPSOS *Polimetro*):

- Weekly interviews on representative sample of Italian voting pop.

- Info on whether internet primary source to form pol. opinion
  - Italy 2017-2018: FB by far the first social media: 60% penetration rate (Twitter 23%), $\sim$ 80% among internet users
- Ideological self-position: Dummy for moderate/non-moderate
- Probability of voting for each party: affective polarization.

Empirical strategy

Difference-in-Differences:

$$Y_{i,m,t} = \alpha + \beta_1 \texttt{Opinion via internet}_{i,m,t} \times \texttt{Post MSI} + \beta_2 \texttt{Opinion via internet}_{i,m,t} + \beta_3 \texttt{Post MSI} + \alpha_m + X_{i,t} + \varepsilon_{i,m,t} \quad (1)$$

- $Y_{i,m,t}$ represents the outcome of interest relative to individual i, leaving in municipality m interviewed in the survey wave t (i.e., probability of declaring a non-moderate political ideology or weighted affective polarization).
- α_m municipality fixed effect
- $X_{i,t}$: socio-demographic control (age, gender, n. of resident family members, level of education, type of occupation, religiosity).
- Observations weighted according to Ipsos sampling weights

Results: MSI and Non-moderate ideology

Table: MSI and non-moderate ideological position

	(1) Non-moderate Ideology	(2) Non-moderate Ideology	(3) Non-moderate Ideology	(4) Non-moderate Ideology
Opinion via internet × Post MSI	0.062***	0.058***	0.051***	0.051***
	(0.016)	(0.015)	(0.014)	(0.018)
Opinion via internet	-0.012	-0.006	-0.012	-0.012
	(0.020)	(0.020)	(0.024)	(0.022)
Post MSI	-0.017*			
	(0.009)			
Observations	25,690	25,690	25,690	25,690
Mean outcome	0.36	0.36	0.36	0.36
SD outcome	0.48	0.48	0.48	0.48
Municipality FE	YES	YES	YES	YES
Date of interview FE	NO	YES	NO	NO
Province-Date of interview FE	NO	NO	YES	YES
Cluster SE	Region	Region	Region	Province

Note: Time horizon: June 2017-June 2018 . Robust Standard Errors in parenthesis. ** p<0.01, ** p<0.05, * p<0.1

Results: MSI and Affective Polarization

Table: MSI and Affective Polarization

	(1) Affective Polarization	(2) Affective Polarization	(3) Affective Polarization	(4) Affective Polarization
Opinion via internet × Post MSI	0.054**	0.055**	0.073***	0.073***
	(0.024)	(0.024)	(0.019)	(0.025)
Opinion via internet	-0.012	-0.011	-0.006	-0.006
	(0.023)	(0.022)	(0.023)	(0.025)
Post MSI	0.118***			
	(0.020)			
Observations	14,499	14,499	14,499	14,499
Mean outcome	1.29	1.29	1.29	1.29
SD outcome	0.61	0.61	0.61	0.61
Municipality FE	YES	YES	YES	YES
Date of interview FE	NO	YES	NO	NO
Province-Date of interview FE	NO	NO	YES	YES
Cluster SE	Region	Region	Region	Province

Note: Time horizon: June 2017-June 2018 . Robust Standard Errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Conclusions

A higher η ($\uparrow$ weight on highlights in the ranking algorithm)

- Assuming bimodal propensity to highlight (Bakshy et al. 2015):
  - increases engagement
  - increases polarization
  - increases misinformation.
- Higher ideological extremism & affective polarization in Italy

$\Rightarrow$ Theoretical & Empirical evidence on adverse effects of FB 2018 MSI update

- A lower λ ($\uparrow$ personalization) increases engagement & polarization.

$\Rightarrow$ Theoretical support for "filter bubble" (Pariser, 2011)

literature