

Equitable Persuasion in Incentivized Deliberation

An impossible tradeoff?

Emaad Manzoor

George H. Chen

Dokyun Lee

Michael D. Smith

**Carnegie
Mellon
University**

deliberation (noun) / di-,li-bə-'rā-shən

extended conversation among two
or more people to come to a better
understanding of some issue

(Beauchamp, 2020)

Deliberation Online




WIKIPEDIA
The Free Encyclopedia

[Article](#)[Talk](#)[Read](#)[Edit](#)[New section](#)[View history](#)

Talk:Coronavirus disease 2019

Discretionary sanctions on the use of preprints [\[edit \]](#)

I am appalled by the use of preprints to support content in this article. The website [MedRxiv](#)  displays a clear disclaimer:


Caution: Preprints are preliminary reports of work that have not been certified by peer review. They should not be relied on to guide clinical practice or health-related behavior and should not be reported in news media as established information.

I'm giving notice that tomorrow I intend to place a [general sanction](#) on the page to prohibit the use of preprints as sources in this article. This ought to be simply a matter of respecting our guidelines on [WP:Reliable sources](#) and [WP:MEDRS](#), but it now seems necessary. I'm naturally willing to hear reasons why discretionary sanctions should not be necessary to enforce our basic sourcing guidelines. --[RexxS](#) [\(talk\)](#) 21:51, 11 May 2020 (UTC)

- **Support**, obviously. [Boing! said Zebedee](#) [\(talk\)](#) 22:08, 11 May 2020 (UTC)
- **Support**. We should not be using preprints EVER. [MartinezMD](#) [\(talk\)](#) 22:53, 11 May 2020 (UTC)
- This is a [WP:point](#), against [WP:5P5](#), [WP:5P4](#), [WP:5P3](#) and potentially [WP:5P2](#).
This is an article about a current event. Our main source in the contested chapter (IFR) say I quote loosely : "Since yesterday [...] one research group has provided a correction of their estimate of the Infection-Fatality Ratio (IFR)". Since yesterday... Is that the pinnacle of peer review we strive for ? We have to deal with research that change daily, there is no need to put the big administrator boots and add yet another banner on top of this page. Just to state the obvious that peer reviewed source would be preferable. Everyone here agree. [lluvalar](#) [\(talk\)](#) 22:56, 11 May 2020 (UTC)
- **Oppose**. What we're up against are bat shit crazy conspiracy theories. That's the reality of the situation. We're also at risk of irrelevancy due to the 24-hour news cycle and social media.

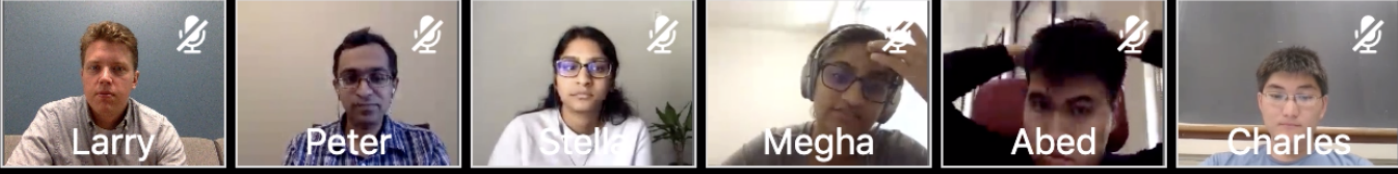
Deliberation Online

Electoral Reform



Current speaker
Akash

Next speakers
Stella
Larry



Cancel Request to Speak

You are in the queue.

Conclude Discussion

[Report a User](#)

Help Log Out

Agenda

- Previous agenda items
- Substitute the national popular vote for the Electoral College through a constitutional amendment**
[View details](#)
 - Pro** Every vote will count in whatever state it is cast.
 - Pro** Instead of the campaign focusing on a few critical swing states we will have a truly national election.
 - Con** The constitution is hard to change. There are more urgent priorities.
 - Con** The national popular vote will be hard to count with enough accuracy if there is a close election. We could have long recounts on a national basis.
 - Con** Without the electoral college, small towns and rural areas would be ignored and candidates would only campaign where there are big concentrations of voters.
- Questions


Stanford Online Deliberation Platform

cdd.stanford.edu

Deliberation Online

Project roadmap? #254

 **Open** aatkinson opened this issue on Jan 29, 2019 · 8 comments

 **aatkinson** commented on Jan 29, 2019



Hi,


The goals are ambitious, the codebase is in flux, and future directions are outlined in your paper <https://arxiv.org/pdf/1812.08729.pdf>, but it's unclear what the roadmap is for this project.

What release cadence can we expect, and what features are prioritized?

I'm also curious if this has a substantial backing by Facebook / the PyTorch team.

Thanks

 3 


 **ahhegazy** commented on Jan 31, 2019 • edited Contributor


Great questions :)

As for the road map, we are mainly working on 5 main areas (roughly sorted according to priority):

- Enriching and Improving our stack with more tasks, models and techniques
- Improve Usability and have better integration with ipython notebooks
- Performance optimizations for training and inference
- Explore model interpretability techniques


For the release cadence, we will do our best to have a release every month. And yeah as we mentioned in our release blog post [here](#) this project powers some of the core projects in production at Facebook and it indeed has a substantial backing from different teams in Facebook AI.




 **ibulu** commented on Feb 1, 2019



Thanks for the answer @ahhegazy! Does integration with fairseq fall under the 1st area?


I am curious about the fairest integration as well. Any plans on that?





 **ahhegazy** commented on Feb 1, 2019 Contributor

Yes, we are working on integrating with translate: https://github.com/pytorch/translate/tree/master/pytorch_translate which is Fairseq models with production support

 3 

 **padipadou** commented on Feb 6, 2019

Hello @ahhegazy, and thanks for sharing this amazing work, just a quick question, can we have an idea about timing regarding fairseq integration ? Thanks a lot !

 3 

github
SOCIAL CODING



Verified



Verified



Verified



Verified



Verified

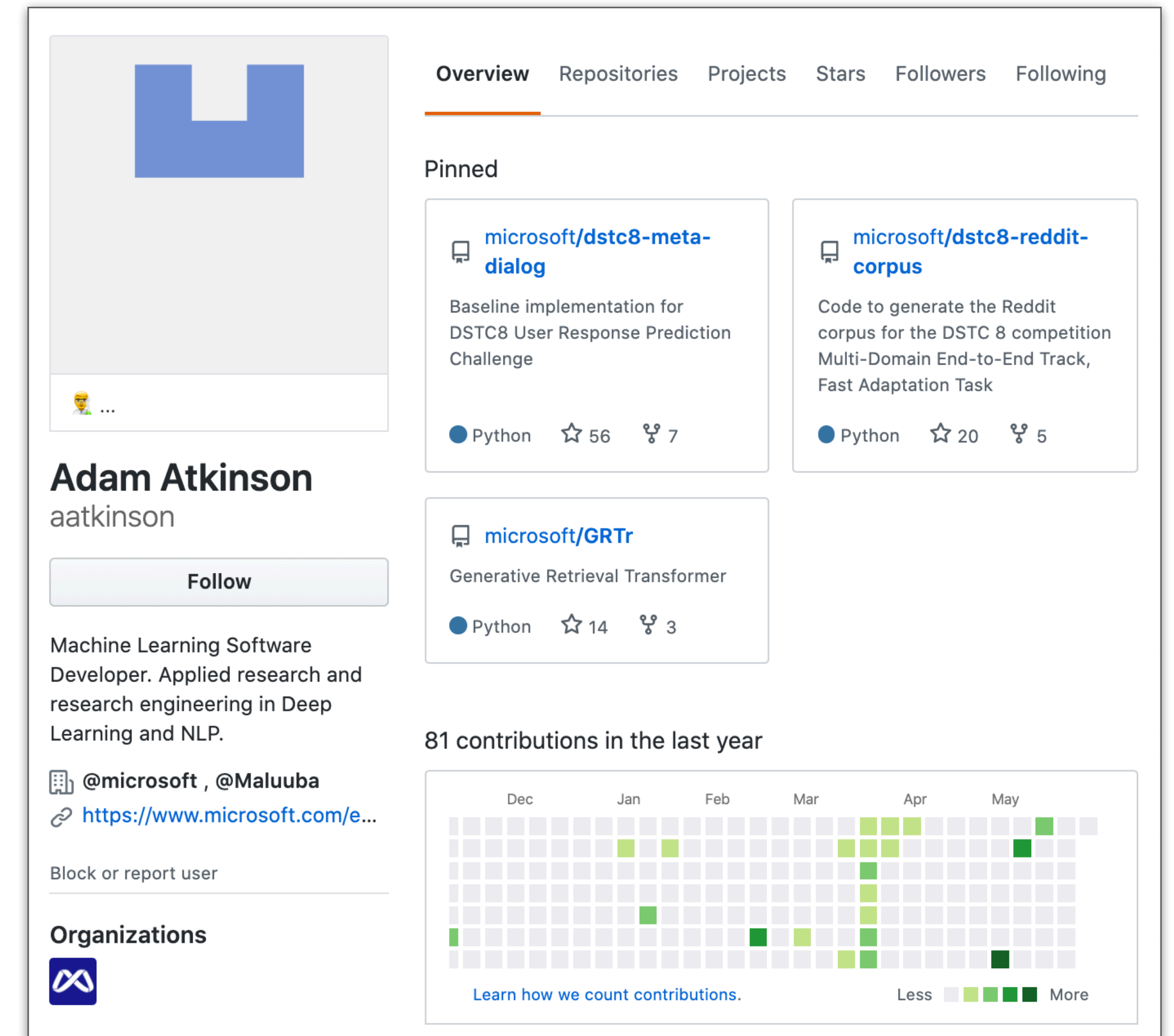
Reputation Indicators

The image displays two GitHub profile pages side-by-side, illustrating reputation indicators. The left profile is for Adam Atkinson (@aatkinson), a Machine Learning Software Developer at Microsoft. His pinned repositories include `microsoft/dstc8-meta-dialog` (Python, 56 stars, 7 forks) and `microsoft/dstc8-corpus` (Python, 20 stars). He has 81 contributions in the last year, shown in a heatmap. The right profile is for padipadou, who has pinned repositories like `higgs_challenge` (Python) and `rakuten_2018` (Jupyter Notebook). Their heatmap shows 0 contributions in the last year. Both profiles have a 'Follow' button and a 'Block or report user' link.

Used by project maintainers to prioritize issues and evaluate new contributors (Marlow et al, 2013)

Reputation Indicators

- + Incentivize engagement
- Distort persuasive equity?



Q. Does reputation
have persuasive power
in deliberation online?

Preview of Findings

Reputation is
persuasive

+10 reputation units →
+26% persuasion rate

Patterns in effect heterogeneity
consistent with **reference cues theory**
(Bilancini & Boncinelli, 2018)

Empirical Strategy

- I. Identifying opinion-change
- II. Disentangling non-reputation factors
- III. Handling unobserved confounders
- IV. Controlling for text

Empirical Strategy

- I. Identifying opinion-change
- II. Disentangling non-reputation factors
- III. Handling unobserved confounders
- IV. Controlling for text

I. Identifying Opinion-Change

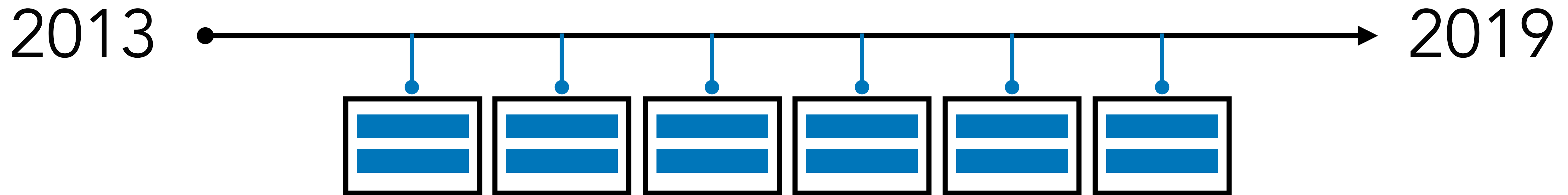
Persuasion: Empirical Evidence.

DellaVigna & Gentzkow. *Annual Review of Economics*. 2010.

Typically **unobserved** —
challenging to identify

I. Identifying Opinion-Change

Our strategy: Dataset of online deliberation from **ChangeMyView**



>1 million debates between >800,000 members
>20 moderators enforce high-quality deliberation

Poster

Posted by u/togtogtog 4Δ 12 hours ago

CMV: Most of us think we live environmentally responsible lives, but most of us don't.

Deltas(s) from OP

Each of us may have things that we do to be environmentally responsible. We may not use plastic straws, or be vegan, or cycle to work.

However, while we are happy to do the things that are fairly easy, we are reluctant to do the harder things: to have less children, or not fly, or not have a car, or not have a smartphone.

In our own heads, we think we are environmentally responsible because we recycle, or buy organic vegetables, or because we use a reusable cup (otherwise known as a cup).

But we ignore the ways in which we are not environmentally responsible, and blame it on the way society is structured, or on politicians, or as being impractical.

Challenger

Reputation

miguelguajiro 110Δ Score hidden · 12 hours ago

By responsible, do you mean sustainable? And how do you conclude that most people believe their lives on the whole are environmentally sustainable? Could it be that people make the easy responsible choices while also aware that their lives as a whole aren't sustainable?

Reply Give Award Share Report Save

↑ togtogtog 4Δ Score hidden · 11 hours ago

↓ Now that is a good point. Maybe people simply don't think they are living sustainable lives and also, many people simply don't think about it one way or the other.

I guess I meant that those of us who *do* think we are living in an environmentally friendly way simply are NOT living sustainably by any means. But I wasn't very clear in how I expressed this.

Δ

●

└─

Indicator of successful persuasion

Explicit indicators of successful persuasion
provided by opinion-holders (posters)

Poster

Posted by u/togtogtog 4Δ 12 hours ago

CMV: Most of us think we live environmentally responsible lives, but most of us don't.

Deltas(s) from OP

Each of us may have things that we do to be environmentally responsible. We may not use plastic straws, or be vegan, or cycle to work.

However, while we are happy to do the things that are fairly easy, we are reluctant to do the harder things: to have less children, or not fly, or not have a car, or not have a smartphone.

In our own heads, we think we are environmentally responsible because we recycle, or buy organic vegetables, or because we use a reusable cup (otherwise known as a cup).

But we ignore the ways in which we are not environmentally responsible, and blame it on the way society is structured, or on politicians, or as being impractical.

Challenger

Reputation

miguelguajiro 110Δ Score hidden · 12 hours ago

By responsible, do you mean sustainable? And how do you conclude that most people believe their lives on the whole are environmentally sustainable? Could it be that people make the easy responsible choices while also aware that their lives as a whole aren't sustainable?

Reply Give Award Share Report Save

↑ togtogtog 4Δ Score hidden · 11 hours ago

↓ Now that is a good point. Maybe people simply don't think they are living sustainable lives and also, many people simply don't think about it one way or the other.

I guess I meant that those of us who *do* think we are living in an environmentally friendly way simply are NOT living sustainably by any means. But I wasn't very clear in how I expressed this.

Δ

Indicator of successful persuasion

Prominent display of reputation based on number of individuals persuaded previously

Empirical Strategy

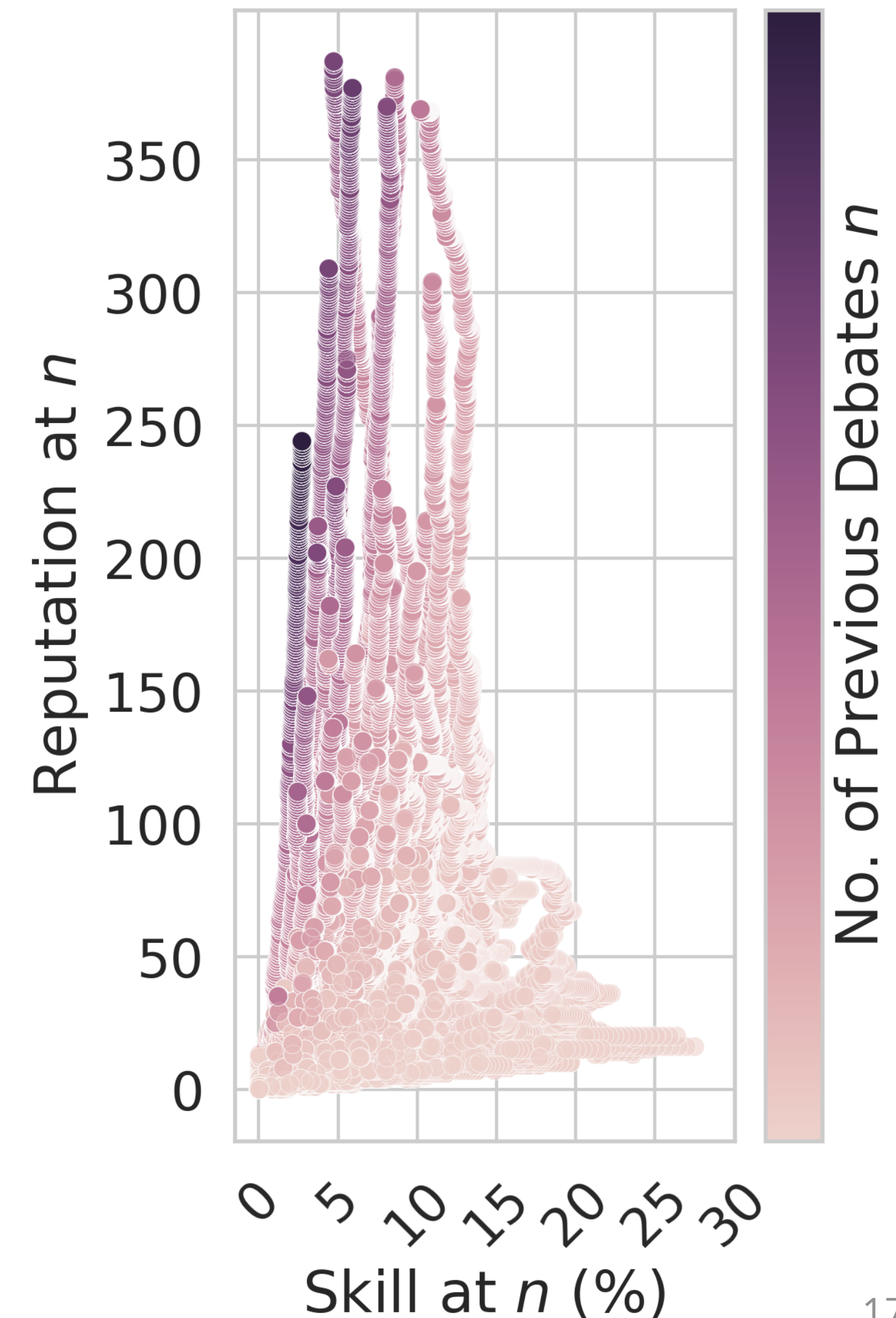
- I. Identifying opinion-change
- II. Disentangling non-reputation factors
- III. Handling unobserved confounders
- IV. Controlling for text

II. Disentangling Non-Reputation Factors

Exploit multiple debates per challenger

$$\text{skill} = \frac{\text{no. posters persuaded previously}}{\text{no. previous debates}}$$

Controls for **time-invariant** challenger characteristics that affect persuasion



II. Disentangling Non-Reputation Factors

Exploit **multiple responses per opinion** to control for **opinion fixed-effects**

Addresses confounding arising from **endogenous opinion selection**

Opinion



r_1



r_2



r_3



Each challenger's response → a debate

Empirical Strategy

- I. Identifying opinion-change
- II. Disentangling non-reputation factors
- III. Handling unobserved confounders
- IV. Controlling for text

Empirical Strategy

- I. Identifying opinion-change
- II. Disentangling non-reputation factors
- III. Handling unobserved confounders
- IV. Controlling for text

III. Handling Unobserved Confounders

Main concern

Time-varying challenger characteristics **correlated with persuasion**

Example: users improving their rhetorical ability with platform experience

III. Handling Unobserved Confounders

Instrument intuition

- **Higher** (worse) position \rightarrow **lower** persuasion probability
- **Reputation** \approx no. of posters persuaded previously

Opinion



*Decreasing
attention,
argument
space*

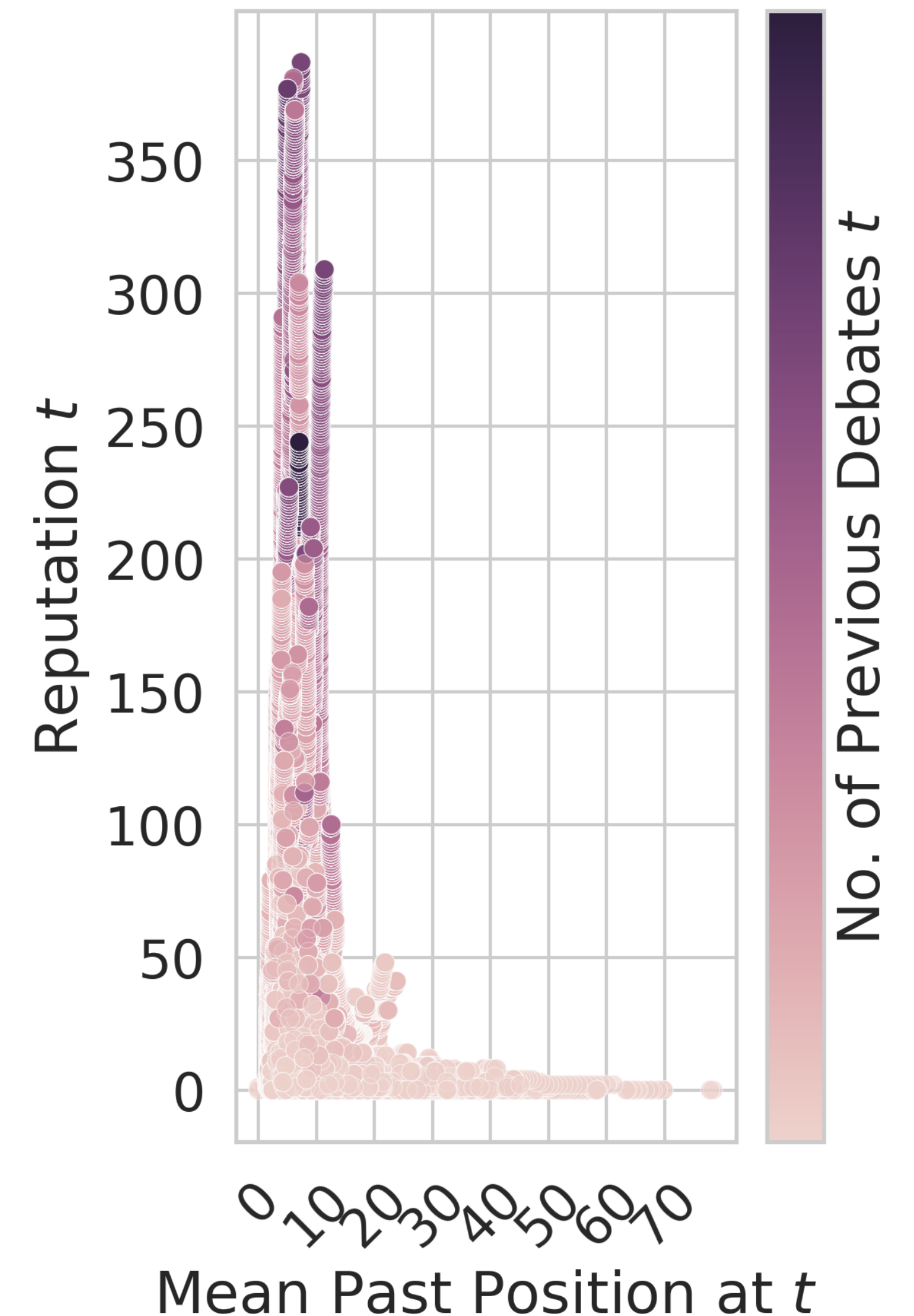
III. Handling Unobserved Confounders

Instrument definition

Mean past position of challenger before the present debate

First-stage F-statistic > 3000

Similar to the Fox News channel position instrument (Martin & Yurukoglu, 2017)



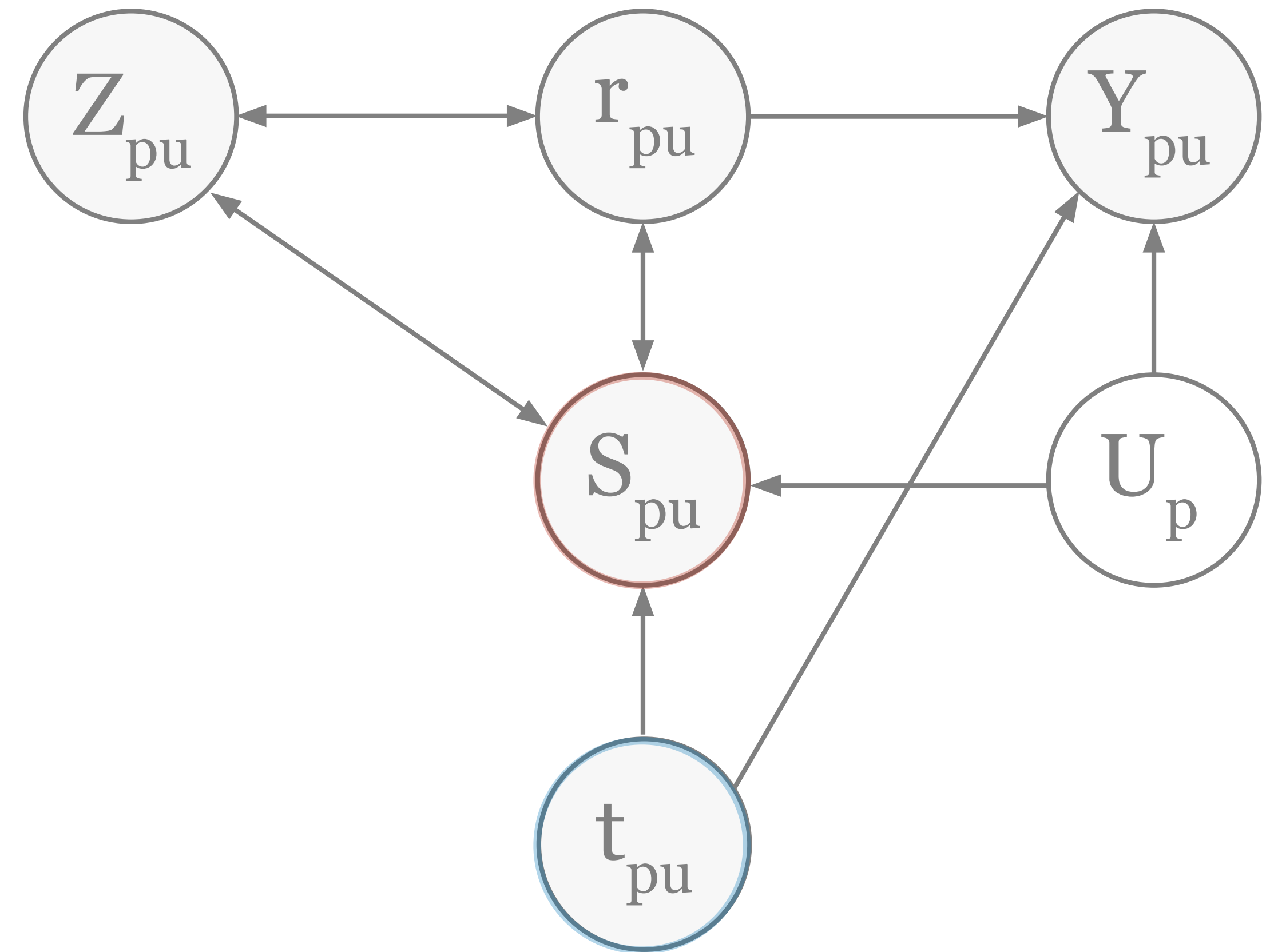
III. Handling Unobserved Confounders

Immediate concern

Users **selecting** opinions to challenge based on their **anticipated** response position



Must control for response position in the **present** debate



(see paper for details)

Empirical Strategy

- I. Identifying opinion-change
- II. Disentangling non-reputation factors
- III. Handling unobserved confounders
- IV. Controlling for text

Empirical Strategy

- I. Identifying opinion-change
- II. Disentangling non-reputation factors
- III. Handling unobserved confounders
- IV. Controlling for text

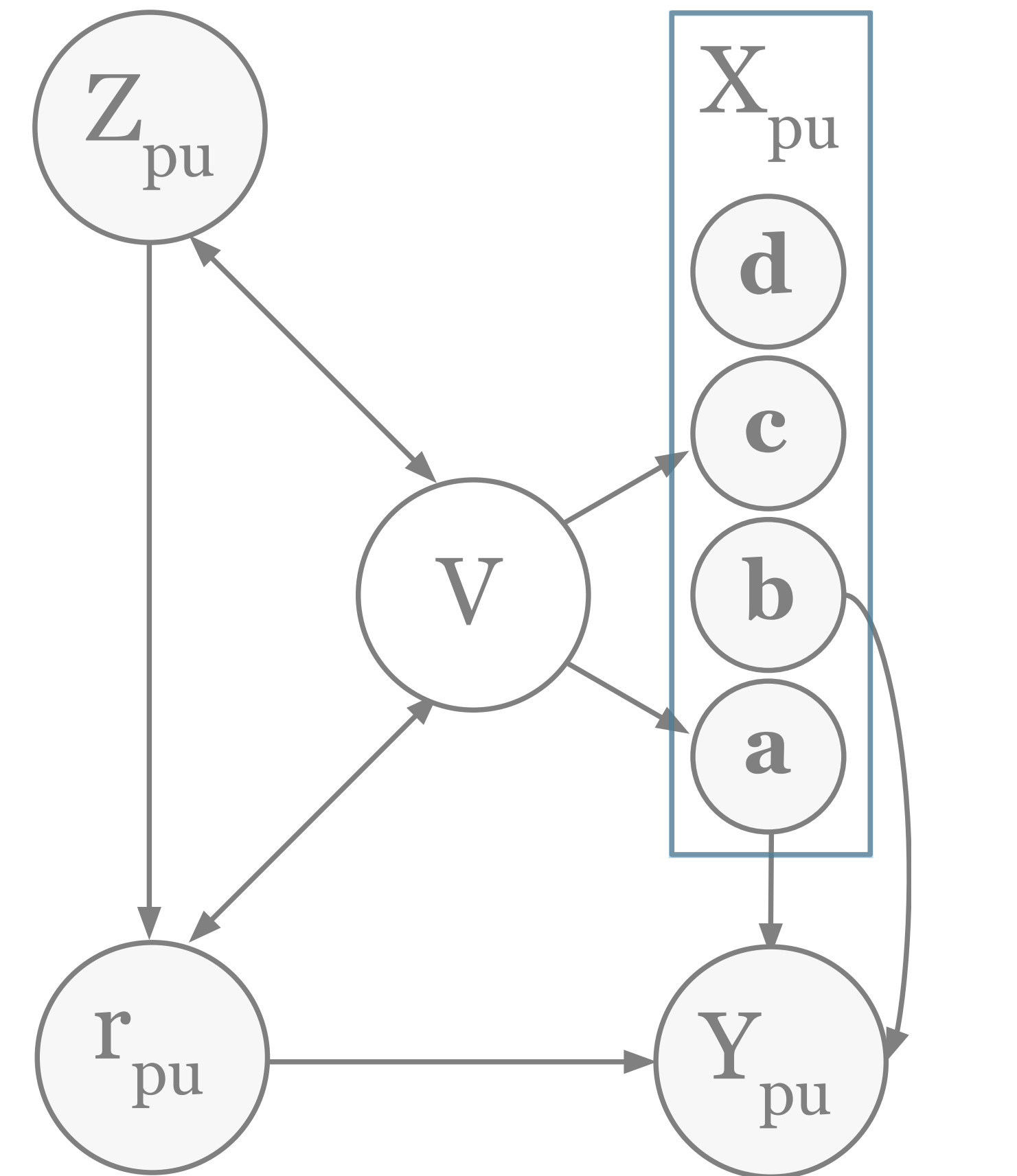
IV. Controlling for Text

Why control for text?

Instrument confounders must affect both instrument **and** outcome

Are likely to affect the outcome
through the response text

NLP approaches: No guarantees on retaining confounders or inference



(see paper for details)

IV. Controlling for Text

Our approach: Partially-linear IV model, estimated via double machine-learning (Chernozhukov et. al., 2016)

$$Y_{pu} = \beta_1 r_{pu} + \beta_2 s_{pu} + \beta_3 t_{pu} + g(\tau_p, X_{pu}) + \epsilon_{pu}$$

$$\mathbb{E}[\epsilon_{pu} | Z_{pu}, \tau_p, s_{pu}, t_{pu}, X_{pu}] = 0$$

$$Z_{pu} = \alpha_1 s_{pu} + \alpha_2 t_{pu} + h(\tau_p, X_{pu}) + \epsilon'_{pu}$$

$$\mathbb{E}[\epsilon'_{pu} | \tau_p, s_{pu}, t_{pu}, X_{pu}] = 0$$

IV. Controlling for Text

Our approach: Partially-linear IV model, estimated via double machine-learning (Chernozhukov et. al., 2016)

$$Y_{pu} = \beta_1 r_{pu} + \beta_2 s_{pu} + \beta_3 t_{pu} + g(\tau_p, X_{pu}) + \epsilon_{pu}$$

$$Z_{pu} = \alpha_1 s_{pu} + \alpha_2 t_{pu} + h(\tau_p, X_{pu}) + \epsilon'_{pu}$$

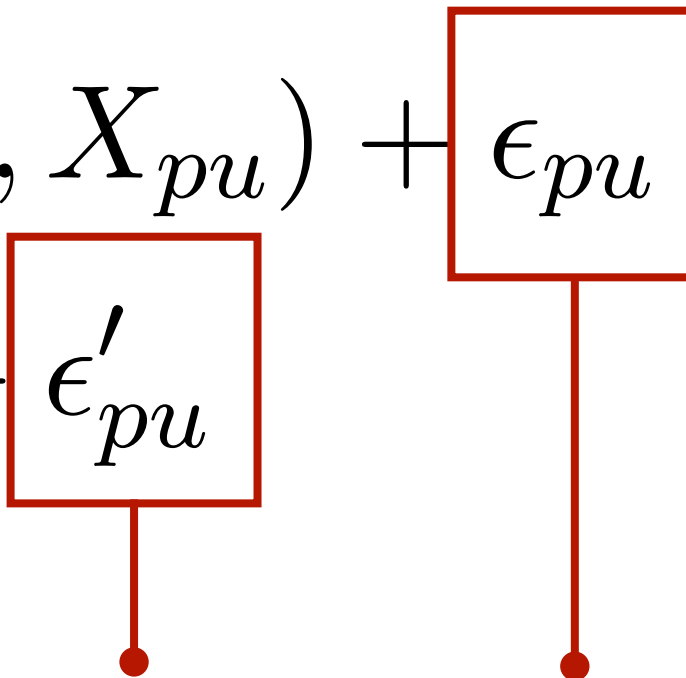
$$\mathbb{E}[\epsilon_{pu} | Z_{pu}, \tau_p, s_{pu}, t_{pu}, X_{pu}] = 0$$

$$\mathbb{E}[\epsilon'_{pu} | \tau_p, s_{pu}, t_{pu}, X_{pu}] = 0$$

Standard
instrumental
variable
assumptions

IV. Controlling for Text

Our approach: Partially-linear IV model, estimated via double machine-learning (Chernozhukov et. al., 2016)

$$Y_{pu} = \beta_1 r_{pu} + \beta_2 s_{pu} + \beta_3 t_{pu} + g(\tau_p, X_{pu}) + \epsilon_{pu}$$
$$Z_{pu} = \alpha_1 s_{pu} + \alpha_2 t_{pu} + h(\tau_p, X_{pu}) + \epsilon'_{pu}$$


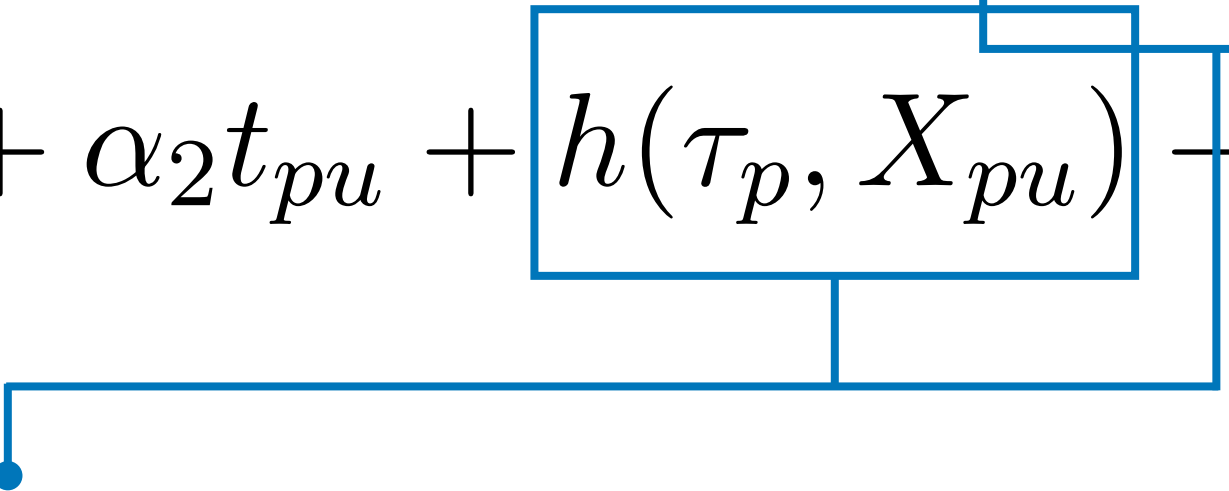
$$\mathbb{E}[\epsilon_{pu} | Z_{pu}, \tau_p, s_{pu}, t_{pu}, X_{pu}] = 0$$

$$\mathbb{E}[\epsilon'_{pu} | \tau_p, s_{pu}, t_{pu}, X_{pu}] = 0$$

**No distributional
assumptions** placed
on error terms (eg.
Gaussian, Gumbel)

IV. Controlling for Text

Our approach: Partially-linear IV model, estimated via double machine-learning (Chernozhukov et. al., 2016)

$$Y_{pu} = \beta_1 r_{pu} + \beta_2 s_{pu} + \beta_3 t_{pu} + g(\tau_p, X_{pu}) + \epsilon_{pu}$$
$$Z_{pu} = \alpha_1 s_{pu} + \alpha_2 t_{pu} + h(\tau_p, X_{pu}) + \epsilon'_{pu}$$


$$\mathbb{E}[\epsilon_{pu} | Z_{pu}, \tau_p, s_{pu}, t_{pu}, X_{pu}] = 0$$

$$\mathbb{E}[\epsilon'_{pu} | \tau_p, s_{pu}, t_{pu}, X_{pu}] = 0$$

Non-parametric nuisance functions of the opinion fixed-effects τ_p and text X_{pu}

Estimated via machine-learning

IV. Controlling for Text

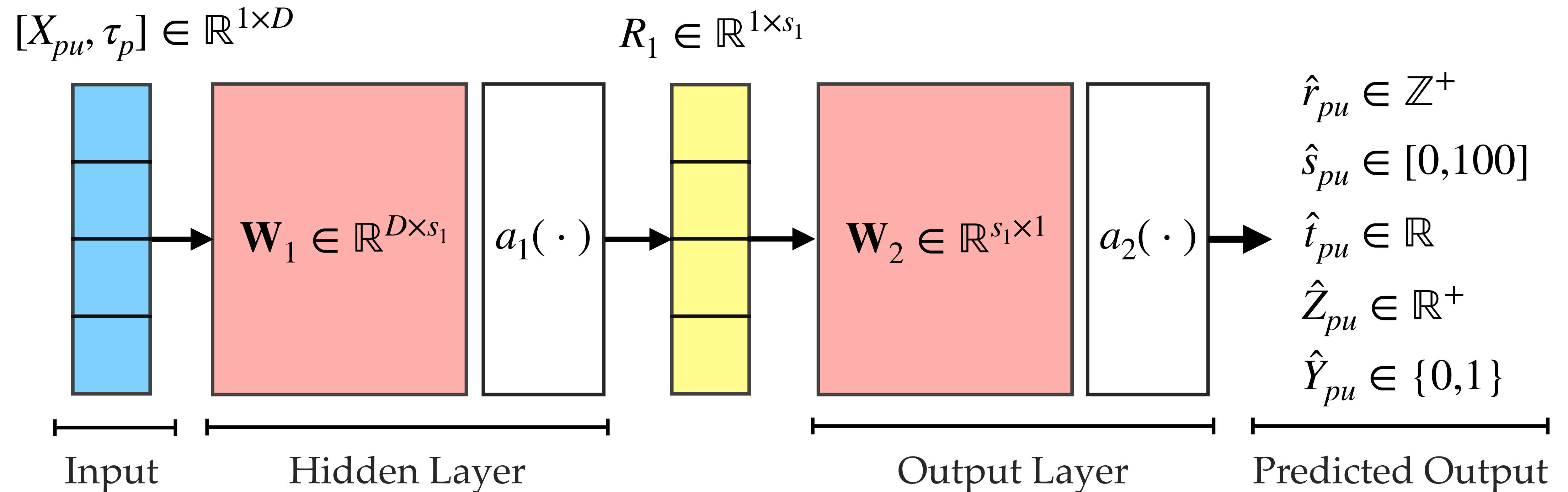
Our approach: Partially-linear IV model, estimated via double machine-learning (Chernozhukov et. al., 2016)

$$\begin{aligned} Y_{pu} &= \beta_1 r_{pu} + \beta_2 s_{pu} + \beta_3 t_{pu} + g(\tau_p, X_{pu}) + \epsilon_{pu} & \mathbb{E}[\epsilon_{pu} | Z_{pu}, \tau_p, s_{pu}, t_{pu}, X_{pu}] &= 0 \\ Z_{pu} &= \alpha_1 s_{pu} + \alpha_2 t_{pu} + h(\tau_p, X_{pu}) + \epsilon'_{pu} & \mathbb{E}[\epsilon'_{pu} | \tau_p, s_{pu}, t_{pu}, X_{pu}] &= 0 \end{aligned}$$

Consistent estimates, valid inference
if **product** of nuisance function
convergence rates is at least $n^{-1/2}$

IV. Controlling for Text

Nuisance functions: Deep ReLU neural networks



Valid inference with double ML (Farrell et. al., 2018)

Results

Reputation is **persuasive**

+10 reputation units → +26%

persuasion rate increase

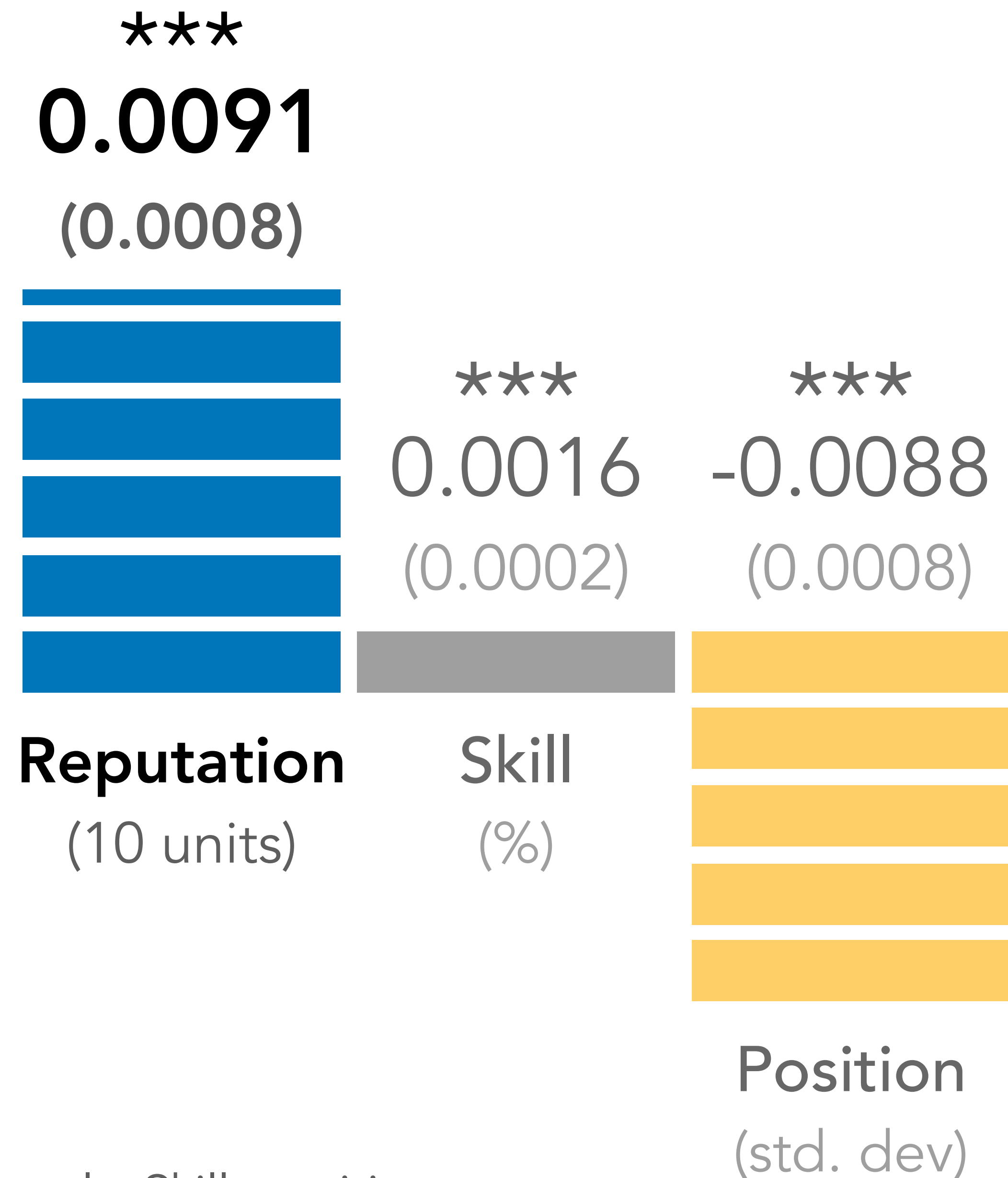
over the platform average

persuasion rate ($\approx 3.5\%$)

Estimated Local Average
Treatment Effect (LATE)

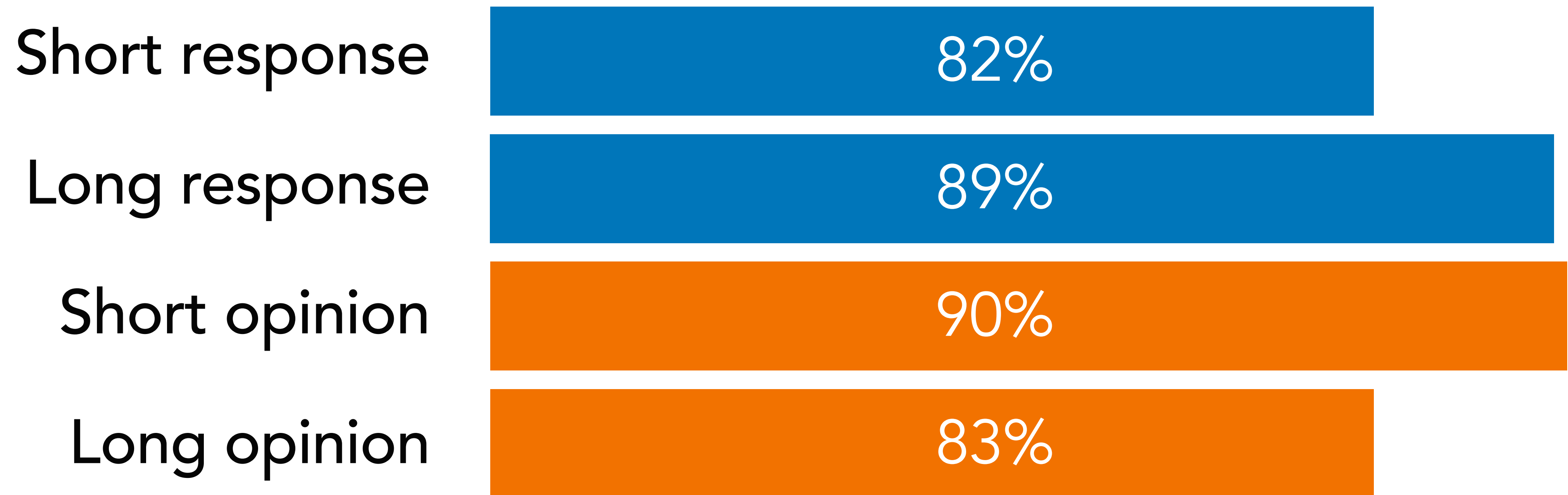
Outcome: Debate success
Treatment: Reputation

Controls: Skill, position, text
Includes opinion fixed-effects



Results

Persuasive power **increases** with **cognitive load** and **decreases** with **issue-involvement** of opinion-holder



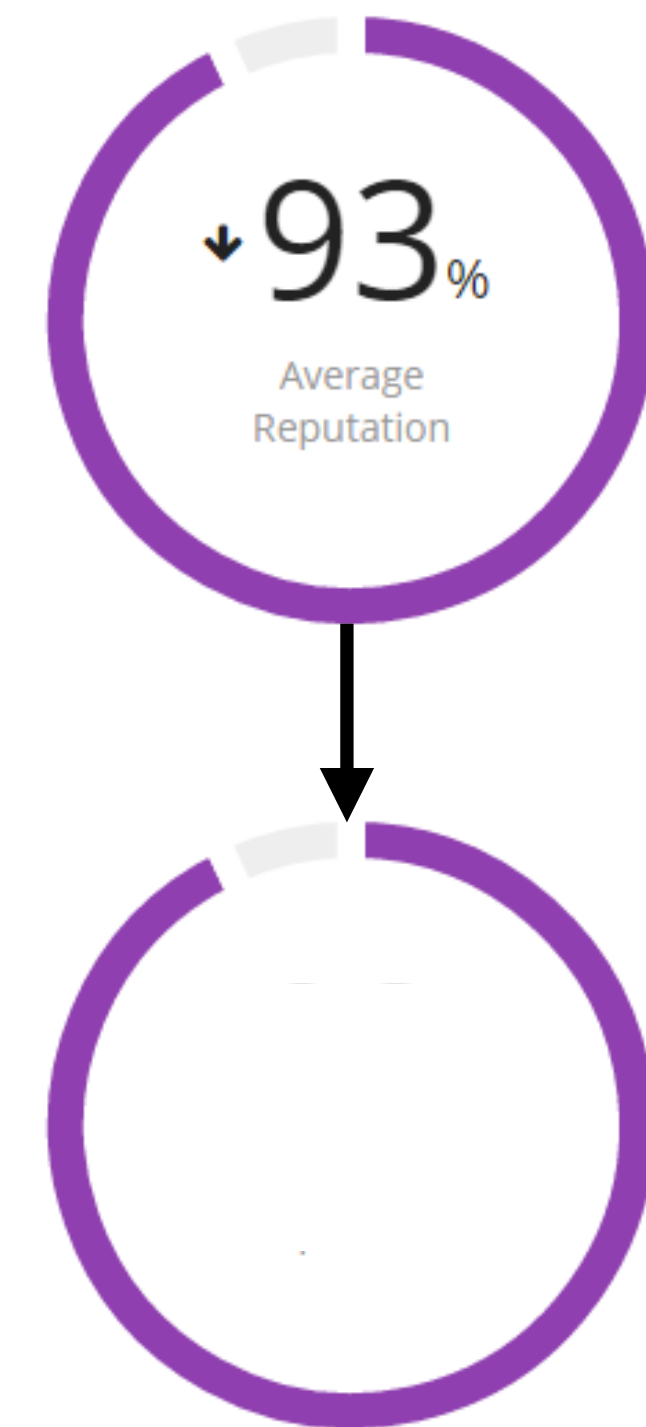
Reputation effect-share (vs skill)

Implications for Deliberation Platforms

Consistent with **reference cues** theory of persuasion (Bilancini & Boncinelli, 2018)

Reference cues used if they (i) have **lower cognitive cost**, and (ii) are **accurate proxies**

Potential strategy: Manipulate perceived reference cue accuracy



Preprint, code & data:
emaadmanzoor.com/ethos/

Emaad Manzoor
George H. Chen
Dokyun Lee
Michael D. Smith

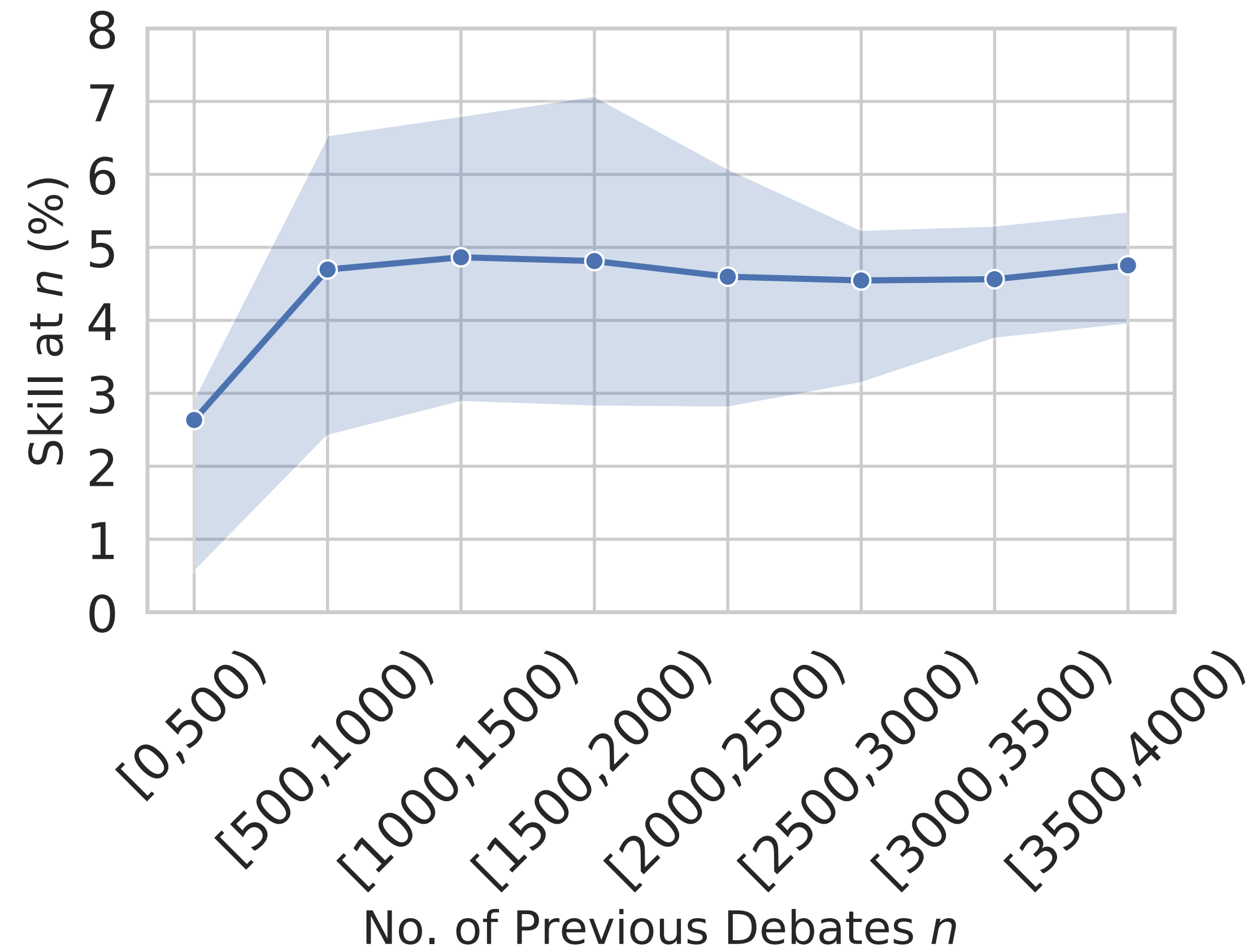
**Carnegie
Mellon
University**

Descriptive Statistics

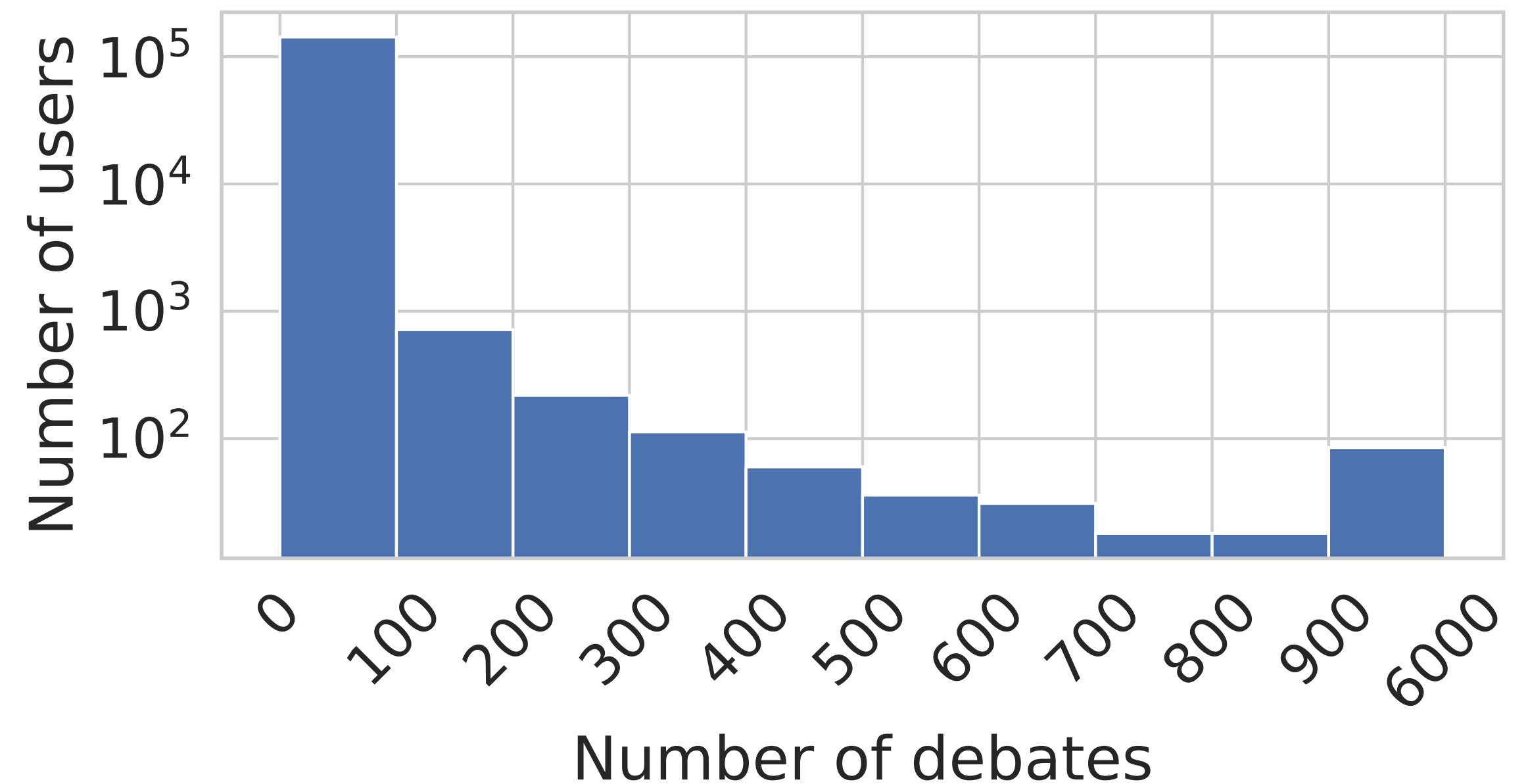
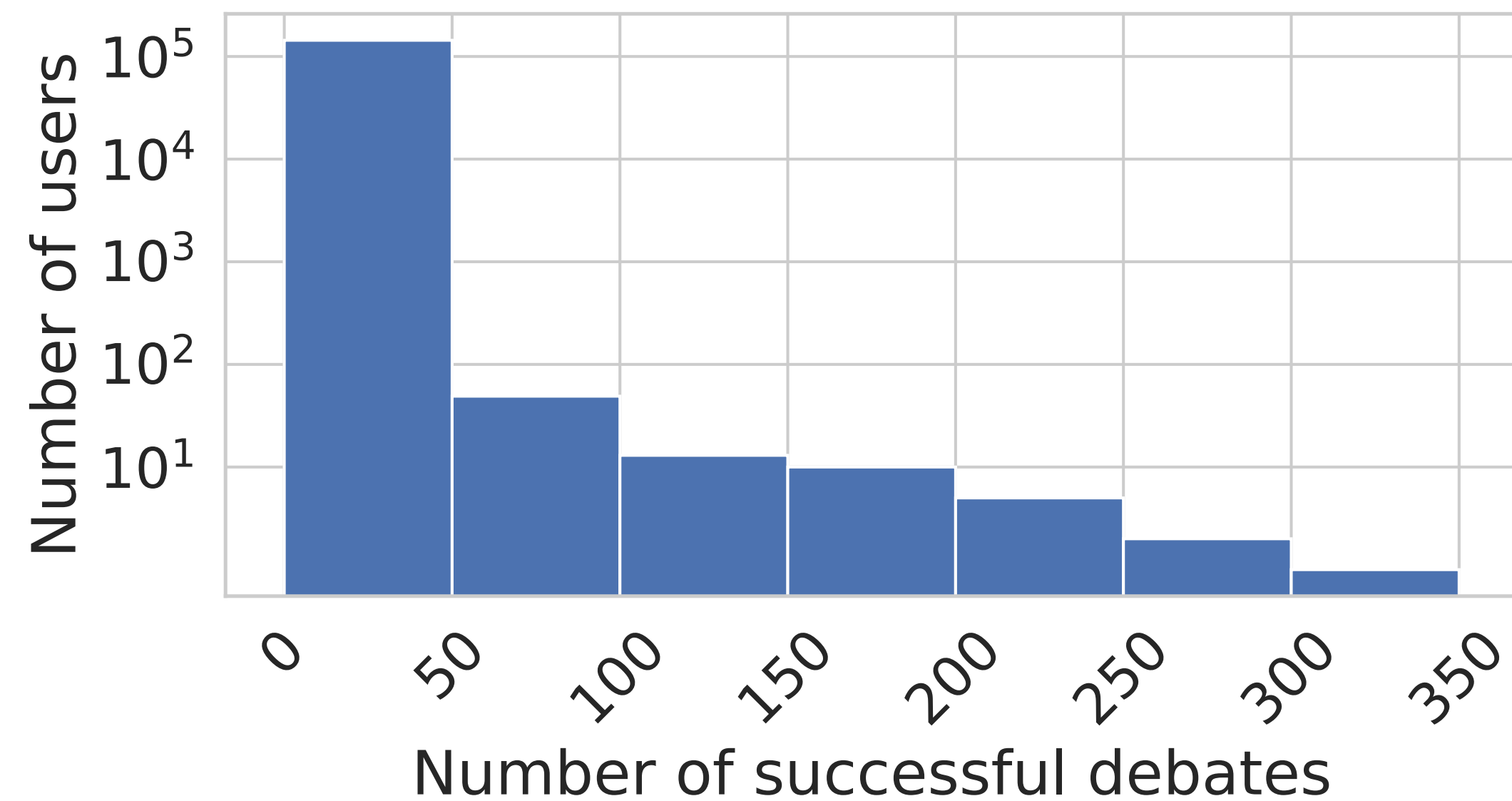
	Mean	Standard Deviation	Median
<i>Statistics of challengers in each debate</i>			
Reputation r_{pu}	15.9	43.4	1.0
Skill s_{pu} (%)	3.0	3.7	1.6
Position t_{pu}	14.8	24.3	8.0
Mean past position Z_{pu}	10.4	13.0	7.5
Number of past debates $\sum_{p' < p} S_{p'u}$	244.4	591.7	24.00
<i>Statistics of overall dataset</i>			
Number of opinions	91,730		
Opinions conceded	21,576		
Opinions leading to more than 1 debate	84,998	<i>(number of clusters with opinion fixed-effects)</i>	
Number of debates	1,026,201		
Successful debates	36,187		
Multi-party debates	348,041		
Number of debates per opinion	11.2	12.7	9
Successful debates per opinion	0.4	0.9	0
Number of unique posters	60,573		
Opinions per poster	1.5	2.4	1
Number of unique challengers	143,891		
Challengers with more than 1 debate	64,871	<i>(number of clusters with user fixed-effects)</i>	
Number of debates per challenger	7.1	58.5	1
Successful debates per challenger	0.3	3.2	0

Table 1: Descriptive Statistics. Debates from March 1, 2013 to October 10, 2019.

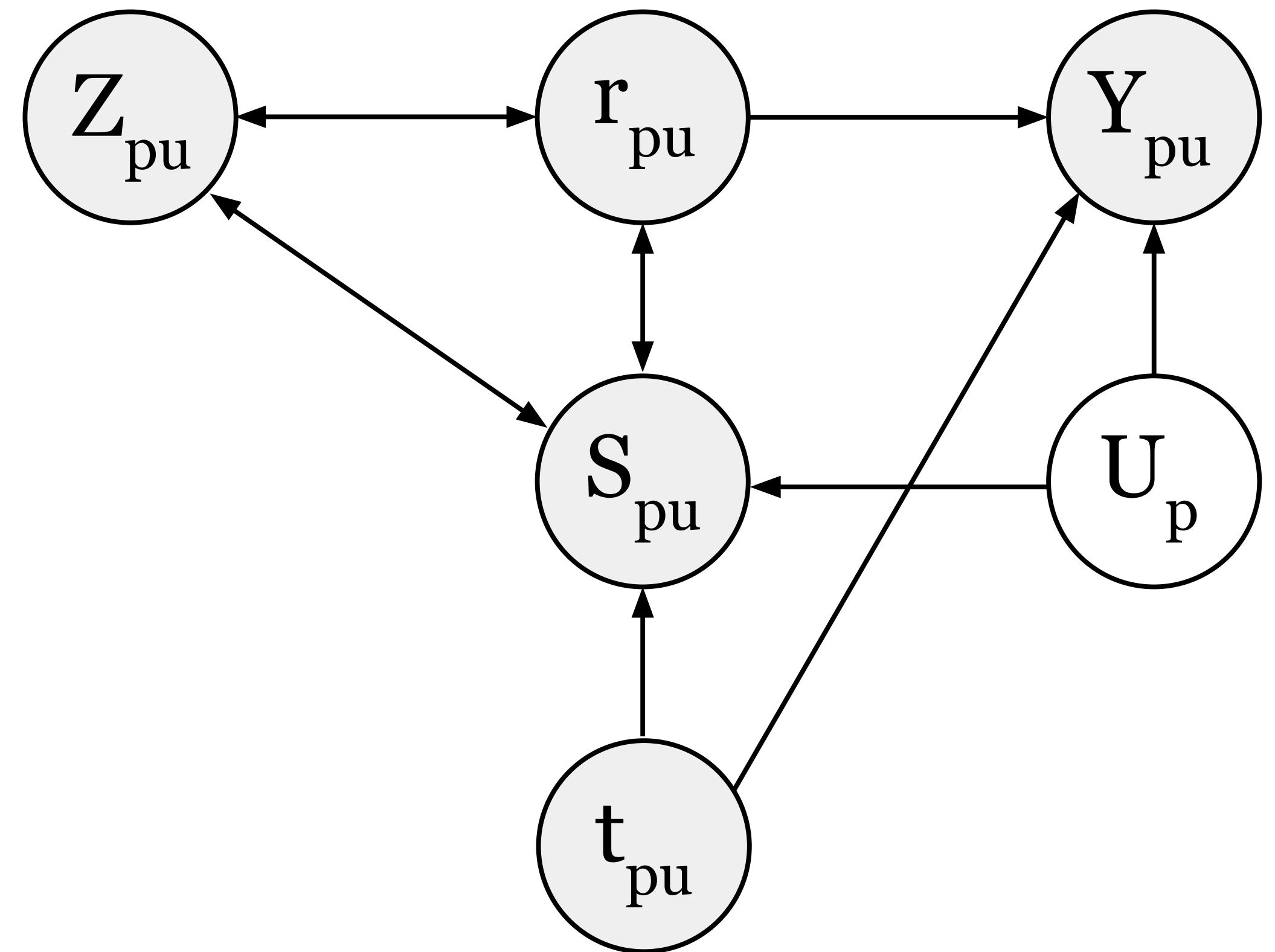
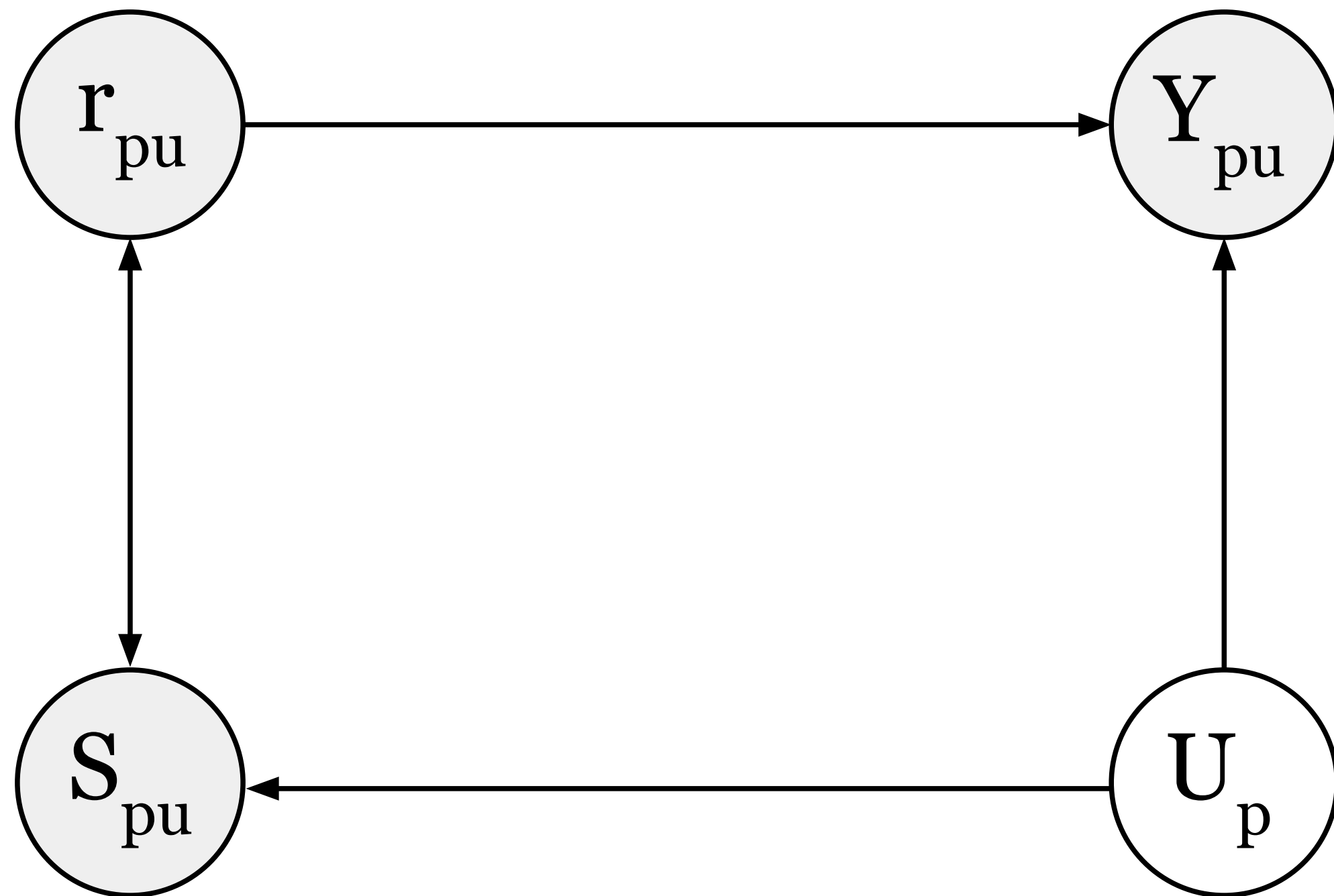
Skill vs. Experience



Debate Participation and Success



Endogenous Opinion Selection



Instrument First-Stage

	Dependent Variable: Reputation r_{pu}
Mean past position Z_{pu}	−0.1833 (0.003)***
Skill s_{pu} (percentage)	2.3055 (0.012)***
Position t_{pu} (std. deviations)	−1.7354 (0.067)***
Opinion fixed-effects (τ_p)	✓
Instrument F-Statistic	3, 338.7
No. of debates	1, 019, 469
R^2	0.22

Note: Standard errors displayed in parentheses. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 5: First-stage estimates. Mean past position as an instrument for reputation.

Double ML Estimation Procedure

1. Estimate the following conditional expectation functions on sample S' :

- i. $l(X_{pu}, \tau_p) = \mathbb{E}[Y_{pu}|X_{pu}, \tau_p]$ to get $\hat{l}(\cdot)$.
- ii. $q(X_{pu}, \tau_p) = \mathbb{E}[Z_{pu}|X_{pu}, \tau_p]$ to get $\hat{q}(\cdot)$.
- iii. $m_r(X_{pu}, \tau_p) = \mathbb{E}[r_{pu}|X_{pu}, \tau_p]$ to get $\hat{m}_r(\cdot)$.
- iv. $m_s(X_{pu}, \tau_p) = \mathbb{E}[s_{pu}|X_{pu}, \tau_p]$ to get $\hat{m}_s(\cdot)$.
- v. $m_t(X_{pu}, \tau_p) = \mathbb{E}[t_{pu}|X_{pu}, \tau_p]$ to get $\hat{m}_t(\cdot)$.

2. Estimate the following residuals on sample S :

- i. $\tilde{Y}_{pu} = Y_{pu} - \hat{l}(X_{pu}, \tau_p)$.
- ii. $\tilde{Z}_{pu} = Z_{pu} - \hat{q}(X_{pu}, \tau_p)$.
- iii. $\tilde{r}_{pu} = r_{pu} - \hat{m}_r(X_{pu}, \tau_p)$.
- iv. $\tilde{s}_{pu} = s_{pu} - \hat{m}_s(X_{pu}, \tau_p)$.
- v. $\tilde{t}_{pu} = t_{pu} - \hat{m}_t(X_{pu}, \tau_p)$.

3. Run a two-stage least-squares regression of \tilde{Y}_{pu} on $\tilde{r}_{pu}, \tilde{s}_{pu}, \tilde{t}_{pu}$ using \tilde{Z}_{pu} as an instrument for \tilde{r}_{pu} to obtain the estimated local average treatment effects of reputation, skill and position on debate success.

Neural Models of Text

Prediction target	Number of Hidden layers	Activation Functions		Loss Function
		Hidden Layer	Output Layer	
Debate success $Y_{pu} \in \{0, 1\}$	5	ReLU	Sigmoid	Binary Cross-Entropy
Reputation $r_{pu} \in \mathbb{Z}^+$	3	ReLU	Rectifier	Mean squared error
Skill $s_{pu} \in [0, 100]$ (percentage)	3	ReLU	Sigmoid	Mean squared error
Position $t_{pu} \in \mathbb{R}$ (standardized)	3	ReLU	Identity	Mean squared error
Instrument $Z_{pu} \in \mathbb{R}^+$	5	ReLU	Rectifier	Mean squared error

Table 7: Architectural hyperparameters. The input layer matrix \mathbf{W}_1 of each neural network has size $89,924 \times 4,926$, where 89,924 is the dimensionality of the input vector (the vocabulary size + the number of unique opinion clusters) and 4,926 is the dimensionality of X_{pu} (the vocabulary size). Each of the h hidden layer matrices $\mathbf{W}_2, \dots, \mathbf{W}_h$ has size $4,926 \times 4,926$, and the output layer matrix \mathbf{W}_{h+1} has size $4,926 \times 1$.

Neural Models of Text

Prediction target	Learning Rate	Batch Size	Weight-Decay	Subsample Loss		
				Train	Validation	Inference
Debate success $Y_{pu} \in \{0, 1\}$	0.0001	50,000	10000	0.148	0.155	0.152
Reputation $r_{pu} \in \mathbb{Z}^+$	0.0001	50,000	10	39.801	40.406	39.842
Skill $s_{pu} \in [0, 100]$ (percentage)	0.0001	50,000	10	3.672	3.764	3.707
Position $t_{pu} \in \mathbb{R}$ (standardized)	0.0001	50,000	10	0.658	0.789	0.796
Instrument $Z_{pu} \in \mathbb{R}^+$	0.0001	50,000	10000	12.389	13.370	13.217

Table 8: Optimization hyperparameters. The subsample losses on S'_{train} , S'_{val} and S are reported after training each neural network with the selected hyperparameters for at most 5,000 mini-batch iterations (with early-stopping) on S'_{train} . The binary cross-entropy subsample loss is reported for the network predicting Y_{pu} and the *root* mean squared prediction error is reported for the other networks.

Effect of Experience

$$Y_{pu} = \rho_u + m_{pu} + \theta_1 \sum_{p' < p} S_{p'u} + \theta_2 t_{pu} + \epsilon_{pu}$$

Dependent Variable: Debate Success Y_{pu}	
No. of opinions challenged previously $\sum_{p' < p} S_{p'u}$	-1×10^{-6} (0.7×10^{-6})
Position t_{pu} (std. deviations)	-0.0107 (0.0003)***
User fixed-effects (ρ_u)	✓
Month-year fixed-effects (m_{pu})	✓
No. of debates	947, 181
R^2	0.07

Note: Standard errors displayed in parentheses. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 3: Estimated effect of past experience on debate success.