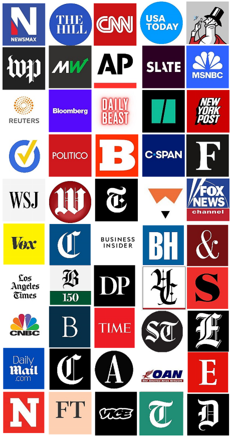
Personalization, Engagement, and Content Quality on Social Media: An Evaluation of Reddit's News Feed

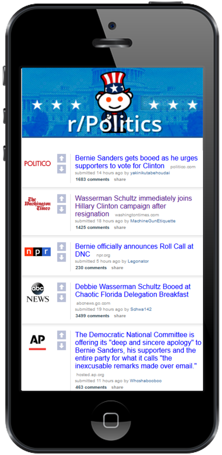
Alex Moehring

February 13, 2025

Social media platforms are gatekeepers of information



Social media platforms are gatekeepers of information



Widespread belief that less credibile content is more engaging

"One of the biggest issues social networks face is that, when left unchecked, people will engage disproportionately with more sensationalist and provocative content.

...

The category we're most focused on is click-bait and misinformation. People consistently tell us these types of content make our services worse -- even though they engage with them."

— Mark Zuckerberg, 2018

Research objectives

1. How does the ranking algorithm affect engagement?
   – Estimate treatment effect of rank using a regression discontinuity
2. Estimate users preferences to comment on various publishers
   – Estimate a micro-founded choice model of engagement decisions
3. Do engagement based rankings promote low credibility publishers?
   – Use choice model to estimate engagement patterns of counterfactual ranking algorithms
4. What is the cost of including credibility in ranking algorithm objective?
   – Plot the efficient frontier between credibility and quantity

PURDUE UNIVERSITY

Research objectives

1. How does the ranking algorithm affect engagement?
   - Estimate treatment effect of rank using a regression discontinuity
2. Estimate users preferences to comment on various publishers
   - Estimate a micro-founded choice model of engagement decisions
3. Do engagement based rankings promote low credibility publishers?
   - Use choice model to estimate engagement patterns of counterfactual ranking algorithms
4. What is the cost of including credibility in ranking algorithm objective?
   - Plot the efficient frontier between credibility and quantity

Research objectives

1. How does the ranking algorithm affect engagement?
   - Estimate treatment effect of rank using a regression discontinuity
2. Estimate users preferences to comment on various publishers
   - Estimate a micro-founded choice model of engagement decisions
3. Do engagement based rankings promote low credibility publishers?
   - Use choice model to estimate engagement patterns of counterfactual ranking algorithms
4. What is the cost of including credibility in ranking algorithm objective?
   - Plot the efficient frontier between credibility and quantity

Research objectives

1. How does the ranking algorithm affect engagement?
   - Estimate treatment effect of rank using a regression discontinuity
2. Estimate users preferences to comment on various publishers
   - Estimate a micro-founded choice model of engagement decisions
3. Do engagement based rankings promote low credibility publishers?
   - Use choice model to estimate engagement patterns of counterfactual ranking algorithms
4. What is the cost of including credibility in ranking algorithm objective?
   - Plot the efficient frontier between credibility and quantity

Preview of findings

1. How does the ranking algorithm affect engagement?
   - **Ranking algorithm steers engagement to promoted articles**
2. Understand users preferences to comment on various publishers
   - **Substantial heterogeneity in preferences for publisher slant and credibility**
3. Do engagement based rankings promote low credibility publishers?
   - **Small positive impact on credibility of content majority of users engage with**
   - **Large increase in low-credibility engagement for subset of users**
4. What is the cost of including credibilitly in ranking algorithm objective?
   - **Good news: Platform can substantially increase credibility for modest cost**

Preview of findings

1. How does the ranking algorithm affect engagement?
   - **Ranking algorithm steers engagement to promoted articles**
2. Understand users preferences to comment on various publishers
   - **Substantial heterogeneity in preferences for publisher slant and credibility**
3. Do engagement based rankings promote low credibility publishers?
   - **Small positive impact on credibility of content majority of users engage with**
   - **Large increase in low-credibility engagement for subset of users**
4. What is the cost of including credibilitly in ranking algorithm objective?
   - **Good news: Platform can substantially increase credibility for modest cost**

Preview of findings

1. How does the ranking algorithm affect engagement?
   - **Ranking algorithm steers engagement to promoted articles**
2. Understand users preferences to comment on various publishers
   - **Substantial heterogeneity in preferences for publisher slant and credibility**
3. Do engagement based rankings promote low credibility publishers?
   - **Small positive impact on credibility of content majority of users engage with**
   - **Large increase in low-credibility engagement for subset of users**
4. What is the cost of including credibilitly in ranking algorithm objective?
   - **Good news: Platform can substantially increase credibility for modest cost**

Preview of findings

1. How does the ranking algorithm affect engagement?
   - **Ranking algorithm steers engagement to promoted articles**
2. Understand users preferences to comment on various publishers
   - **Substantial heterogeneity in preferences for publisher slant and credibility**
3. Do engagement based rankings promote low credibility publishers?
   - **Small positive impact on credibility of content majority of users engage with**
   - **Large increase in low-credibility engagement for subset of users**
4. What is the cost of including credibilitly in ranking algorithm objective?
   - **Good news: Platform can substantially increase credibility for modest cost**

PURDUE UNIVERSITY

Outline

1. Setting and Data
2. Estimating position effects
3. Analysis of counterfactual algorithms
4. Conclusion
5. Appendix

Research setting: Reddit politics community

Background:

- Reddit is a large social media platform
  - Organized into virtual communities
- Politics: Centers on US political news

Ideal setting:

- Large community
- Feasible
- Neutral measure of credibility

Data

Combine three types of data:

1. State of the feed at various points in time
   - Historical snapshots of subreddit feed from Wayback Machine

2. User engagement data
   - Panel of user-level comments [Baumgartner et al., 2020]

3. Vertical and horizontal publisher features
   - Publisher slant and credibility ratings [Robertson et al., 2018, Lin et al., 2022]

Data

Combine three types of data:

1. State of the feed at various points in time
   – Historical snapshots of subreddit feed from Wayback Machine

2. User engagement data
   – Panel of user-level comments [Baumgartner et al., 2020]

3. Vertical and horizontal publisher features
   – Publisher slant and credibility ratings [Robertson et al., 2018, Lin et al., 2022]

PURDUE UNIVERSITY

Data

Combine three types of data:

1. State of the feed at various points in time
   - Historical snapshots of subreddit feed from Wayback Machine
2. User engagement data
   - Panel of user-level comments [Baumgartner et al., 2020]
3. Vertical and horizontal publisher features
   - Publisher slant and credibility ratings [Robertson et al., 2018, Lin et al., 2022]

Data

Combine three types of data:

1. State of the feed at various points in time
   – Historical snapshots of subreddit feed from Wayback Machine

2. User engagement data
   – Panel of user-level comments [Baumgartner et al., 2020]

3. Vertical and horizontal publisher features
   – Publisher slant and credibility ratings [Robertson et al., 2018, Lin et al., 2022]

PURDUE UNIVERSITY

Outline

Estimating causal effect of rank

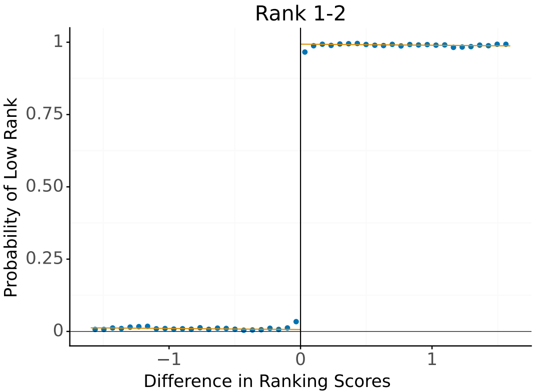
- Article rank is endogenous
- Source code reveals ranking algorithm
- Algorithm creates ranking score for each article
  - Combines article age with votes

Estimating causal effect of rank

- Article rank is endogenous
- Source code reveals ranking algorithm
- Algorithm creates ranking score for each article
  - Combines article age with votes

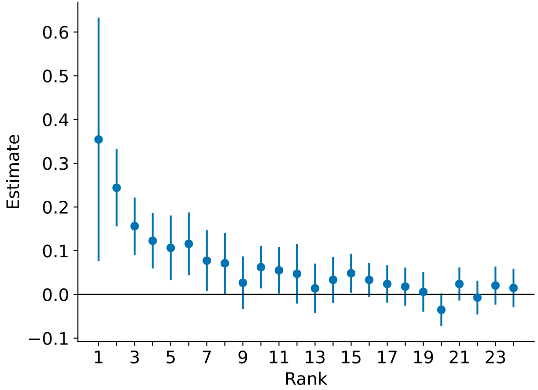
```
cpdef double _hot(long ups, long downs, double date):
    """The hot formula. Should match the equivalent function in postgres."""
    s = score(ups, downs)
    order = log10(max(abs(s), 1))
    if s > 0:
        sign = 1
    elif s < 0:
        sign = -1
    else:
        sign = 0
    seconds = date - 1134028003
    return round(sign * order + seconds / 45000, 7)
```

First stage confirms posts ranked according to reconstructed algorithm



▸ Other Positions ▸ Balance ▸ Discontinuity Frontier

Algorithm focuses engagement on highly ranked posts



RD Outcome Plots | Bandwidth Robustness | Window Robustness | Donut Robustness | Effects by Category | Balance: Post Age | Discontinuity Frontier

Outline

1. Setting and Data
2. Estimating position effects
3. Analysis of counterfactual algorithms
4. Conclusion
5. Appendix

Discrete-choice model of engagement decisions

Choice model:

$$d_{ijt} = 1\left[v_{ijt} = 1\right] 1\left[u_{ijt} \geq u_{i0t}\right]$$

- d_{ijt}: comment decision by user i on article j in period t
- v_{ijt}: independent Bernoulli draw with probability $p\left(r_{jt}, t\right) = p_t p_r$
- u_{ijt}: utility for commenting on article j
  - $u_{ijt} = x_j' \beta_i + \xi_{jt} + \varepsilon_{ijt}$ where ε_{ijt} ∼Type 1 EV
- See paper for identification and estimation

Endogenous Search

Discrete-choice model of engagement decisions

Choice model:

$$d_{ijt} = 1\left[v_{ijt} = 1\right] 1\left[u_{ijt} \geq u_{i0t}\right]$$

- d_{ijt}: comment decision by user i on article j in period t
- v_{ijt}: independent Bernoulli draw with probability $p\left(r_{jt}, t\right) = p_t p_r$
- u_{ijt}: utility for commenting on article j
  - $u_{ijt} = x_j' \beta_i + \xi_{jt} + \varepsilon_{ijt}$ where ε_{ijt} ∼Type 1 EV
- See paper for identification and estimation

▸ Endogenous Search

Discrete-choice model of engagement decisions

Choice model:

$$d_{ijt} = 1\left[v_{ijt} = 1\right] 1\left[u_{ijt} \geq u_{i0t}\right]$$

- d_{ijt}: comment decision by user i on article j in period t
- v_{ijt}: independent Bernoulli draw with probability $p\left(r_{jt}, t\right) = p_t p_r$
- u_{ijt}: utility for commenting on article j
  - $u_{ijt} = x_j' \beta_i + \xi_{jt} + \varepsilon_{ijt}$ where ε_{ijt} ∼Type 1 EV
- See paper for identification and estimation

Endogenous Search

Discrete-choice model of engagement decisions

Choice model:

$$d_{ijt} = 1\left[v_{ijt} = 1\right] 1\left[u_{ijt} \geq u_{i0t}\right]$$

- d_{ijt}: comment decision by user i on article j in period t
- v_{ijt}: independent Bernoulli draw with probability $p\left(r_{jt}, t\right) = p_t p_r$
- u_{ijt}: utility for commenting on article j
  - $u_{ijt} = x_j' \beta_i + \xi_{jt} + \varepsilon_{ijt}$ where ε_{ijt} $\sim$Type 1 EV
- See paper for identification and estimation

Endogenous Search

Counterfactual ranking algorithms

Mechanics:

- Platform estimates engagement probability conditional on exposure ($\hat{P}_{ijt}$)
- Platform can then shift attention to various articles through their ranking algorithm

Counterfactual ranking algorithms:

1. Reddit's actual ranking algorithm
2. Random benchmark
3. Non-personalized engagement maximizing
   - Sort posts in descending order of $E\left[\hat{P}_{ijt}\right]$
4. Personalized engagement maximizing
   - For each user, sort posts in descending order of $\hat{P}_{ijt}$

EM Derivation

Counterfactual ranking algorithms

Mechanics:

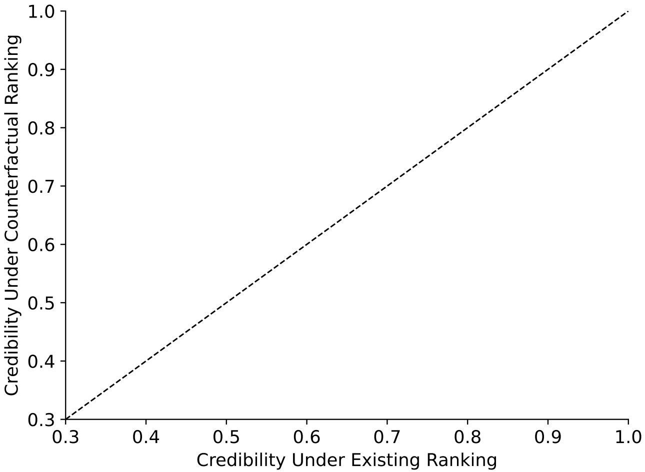
- Platform estimates engagement probability conditional on exposure ($\hat{P}_{ijt}$)
- Platform can then shift attention to various articles through their ranking algorithm

Counterfactual ranking algorithms:

1. Reddit's actual ranking algorithm
2. Random benchmark
3. Non-personalized engagement maximizing
   - Sort posts in descending order of $E\left[\hat{P}_{ijt}\right]$
4. Personalized engagement maximizing
   - For each user, sort posts in descending order of $\hat{P}_{ijt}$

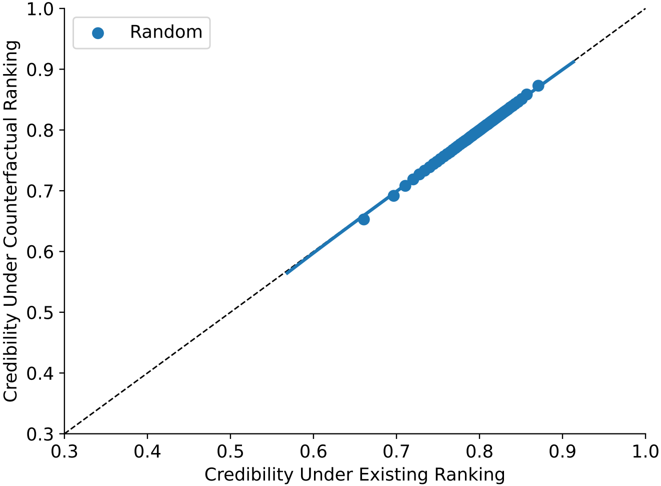
▸ EM Derivation

How does engagement maximization impact credibility?



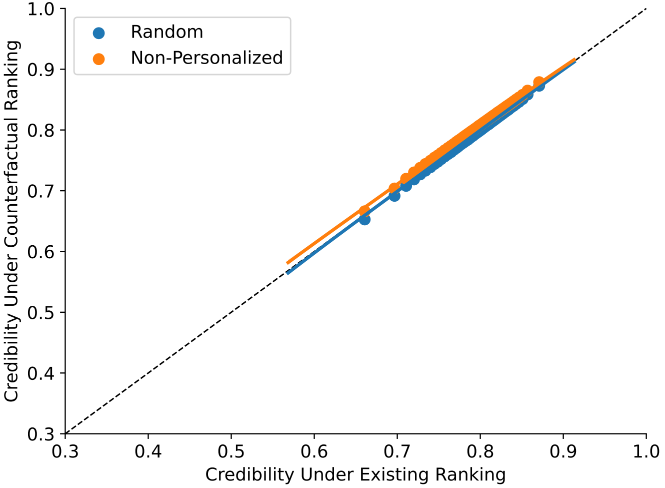
▸ CDFs

How does engagement maximization impact credibility?



▸ CDFs

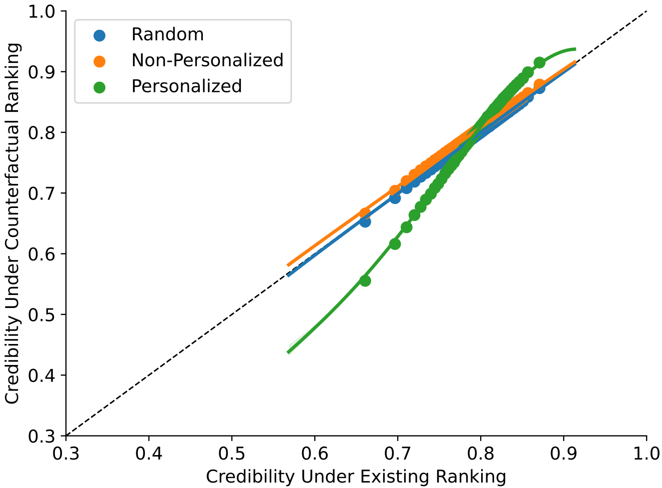
How does engagement maximization impact credibility?



CDFs

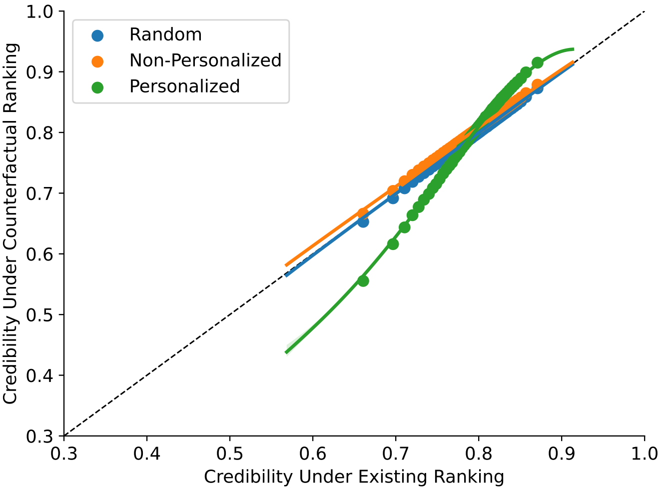
PURDUE UNIVERSITY

How does engagement maximization impact credibility?



▸ CDFs

How does engagement maximization impact credibility?

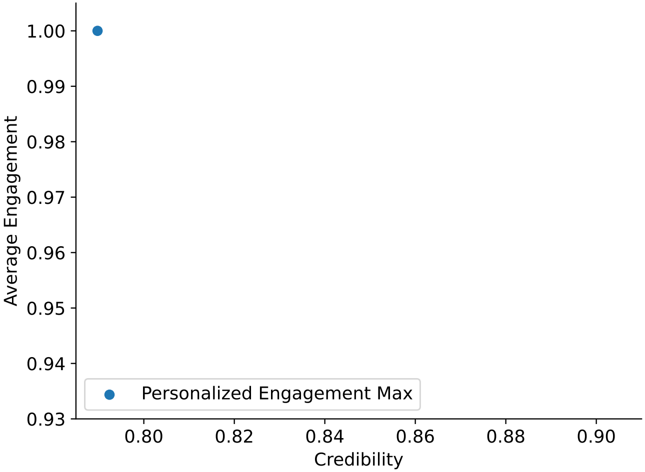


✓ Personalization exacerbates differences in credibility

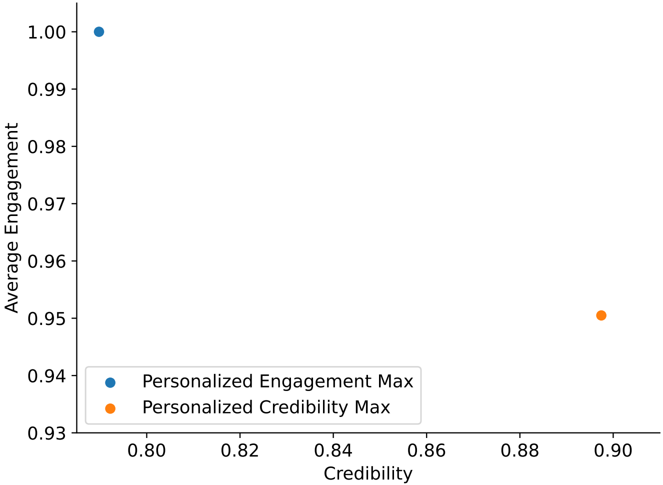
▸ CDFs

PURDUE UNIVERSITY

What is the cost of incorporating credibility?

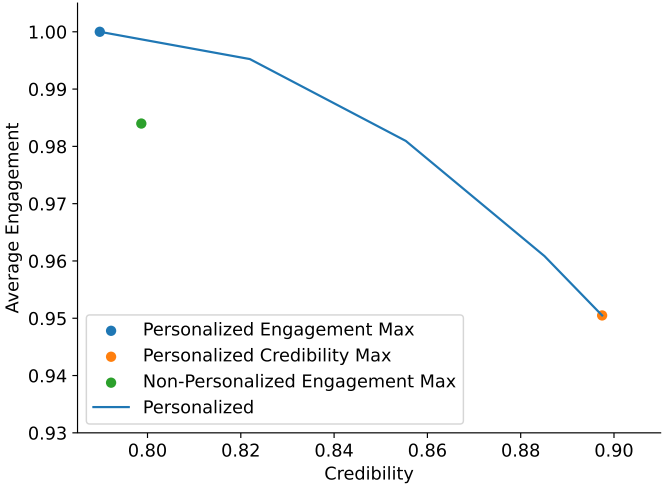


What is the cost of incorporating credibility?

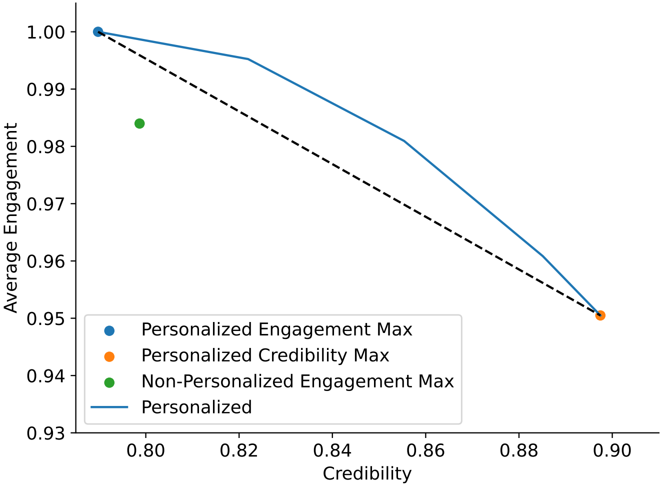


PURDUE UNIVERSITY

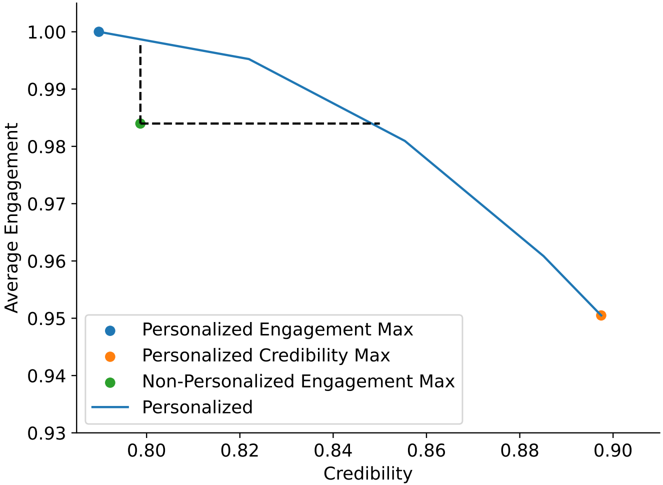
What is the cost of incorporating credibility?



What is the cost of incorporating credibility?



What is the cost of incorporating credibility?



PURDUE UNIVERSITY

Additional findings

1. Personalized engagement maximization reduces engagement diversity
   - Driven by users with preferences for extreme publishers

2. Personalized engagement maximization increases demand for slanted and low-credibility publishers

3. Engagement maximization slightly increases the share of negative sentiment comments
   - Largest increase for users preferring left-leaning and low-credibility publishers

Diversity Market Shares Conclusion

Additional findings

1. Personalized engagement maximization reduces engagement diversity
   - Driven by users with preferences for extreme publishers
2. Personalized engagement maximization increases demand for slanted and low-credibility publishers
3. Engagement maximization slightly increases the share of negative sentiment comments
   - Largest increase for users preferring left-leaning and low-credibility publishers

Diversity Market Shares Conclusion

Additional findings

1. Personalized engagement maximization reduces engagement diversity
   - Driven by users with preferences for extreme publishers
2. Personalized engagement maximization increases demand for slanted and low-credibility publishers
3. Engagement maximization slightly increases the share of negative sentiment comments
   - Largest increase for users preferring left-leaning and low-credibility publishers

▸ Diversity ▸ Market Shares ▸ Conclusion

Additional findings

1. Personalized engagement maximization reduces engagement diversity
   – Driven by users with preferences for extreme publishers
2. Personalized engagement maximization increases demand for slanted and low-credibility publishers
3. Engagement maximization slightly increases the share of negative sentiment comments
   – Largest increase for users preferring left-leaning and low-credibility publishers

Diversity Market Shares Conclusion

Outline

Recap

Objective: Understand **trade-offs** faced when **optimizing for engagement**

Approach: Estimate a discrete choice model of engagement for counterfactual analysis

Results and implications

- Engagement maximization **exacerbates differences** in credibility of news diets
  - Increases market share of low credibility and politically slanted outlets
- Managers can **improve the credibility** of promoted content for **modest cost**
- Personalization opt-outs are an inefficient solution to the problem
  - A better approach is to think about how to align the platform's ranking objective

Recap

Objective: Understand **trade-offs** faced when **optimizing for engagement**

Approach: Estimate a discrete choice model of engagement for counterfactual analysis

Results and implications

- Engagement maximization **exacerbates differences** in credibility of news diets
  - Increases market share of low credibility and politically slanted outlets
- Managers can **improve the credibility** of promoted content for **modest cost**
- Personalization opt-outs are an inefficient solution to the problem
  - A better approach is to think about how to align the platform's ranking objective

Recap

Objective: Understand **trade-offs** faced when **optimizing for engagement**

Approach: Estimate a discrete choice model of engagement for counterfactual analysis

Results and implications

- Engagement maximization **exacerbates differences** in credibility of news diets
  - Increases market share of low credibility and politically slanted outlets
- Managers can **improve the credibility** of promoted content for **modest cost**
- Personalization opt-outs are an inefficient solution to the problem
  - A better approach is to think about how to align the platform's ranking objective

Recap

Objective: Understand **trade-offs** faced when **optimizing for engagement**

Approach: Estimate a discrete choice model of engagement for counterfactual analysis

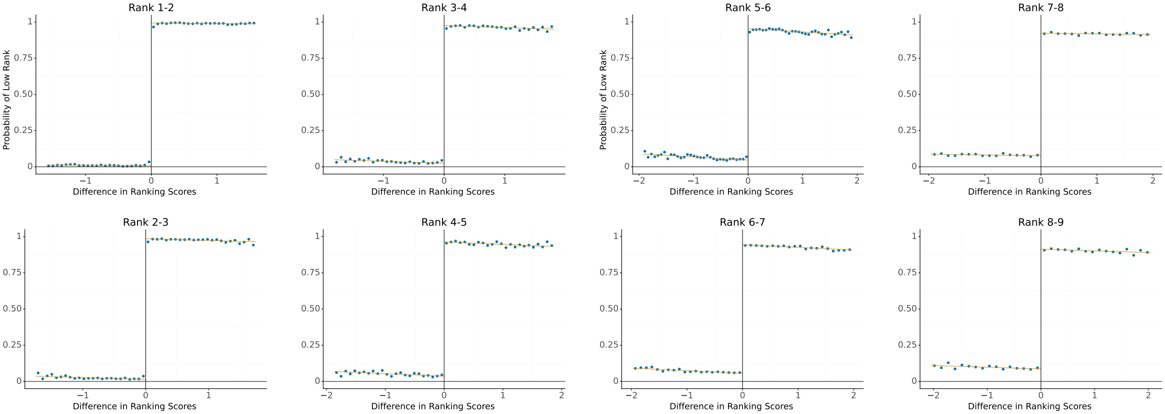
Results and implications

- Engagement maximization **exacerbates differences** in credibility of news diets
  - Increases market share of low credibility and politically slanted outlets
- Managers can **improve the credibility** of promoted content for **modest cost**
- Personalization opt-outs are an inefficient solution to the problem
  - A better approach is to think about how to align the platform's ranking objective

PURDUE UNIVERSITY

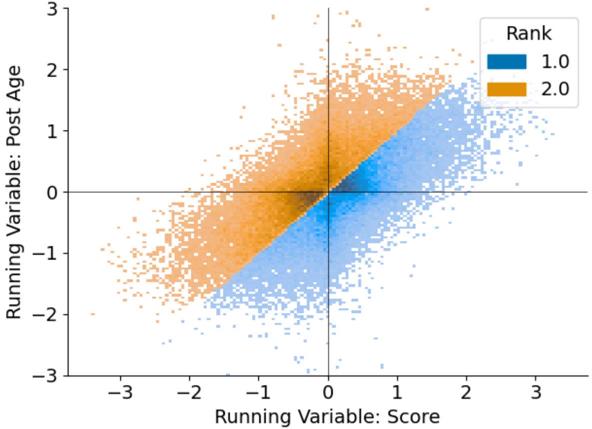
Outline

Regression discontinuity first stage



Back

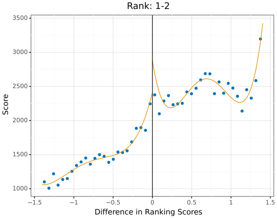
Ranking score joint distribution

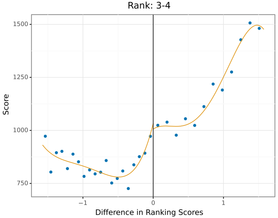


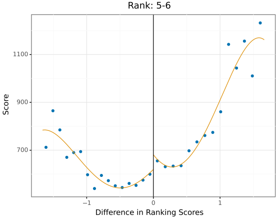
▸ Back

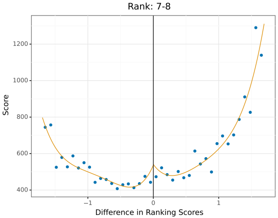
PURDUE UNIVERSITY

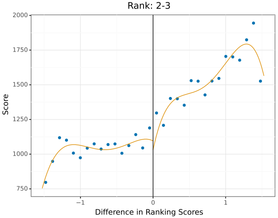
Balance: Score

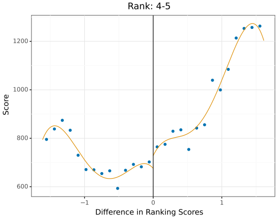


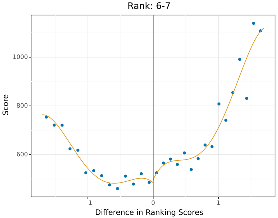


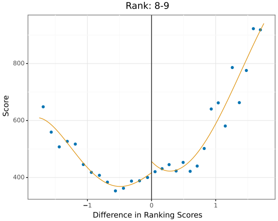




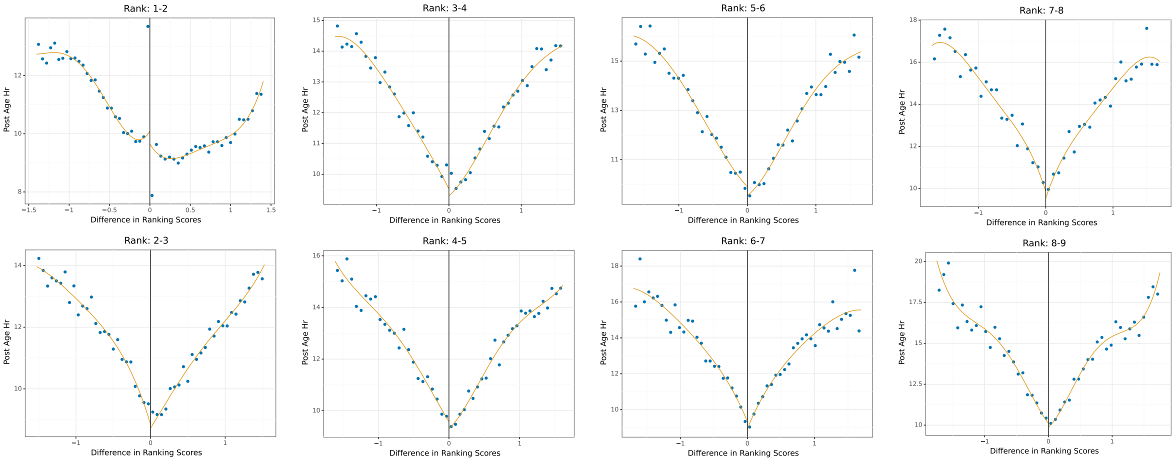








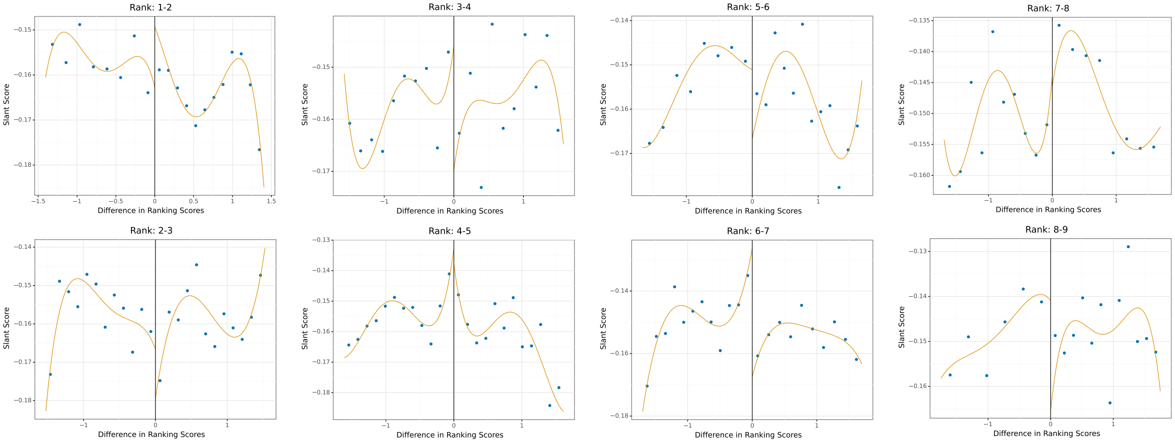
Balance: Post Age



Back

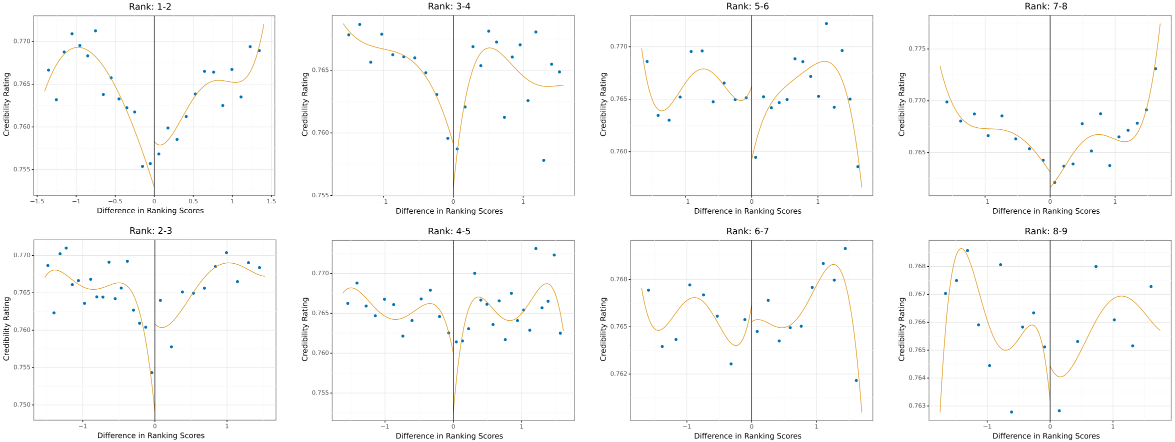
PURDUE
UNIVERSITY

Balance: Slant score



PURDUE
UNIVERSITY

Balance: Credibility rating



PURDUE UNIVERSITY

Balance tests

Figure: Vote score

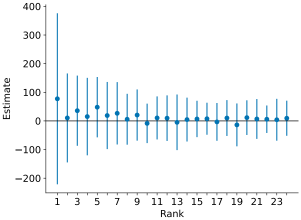


Figure: Slant score

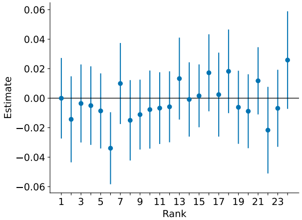


Figure: Post age (hours)

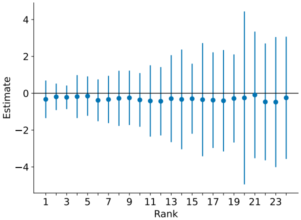
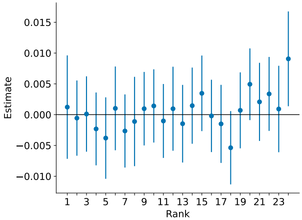
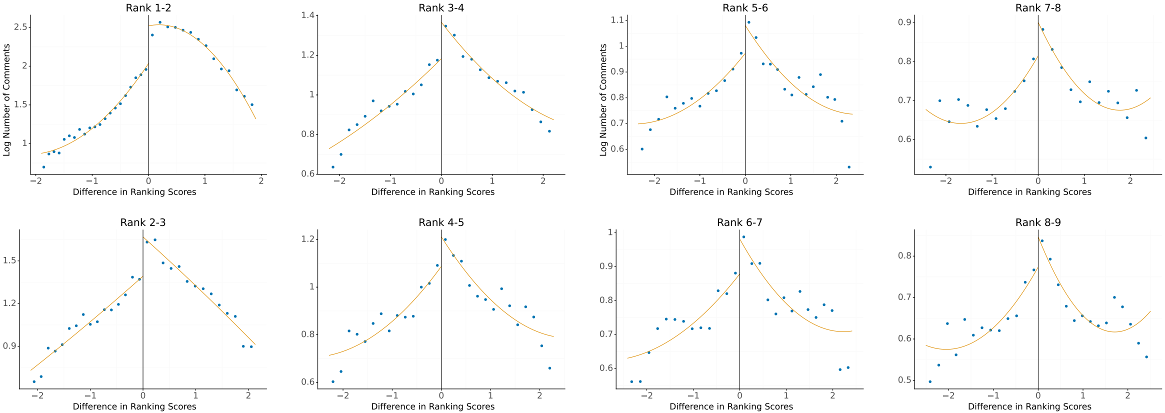


Figure: Credibility rating



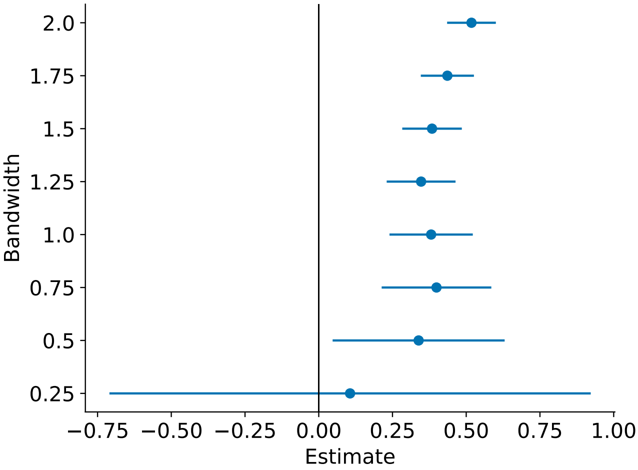
Back

RD Outcome Plots



Back

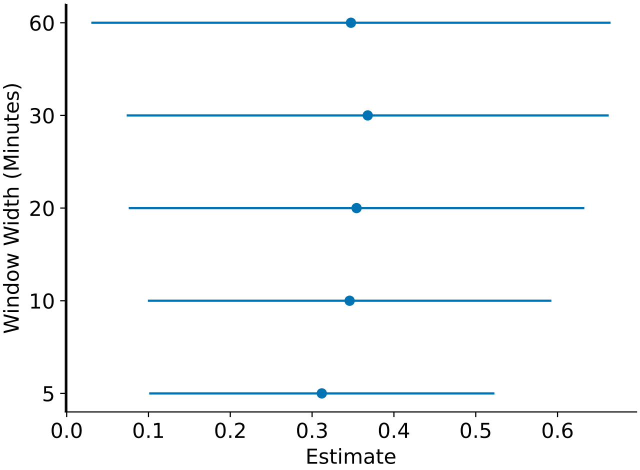
Robustness to bandwidth



Back

PURDUE UNIVERSITY

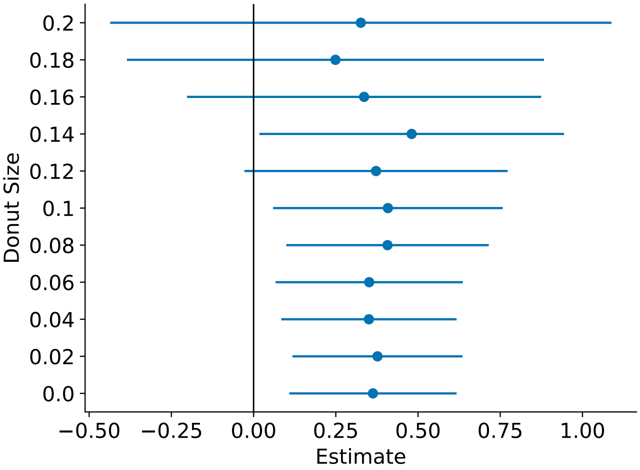
Robustness to window length



Back

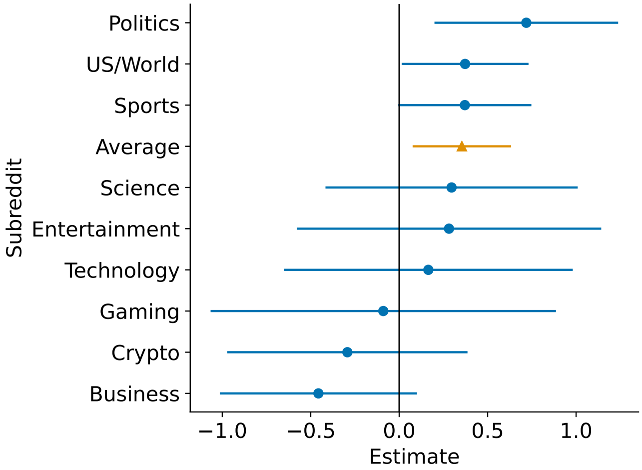
PURDUE UNIVERSITY

Robustness to donut width



Back

Main effects by category



Back

Identification

- Exposure parameters: $p(r,t) = p_r p_t$
  - p_t: Calibrated to the share of active users who visit the platform in each period
  - p_r: Identified by the reduced form treatment effects

$$\tau_r = \log \frac{E\left[d_{ijt}(r)\right]}{E\left[d_{ijt}(r+1)\right]} = \log \frac{p_r}{p_{r+1}}$$

- Preference parameters:
  - Individual preferences identified by repeated decisions of each individual
  - Key assumptions: article features mean independent of $\xi_{jt}, \varepsilon_{ijt}$
  - Assume market shares observed without sampling error

Back

PURDUE UNIVERSITY

Empirical Bayes shrinkage

Challenges:

- Each $\hat{\beta}_i$ estimate is imprecise

Solution:

- Empirical Bayes shrinkage

$$\hat{\beta}_i | \beta_i, \Sigma_i \sim N\left(\beta_i, \Sigma_i\right)$$

$$\beta_i \sim N\left(\mu, \Sigma\right)$$

$$\hat{\mu} = E_n\left[\hat{\beta}_i\right]$$

$$\hat{\Sigma} = E_n\left[\left(\hat{\beta}_i - \hat{\mu}\right)\left(\hat{\beta}_i - \hat{\mu}\right)'\right] - E_n\left[\Sigma_i\right]$$

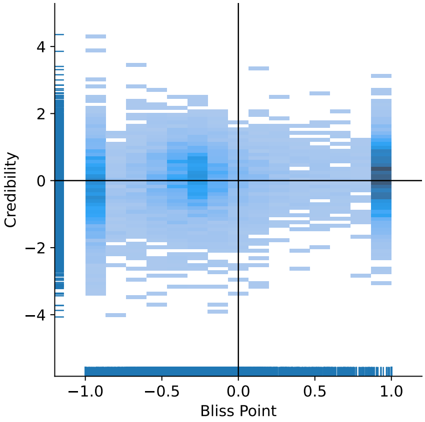
Individual preference estimates

Table: Individual preference estimates

	Mean	Std	1%	25%	50%	75%	99%
Constant	-4.55	0.75	-6.09	-5.06	-4.62	-4.11	-2.49
Slant Score	-0.20	0.64	-1.60	-0.65	-0.20	0.23	1.27
Slant Score^2	-0.30	0.82	-2.10	-0.88	-0.30	0.27	1.59
Credibility Rating	-0.02	0.86	-2.30	-0.53	0.03	0.56	1.86
ξ_{jt}	-0.00	0.93	-2.22	-0.61	0.02	0.61	2.12

Back

Joint distribution of preferences



Back

PURDUE UNIVERSITY

Model fit

Table: Summary of model fit

		Actual		Model	
		Mean	Std	Mean	Std
Total		52.39	38.85	54.05	35.31
Credibility	High	41.55	30.73	42.72	27.88
	Low	10.84	9.62	11.33	8.26
Slant Partition	Strongly Left	13.38	11.86	14.04	10.42
	Left	7.99	6.75	8.14	5.31
	Middle	13.44	10.60	13.63	8.88
	Right	13.63	10.78	13.93	9.15
	Strongly Right	3.95	3.87	4.31	3.32

Summary of model fit

Figure: Total engagement

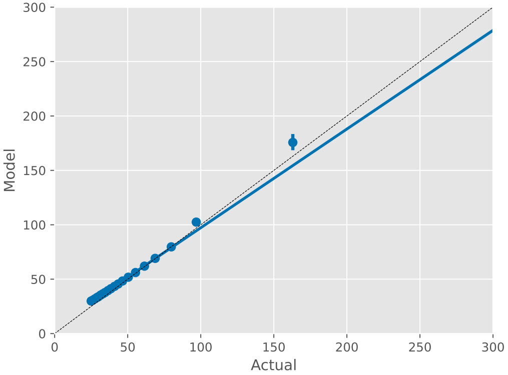
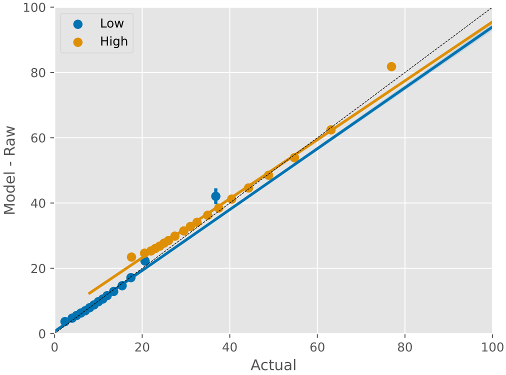
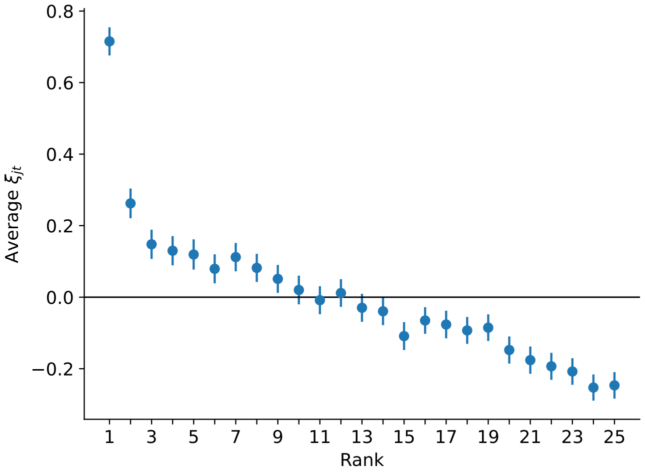


Figure: Engagement by credibility



▸ Back

Correlation of ξ_{jt} and r_{jt}



Analyzing comment text

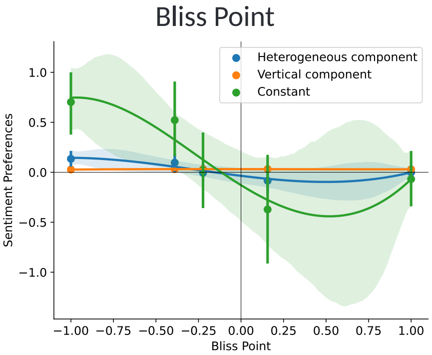
- Calculate sentiment, emotion, and toxicity of text using pre-trained neural network [Pérez et al., 2021]
- Users choose the probability their comment will be perceived negatively

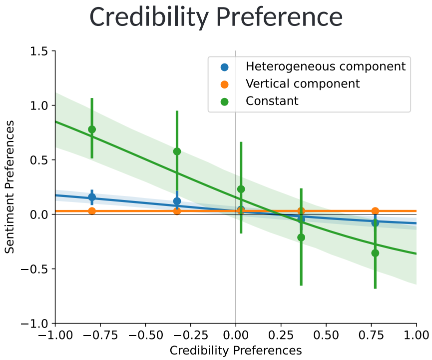
$$\log \frac{b_{ijt}}{1 - b_{ijt}} = \beta^s_{i0} + \beta^s_{i1}(\delta_{ijt} - \xi_{jt}) + \beta^s_{i2}\xi_{jt} + \varepsilon^s_{ijt}$$

- b_{ijt}: probability comment is negative
- δ_{ijt}: User-specific utility component ($x'_{jt}\beta_i$)

Back

Substantial heterogeneity in sentiment preferences





Back

Engagement maximizing algorithm

- Given preferences, platform wants to maximize in each period

$$\arg\max_{r\in\mathcal{R}} \sum_j E[d_{ij}] = \arg\max_{r\in\mathcal{R}} \sum_j p_{r_j} E\left[\frac{\exp \delta_{ij}}{1+\exp \delta_{ij}}\right]$$

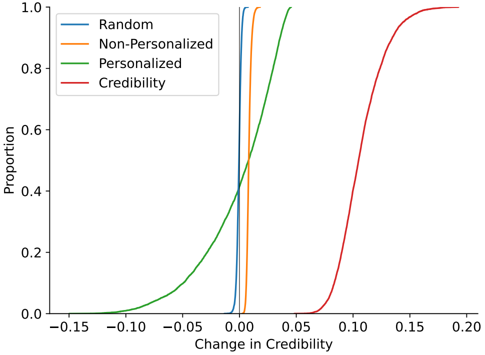
- Optimal allocation ranks posts in descending order of $E\left[\frac{\exp \delta_{ij}}{1+\exp \delta_{ij}}\right]$
  - Assume $p\left(r,t\right)$ weakly decreasing in r

- Applies to personalized: platform ranks in descending order of δ_{ij}

Back

Impact on publisher credibility

Figure: Distribution of credibility change



Back

PURDUE UNIVERSITY

Counterfactual summaries

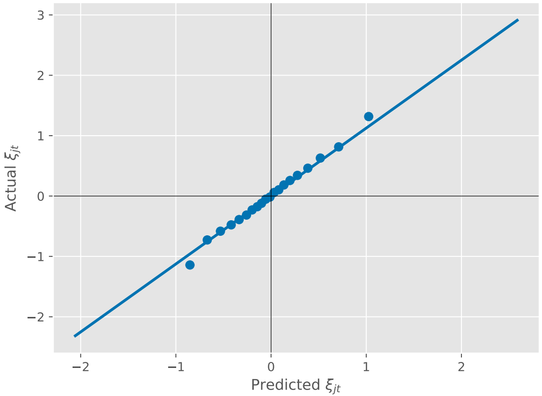
Table: Summary of counterfactual engagement

	Engagement	Dist. to Uniform	Credibility	Neg. Engagement
Intercept	54.048	0.280	0.790	0.512
	(0.387)	(0.001)	(0.001)	(0.002)
Random	-6.468	-0.004	-0.001	-0.001
	(0.043)	(0.000)	(0.000)	(0.000)
Non-Personalized	10.309	-0.008	0.008	0.001
	(0.062)	(0.000)	(0.000)	(0.000)
Personalized	11.357	0.030	-0.001	0.002
	(0.072)	(0.001)	(0.000)	(0.000)
Credibility Max.	8.118	0.008	0.107	0.001
	(0.054)	(0.000)	(0.000)	(0.000)
Observations	41675	41675	41675	41675
R-Squared	0.031	0.031	0.405	0.000

Back

Predicting ξ_{jt}

- Platform must predict ξ_{jt} given article features
- Train a random forest to predict ξ_{jt} given:
  - Vote score
  - Post age
  - Publisher slant & credibility rating
  - Number of existing comments
- Able to predict ξ, unexplained variation persists
  - Model achieves an R^2 of 0.42

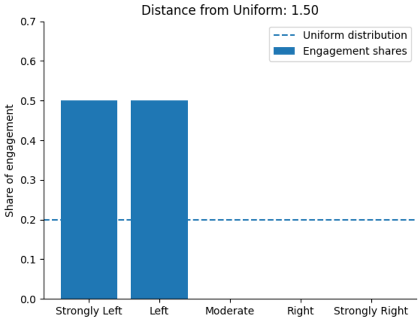


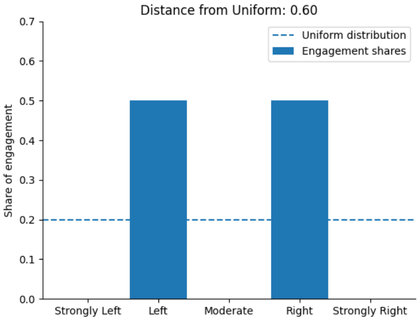
Back

Measuring diversity

How should we measure engagement diversity?

- Split publishers into five partitions of political slant
- Literature uses measures of industry concentration [Claussen et al., 2021, Holtz et al., 2020]
- Calculate first Wasserstein's distance between engagement distribution and uniform

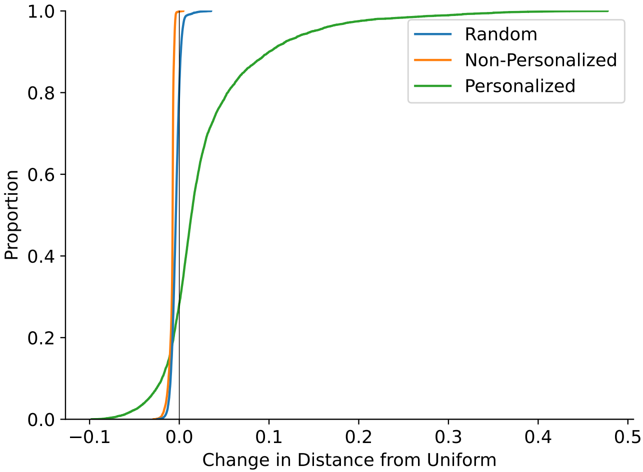




Back

PURDUE UNIVERSITY

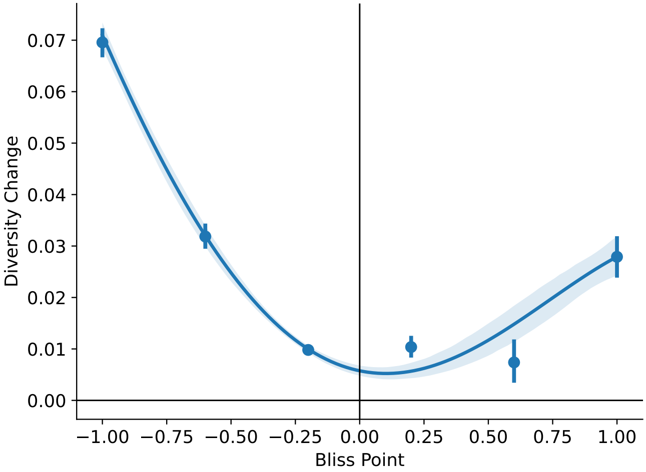
Personalization reduces engagement diversity



Back

PURDUE UNIVERSITY

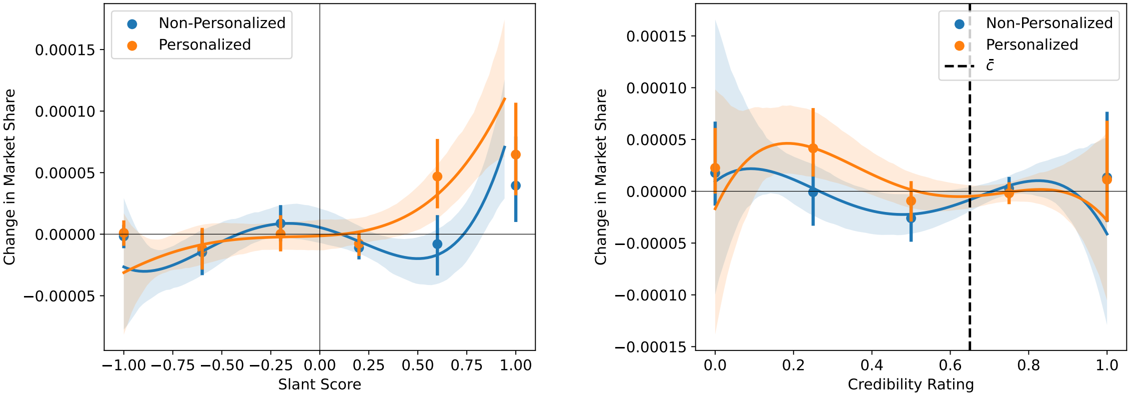
Decrease in diversity concentrated at extremes



Back

PURDUE UNIVERSITY

Impact on publishers



Back