



Malicious Behavior on the Web: Characterization and Detection

Srijan Kumar (@srijankr)

Justin Cheng (@jcccf)

Jure Leskovec (@jure)

Slides are available at <http://snap.stanford.edu/www2017tutorial/>

Tutorial Outline

Malicious users

Trolling

Sockpuppets

Vandals

Misinformation

Fake reviews

Hoaxes

<http://snap.stanford.edu/www2017tutorial>

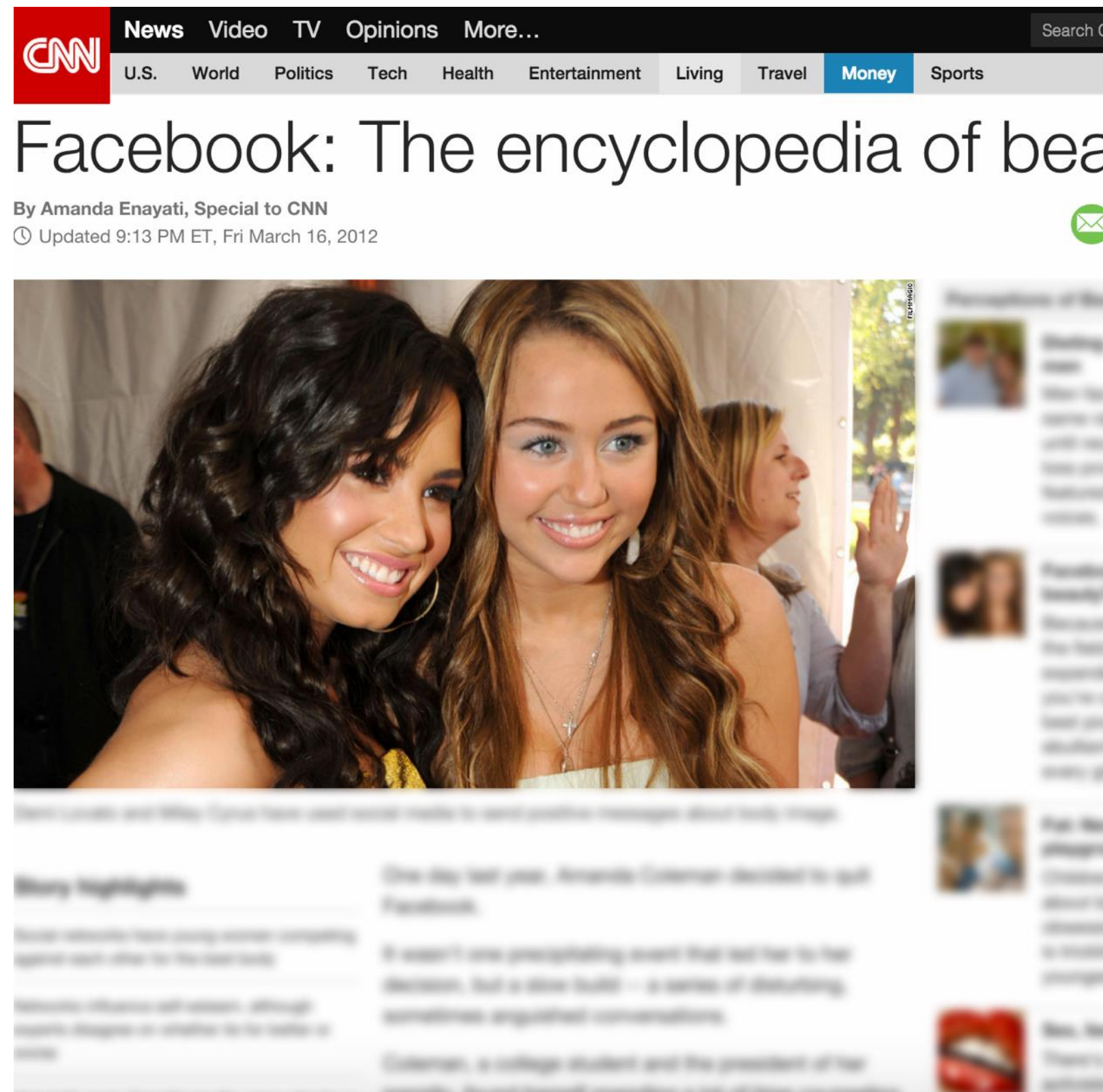
Introduction to Trolling



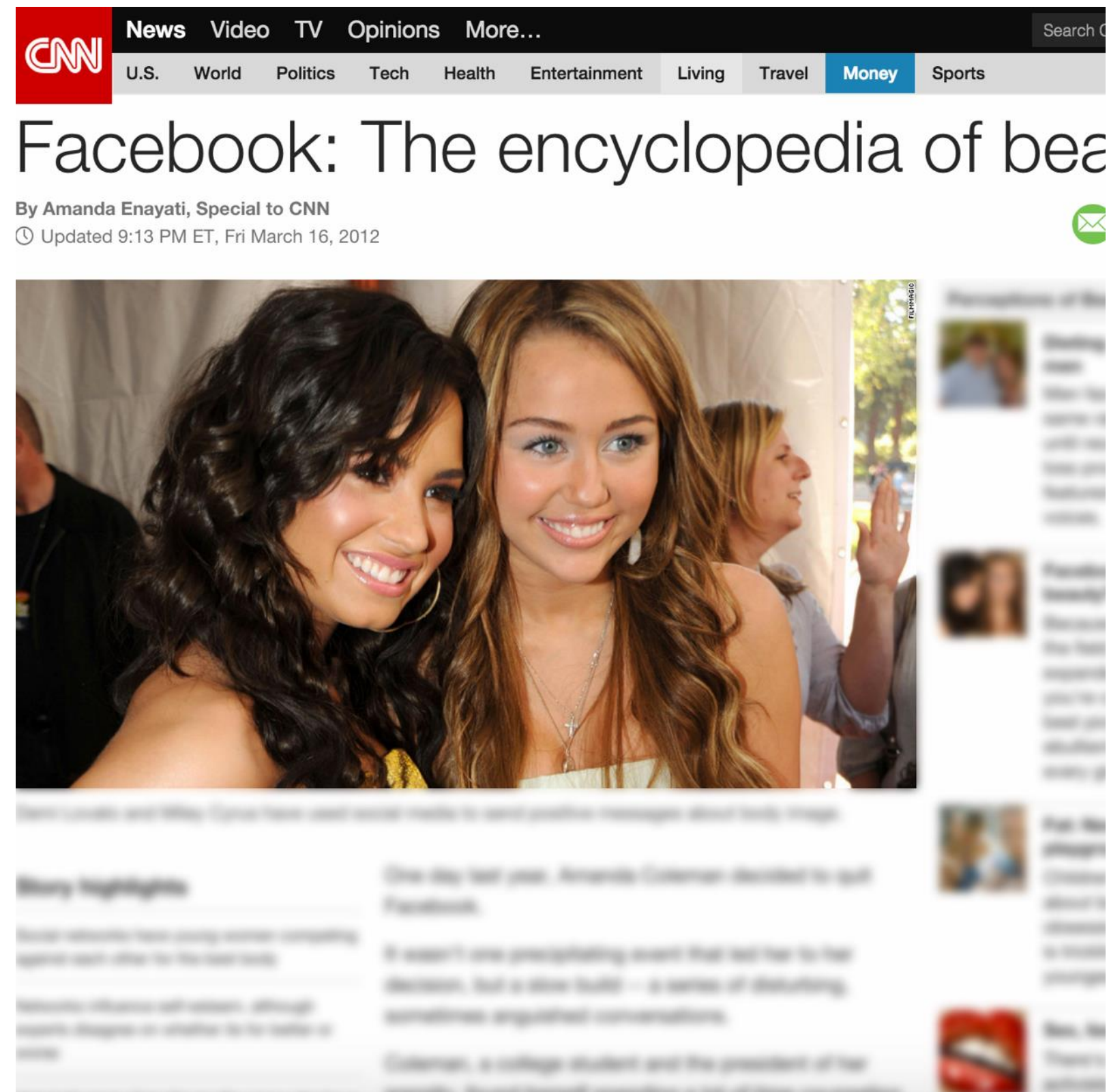
Warning!

This talk contains user posts with strong writing (profanity, sexism, racism, and religious intolerance).

An example from CNN



An example from CNN



It also shows that Islam and Christianity teaching women to dress modest could be right after all.

(These are actual comments)

An example from CNN



It also shows that Islam and Christianity teaching women to dress modest could be right afterall.



Religious nut alert



Clearly that is the only logical conclusion to this article. Now if you'll excuse me, I need to iron my tarp. I have work on Monday, and I want to appear 'modest'.



fail at life. go bomb yourself.

(These are actual comments)

Another example from NY Magazine



Another example from NY Magazine



MOTHERHOOD

February 10, 2016 10:50 a.m.

I'm Voting for Hillary Because of My Daughter

By Laura June

[f](#) Share | [t](#) Tweet | [g+](#) Share | [✉](#) Email | [📧](#)



Hillary is a c***. I am voting with my d*** for Putin.

(These are actual comments)

Another example from NY Magazine



MOTHERHOOD

February 10, 2016 10:50 a.m.

I'm Voting for Hillary Because of My Daughter

By Laura June

[f](#) Share | [t](#) Tweet | [g+](#) Share | [✉](#) Email | [📧](#)



Why we should never have let women vote.

(These are actual comments)

Because trolls are common in
comments sections...

THOUGHT CATALOG®

APRIL 30, 2014

The Worst Parts Of Humanity Live In The Comments Section

By **Caity Mae** • [View Comments](#) • 

Many websites have decided to
complete remove comments!

WHY WE'RE SHUTTING OFF OUR COMMENTS

We're turning comments off for a bit

**Sick of Internet comments? Us, too -
here's what we're doing about it**

Popular Science (2013); The Verge (2015); Chicago Sun-Times (2014)

Trolling research through the years

Troll Example #1: Ultimatego

Trolling as **disrupting a group while remaining undercover**

Trolled users in wedding Usenet newsgroups by being condescending

“People who could not get married in full formal splendor should not have a wedding at all but should simply go to city hall”

Troll Example #2: “Macho Joe”

Trolling as a “**flame war**”

Usenet troll who attacked alt.tv.melrose-place over a period of four months

“Oh BARF!!!! Last thing I want to see - a couple of *fags* making out on prime time! Why does the show need these weirdo characters anyway?”

Troll Example #3: “Kent”

Trolling as luring others into pointless and time-consuming discussions

A hostile male participant in a feminist forum made dozens of posts attacking forum members over a period of eight weeks

Presented himself as “sincerely interested in debating feminism”

But refused to acknowledge others’ points, willfully misinterpreted others’ views and taunted others.

Trolls on Wikipedia

Four trolls identified on Hebrew Wikipedia

Interviews of 15 Wikipedia sysops

Trolls are motivated by boredom, attention seeking, and revenge

Viewed vandalizing Wikipedia as a form of entertainment

Trolls as defined on Wikipedia

In Internet slang, a troll is a person who sows discord on the Internet by starting arguments or upsetting people, by posting inflammatory, extraneous, or off-topic messages in an online community with the intent of provoking readers into an emotional response or of otherwise disrupting normal, on-topic discussion, often for the troll's amusement.

What is trolling?

Engaging in negatively marked online behavior?

Taking pleasure in upsetting others?

Not following the rules?

Disrupting a group while staying undercover?

Our Definition:

Trolling is behavior that occurs
outside community norms

Defined using community guidelines
(e.g., name-calling, personal attacks, profanity, threats,
hate speech, ethnically/racially offensive material)



Several large comment-based
news communities



IGN[®]



76M users, 470M posts, 831M votes
A year of data

Operationalizing trolling



Troll: a user banned in the future

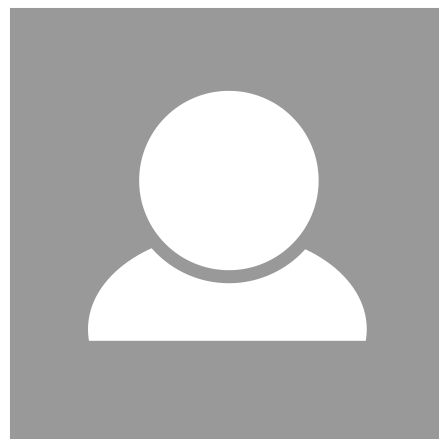


Non-troll: a user who was never banned?

Operationalizing trolling



Troll: a user banned in the future



Non-troll (matched): a user who was never banned, but similarly active.




Our Definition:

Trolling is behavior that occurs
outside community norms

A **post** is trolling if it was flagged/deleted by a moderator.

A **user** is a troll if they were eventually banned by a
moderator.

How common is trolling?

			
Post Deletions by moderators	21.4%	2.3%	2.7%
User Bans by moderators	3.3%	1.7%	2.2%

Research Question:
Why is trolling so prevalent?

Implication:
Understanding trolling helps us design
healthier, more prosocial communities



Prior work

Trolling is largely due to sociopaths



Donath (1999), Hardaker (2010), Buckels, et al. (2014)

Trolling is innate?

Trolls are born and not made

Baker (2001), Herring, et al. (2011), Shachaf-Hara (2010)

Trolls have particular personality types

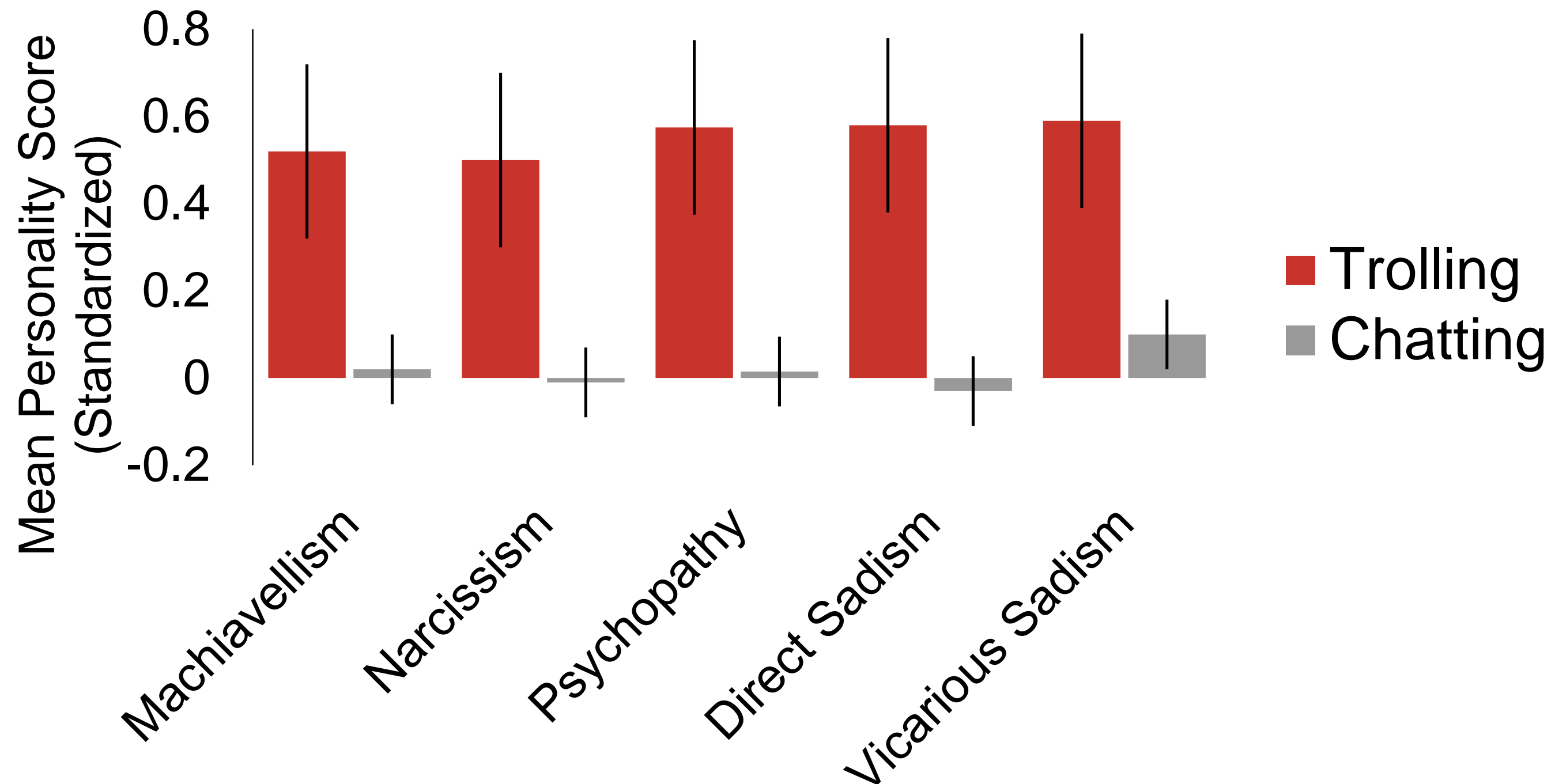
Buckels, et al. (2014), Raine (2002)

Trolls are sociopathic

Rensin (2014), Schwartz (2008)

Is trolling a personality trait?

Surveying 418 MTurk participants, several measures of psychopathic behavior are correlated with trolling!



Buckels, et al. (2014)

Are trolls just a vocal minority?

The Trolls Among Us

WHY THE TROLLS WILL ALWAYS WIN

**Confessions of a former
internet troll**

Donath (1999); Hardaker (2010); Shachaf & Hara (2010); NYT (2008); Wired (2014); Vox (2014)



This work

Trolling is due to ordinary people



How do trolls differ from non-trolls?

Cheng, J., Danescu-Niculescu-Mizil, C. & Leskovec, J. (2015). **Antisocial Behavior in Online Discussion Communities**. ICWSM 2015.

Comparing trolls and non-trolls

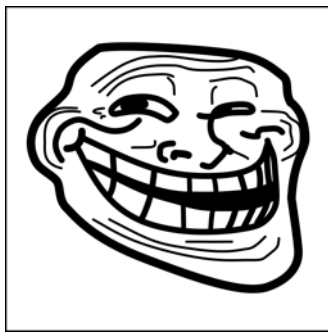


vs.



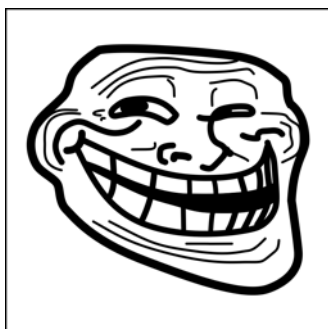
Statistical analysis of linguistic features and posting activity, comparing banned/non-banned users.

Comparing trolls and non-trolls



you get out of MY country, you f***ing a*****

Comparing trolls and non-trolls

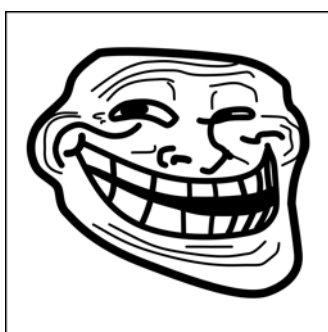


you get out of MY country, you f***ing a*****

More profane

Three times more swear words, $p < 10^{-2}$

Comparing trolls and non-trolls



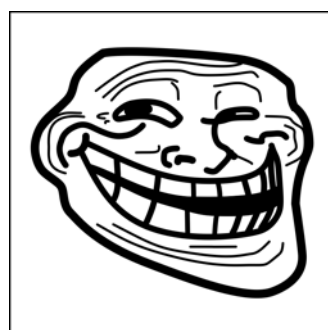
Facebook is for losers with no friends in real life.

Less positive
8% less positive words, $p < 10^{-4}$

Comparing trolls and non-trolls



Penny, once again you show why you are one of the best of the league. Always a class-act...

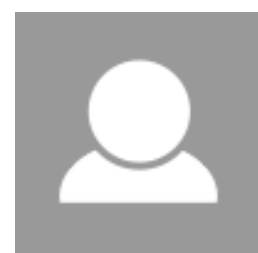


Why...do I not see any articles similar to this about white NBA basketball players?.....every single touchy feely story is about a black ball player.....YOU GUYS MAKE ME SICK AS A READER !

Less similar to previous posts

9% less similar (cosine similarity), $p < 10^{-4}$

not making



How many white NBA players grew up in the inner city? I'm sure there are many stories of charitable efforts by white NBA players, but this isn't a story about a Black NBA



How many white NBA players grew up in the

Comparing trolls and non-trolls

back to their roots to contribute.

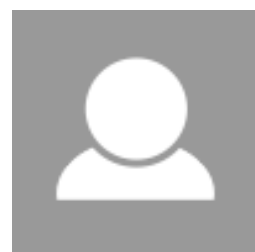


b\c young black men need to see examples from their own race. They need to see that even minorities can succeed and give back

Get more replies from other users

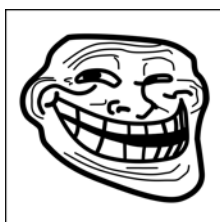
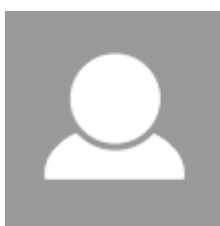
Twice as many replies, $p < 10^{-2}$

guidance.



A story about a millionaire helping kids in his poor neighborhood personally, and being a positive role model makes you sick as a reader. Gotta love conservatives.

Comparing trolls and non-trolls




Republican If you claim you want less government but want to control the bedroom, you're a Republican; If you want to cut Education, you're a Republican; If you want to cut Social Security, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican; If you want to cut Social Security, you're a Republican; If you want to cut Education, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican.

Security, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican If you claim you want less government but want to control the bedroom, you're a Republican; If you want to cut Education, you're a Republican; If you want to cut Social Security, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican; If you claim you want less government but want to control the bedroom, you're a Republican; If you want to cut Education, you're a Republican; If you want to cut Social Security, you're a Republican. If you want to cut Medicare and Medicaid, you're a Republican.

Security, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican. If you claim you want less government but want to control the bedroom, you're a Republican; If you want to cut Education, you're a Republican; If you claim you want less government but want to cut Social Security, you're a Republican; If you want to control the bedroom, you're a Republican; If you want to cut Medicaid, you're a Republican; If you want to cut Medicare, you're a Republican; If you want to cut Social Security, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican; If you want to cut Medicaid and Medicaid, you're a Republican.

want less government but want to control
If you claim you want less government, but
Republican, if you want to cut Education, you're a
bedroom you're a Republican; If you want to cut Education, you're a
to cut Social Security, you're a Republican; If you want to cut Medicare
Republican; If you want to cut Social Security, you're a Republican; If you
and Medicaid, you're a Republican; you're a Republican; If you want to cut
want to cut Medicare and Medicaid, you're a Republican; If you claim you
Social Security, you're a Republican; If you want to cut Medicare and
want less government but want to control the bedroom, you're a
Medicaid, Republican.
Republican, if you want less government, but want to control the
Republican, if you want to cut Education, you're a Republican; If you want
bedroom you're a Republican; If you want to cut Education, you're a
to cut Social Security, you're a Republican; If you want to cut Medicare

 Republican: If you want to cut Social Security, you're a Republican; if you want to cut Medicare and Medicaid, you're a Republican; if you want to cut Social Security, you're a Republican; if you want to cut Medicare and Medicaid, you're a Republican; if you claim you want less government but want to control the bedroom, you're a Republican.

Republican; If you want to cut Education, you're a Republican; If you want to cut Social Security, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican; If you want to control the bedroom, you're a Republican; If you want less government, you're a Republican; If you want to cut Social Security, you're a Republican; If you want to cut Education, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican; If you want to control the bedroom, you're a Republican; If you want less government, you're a Republican.

Republican If you claim you want less government but want to control the bedroom, you're a Republican; If you want to cut Education, you're a Republican; If you want to cut Social Security, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican; If you want to cut Social Security, you're a Republican; If you want to cut Education, you're a Republican; If you want to cut Social Security, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican.

Security, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican If you claim you want less government but want to control the bedroom, you're a Republican; If you want to cut Education, you're a Republican; If you want to cut Social Security, you're a Republican; If you want to cut Medicare and Medicaid, you're a Republican you're a Republican; If you want to cut Social Security, you're a Republican; If you want to cut Medicare and Medicaid, Republican.

Post more per thread

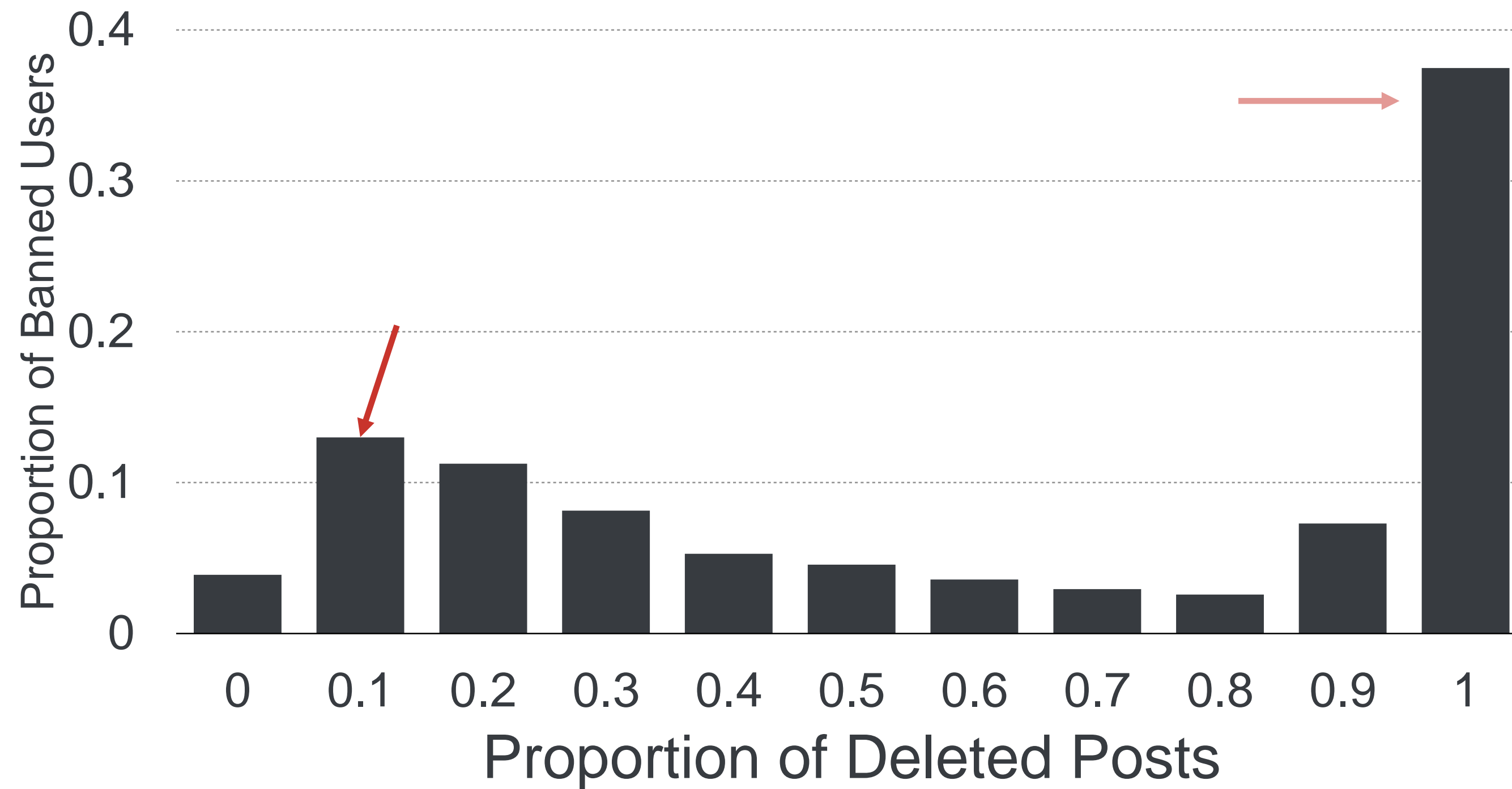
47% more posts, $p < 10^{-2}$

Causes of trolling

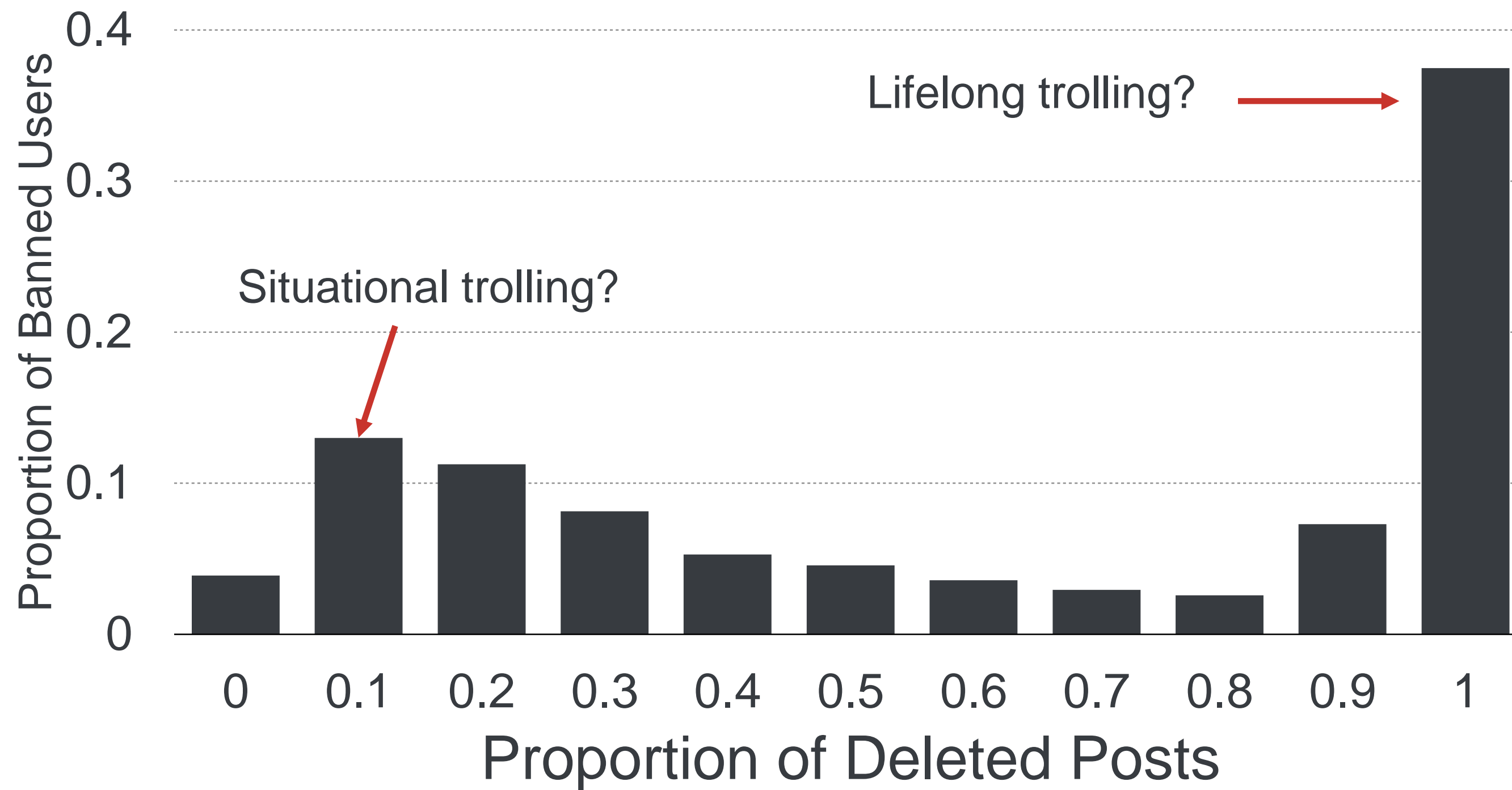
Cheng, J., Bernstein, M.S., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). **Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions**. CSCW 2017.

How much do trolls troll?

The distribution of trolls is bimodal



Are there two types of trolls?



What if trolling behavior is
situational?

Challenge

How to show that trolling is situational?

How to show that trolling is situational?

Observational data isn't causal

How to show that trolling is situational?

Experiments are hard to generalize

How to show that trolling is situational?

Solution: online experiment + observational study

Hypothesis

Anyone can become a troll



Theory 1: “Broken windows”

Norms strongly signal whether further bad behavior happens

Zimbardo (1969); Wilson (1982)



Theory 2: Unpleasant stimuli increase aggression

Aversives can lead people to act aggressively towards others, even if they weren't responsible.

Jones & Bogat (1978); Rotton & Frey (1985)

Online experiment simulating a discussion forum



N=667 (40% female)

Online experiment simulating a discussion forum

Complete a quiz, then participate in a discussion



Quiz



Discussion

Online experiment simulating a discussion forum
**We manipulated quiz difficulty and discussion
context**

Online experiment simulating a discussion forum

The quiz was either easy or difficult



Online experiment simulating a discussion forum

Discussion context was either positive or negative



Easy quiz (positive mood condition)

Qualification Test

Unscramble the following letters to form an English word:

"P A P H Y"

Subtract three thousand from five thousand. Write your answer in words.

How many 'I's are there in these sentences?

"I wanted to enjoy the play but I left."

179.5 seconds left

Difficult quiz (negative mood condition)

Qualification Test

Unscramble the following letters to form an English word:

"D E A N Y O N"

Subtract six thousand, seven hundred eighty-three from eight thousand, eleven. Write your answer in words.

How many 'I's are there in these sentences?

"I really wanted to like this play but I was flumoxxed by the blatant lying so I ultimately left. I really really wanted to like this play but I was flumoxxed by

272.5 seconds left

Positive discussion context condition

News of the Day

I'm Voting for Hillary Because of My Daughter

Back in the 2008 primary season, I supported Hillary Clinton. That choice...

Top Comments Sorted by Best



User1337 · 2 hours ago

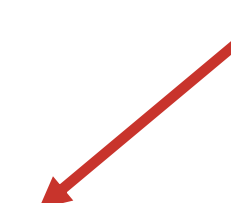
I'm a woman, and i don't think you should vote for a woman just because she is a woman. vote for her because you believe she deserves it.

6  |  · [Reply](#)



User9054 · 3 hours ago

Personally, I'd vote for whoever I think is the best and



Negative discussion context condition

News of the Day

I'm Voting for Hillary Because of My Daughter

Back in the 2008 primary season, I supported Hillary Clinton. That choice...

Top Comments Sorted by Best



User1337 · 2 hours ago

Oh yes. By all means, vote for a Wall Street sellout - - a lying, abuse-enabling, soon-to-be felon as our next President. And do it for your daughter. You're quite the role model.

1 ^ | v · [Reply](#)



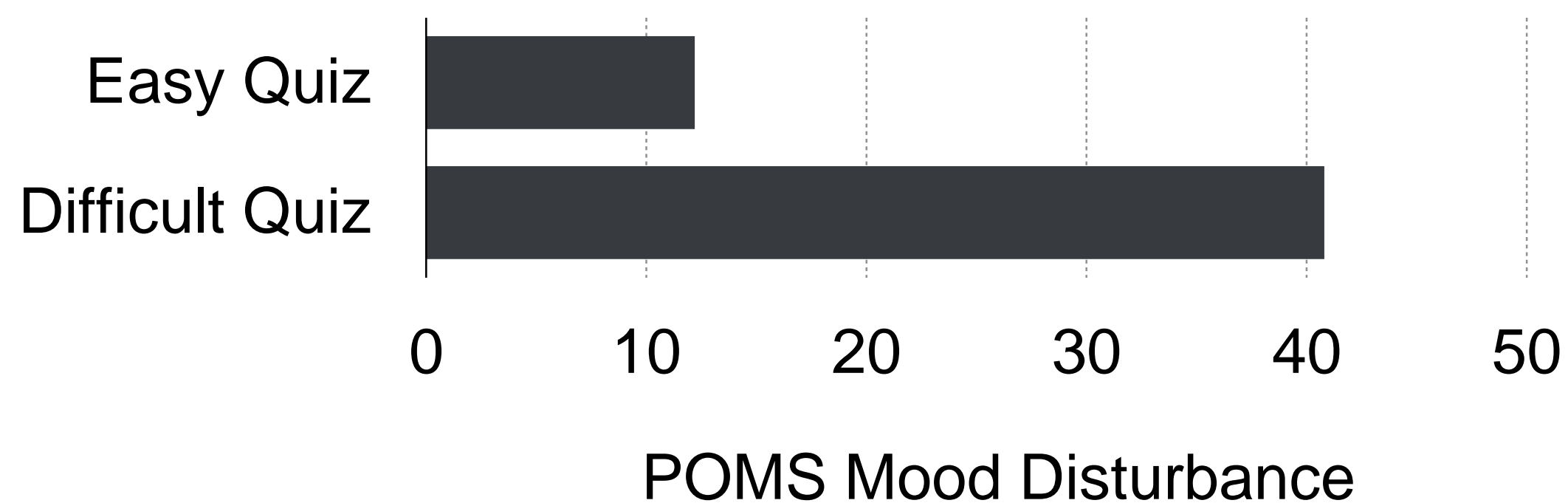
User9054 · 3 hours ago

Hillary is a cunt. I am voting with my dick for Putin. /s



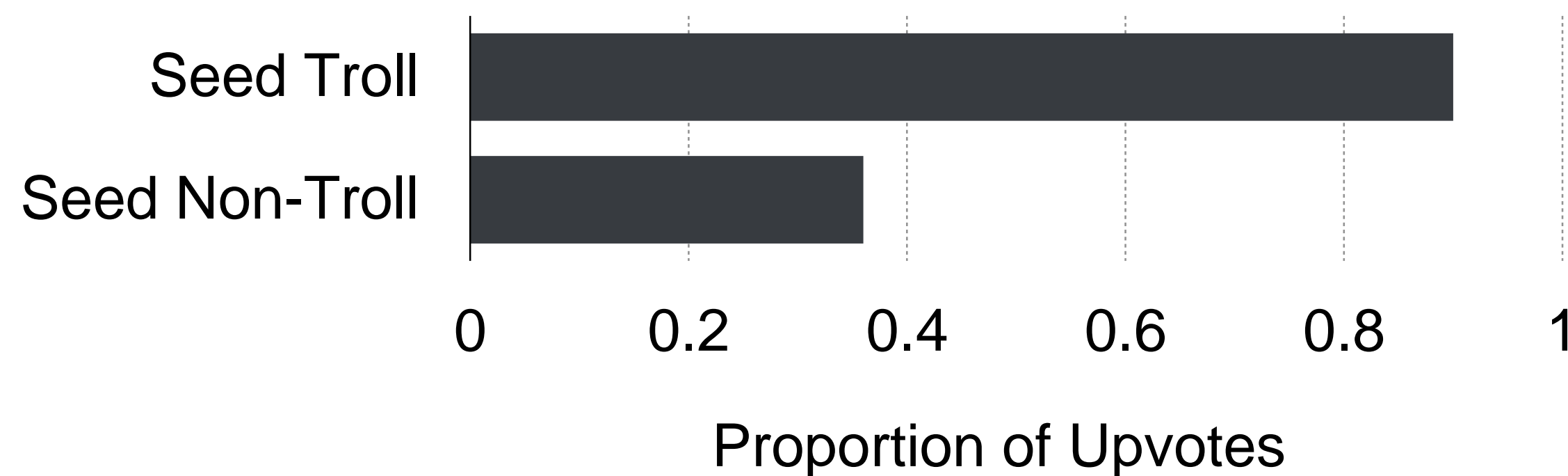
Manipulation checks

People were in a worse mood after the difficult quiz



Manipulation checks

People also perceived seed troll posts as worse



How much trolling was there in each condition?

posts independently labeled by two expert
raters using community guidelines

How much trolling was there in each condition?

% Troll Posts	Positive Mood	Negative Mood
Positive Context		
Negative Context		

Trolling is lowest in the positive conditions...

% Troll Posts	Positive Mood	Negative Mood
Positive Context	35%	
Negative Context		

...increases with either negative
condition...

% Troll Posts	Positive Mood	Negative Mood
Positive Context	35%	49%
Negative Context	47%	

...and almost doubles in the negative conditions

% Troll Posts	Positive Mood	Negative Mood
Positive Context	35%	49%
Negative Context	47%	68%

- ($p < 0.05$ using a mixed effects logistic regression model)

Negative affect also triples

% Neg. Affect Words (LIWC)		Positive Mood	Negative Mood
	Positive Context	1.1%	1.4%
	Negative Context	2.3%	2.9%

• ($p < 0.05$)

Comment from the positive mood/context condition:

Comment from the positive mood/context condition:

“Hillary is a solid candidate. As a woman, I appreciate that she's a woman, but it's not the only reason I think she would do well in office.”

Comment from the negative mood/context
condition:

Comment from the negative mood/context condition:

“Anyone who votes for her is a complete idiot. These supporters are why this country is in such bad shape now. Uneducated people.”

Bad mood and negative context increase trolling

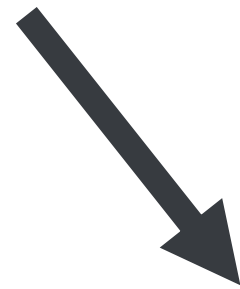
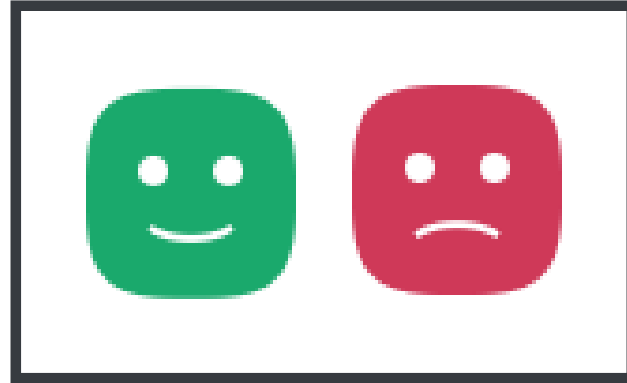


User



Troll or not?

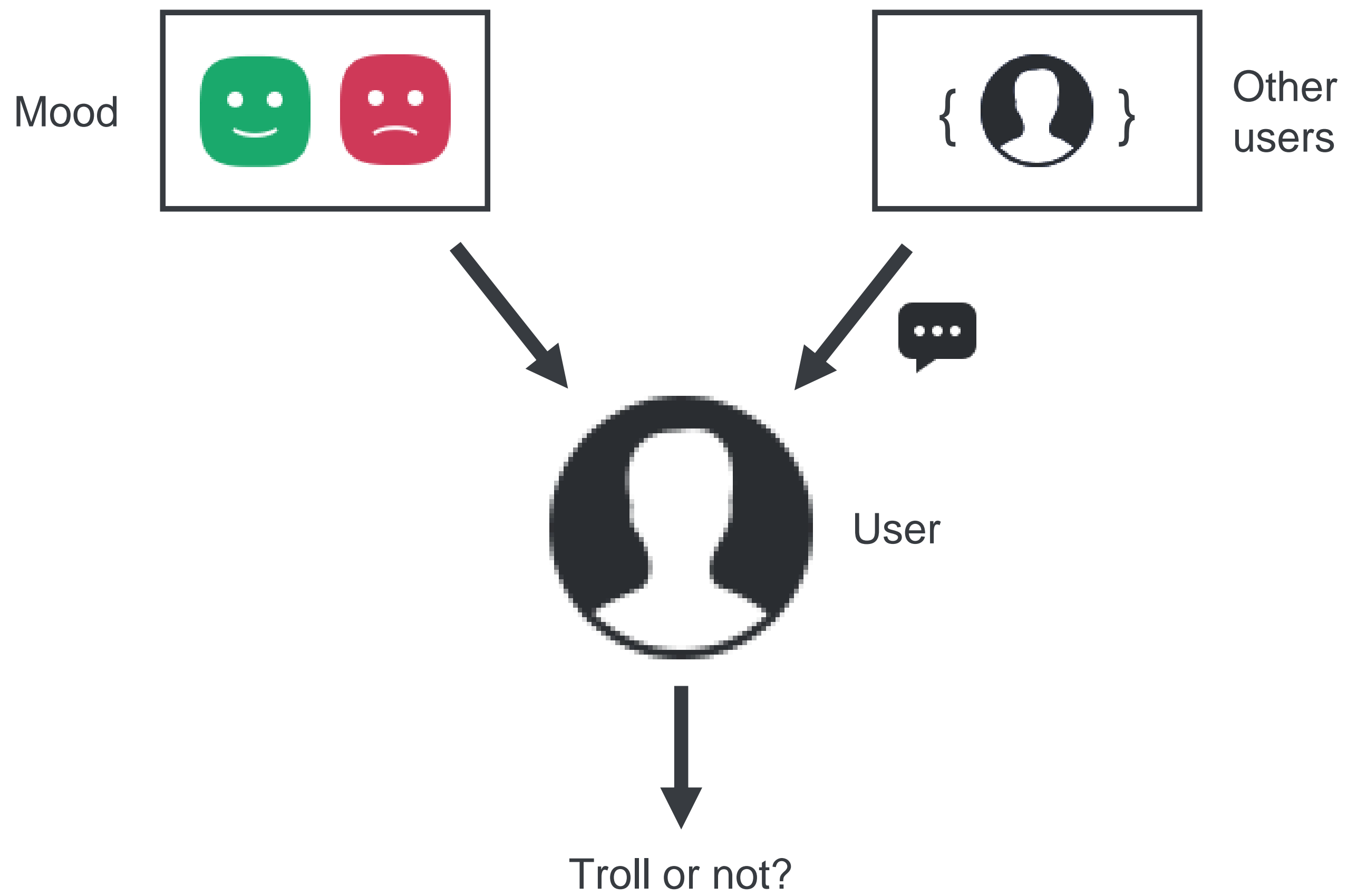
Mood



User



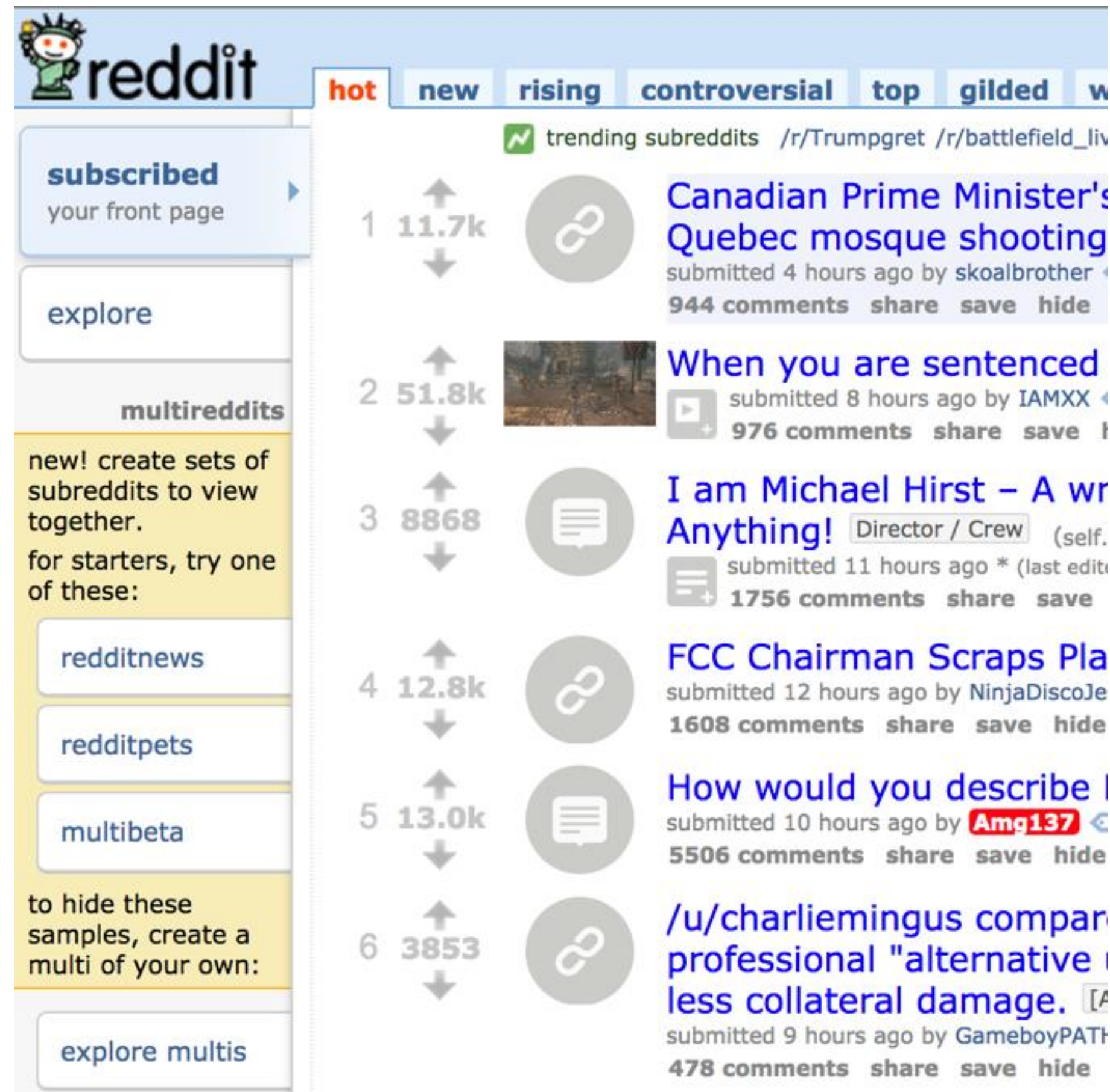
Troll or not?



Could voting reduce bad behavior?



Voting regulates content quality (?)



The screenshot shows the Reddit interface with the 'hot' tab selected. The left sidebar contains navigation options: 'subscribed' (your front page), 'explore', 'multireddits' (with a prompt to create sets of subreddits and examples like redditnews, redditpets, and multibeta), and 'explore multis'. The main content area displays a list of posts ranked by votes, with up and down arrows indicating the voting mechanism. The posts are as follows:

Rank	Votes	Post Title	Submitted	Comments
1	11.7k	Canadian Prime Minister's Quebec mosque shooting	4 hours ago by skoalbrother	944
2	51.8k	When you are sentenced	8 hours ago by IAMXX	976
3	8868	I am Michael Hirst – A wr Anything! Director / Crew (self.)	11 hours ago * (last edit)	1756
4	12.8k	FCC Chairman Scraps Pla	12 hours ago by NinjaDiscoJe	1608
5	13.0k	How would you describe l	10 hours ago by Amg137	5506
6	3853	/u/charliemingus compar professional "alternative less collateral damage.	9 hours ago by GameboyPATh	478

Our claim

Downvoting causes negative
behavior to worsen

How does trolling behavior spread?

Cheng, J., Danescu-Niculescu-Mizil, C. & Leskovec, J. (2014). **How Community Feedback Shapes User Behavior**. ICWSM 2014.

What are the effects of evaluations?





Mark · 7 hours ago

You do not use Facebook, Facebook uses you.

5 ^ | 4 v

· Reply · Share ›

But how do we summarize
feedback?

P

Number of upvotes?

15⁺ 0⁻ vs. 15⁺ 15⁻

Doesn't account for down-votes.

P-N

Difference in up- and downvotes?

5⁺ 0⁻ vs. 50⁺ 45⁻

Doesn't account for proportion of downvotes.

$$P/(P+N)$$

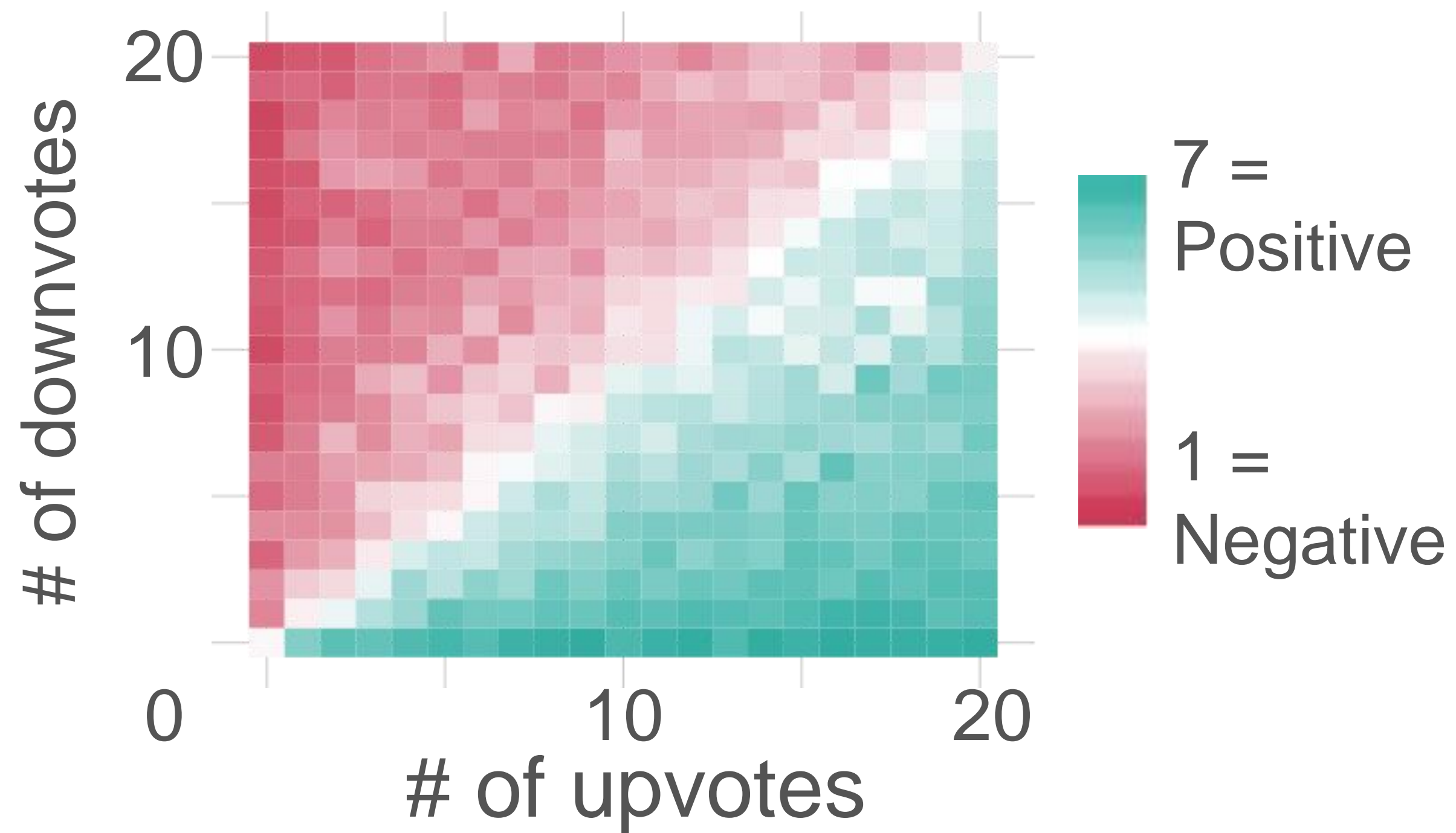
Proportion of up-votes?

4⁺ 1⁻ vs. 40⁺ 10⁻

Doesn't account for total number of votes.

How do we measure how users
perceived these votes?

User ratings are independent of the number of votes



How do these measures correlate with user ratings?

R^2

P

0.410

Number of up-votes

$P-N$

0.879

Difference in up/down-votes

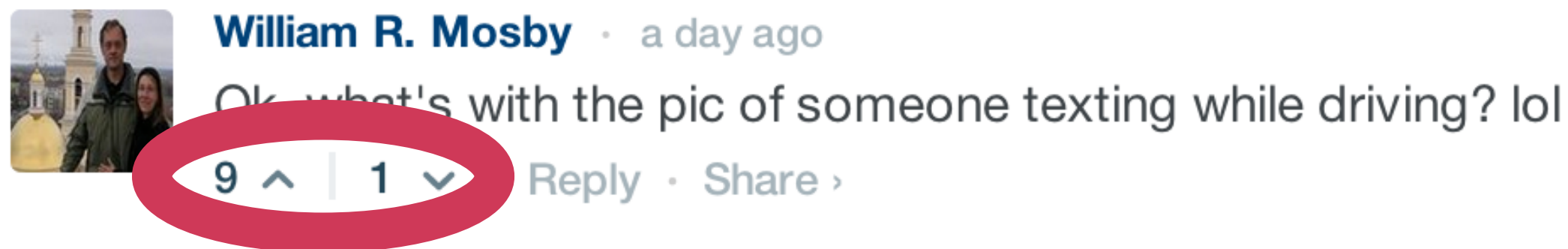
$P/(P+N)$ ✓

0.920

Proportion of up-votes

Defining positive/negative evaluations

A post is **positively evaluated** when the proportion of up-votes is higher than a given threshold (e.g., ≥ 75 th percentile).



$$P / (P+N) = 9 / (9+1) = 0.9 \geq \text{High Threshold}$$

Defining positive/negative evaluations

A post is **negatively evaluated** when the proportion of up-votes is lower than a given threshold (e.g., ≤ 25 th percentile).



$$P / (P+N) = 2 / (2+8) = 0.2 \leq \text{Low Threshold}$$

How will users' behavior change
after an evaluation?

Do users improve?

Operant conditioning predicts that feedback would guide authors towards better behavior.

In other words, up-votes are “reward” stimuli, and down-votes are “punishment” stimuli.

Skinner (1938)

Or do they get worse?

Feedback can have negative effects. People given only positive feedback tend to become complacent.

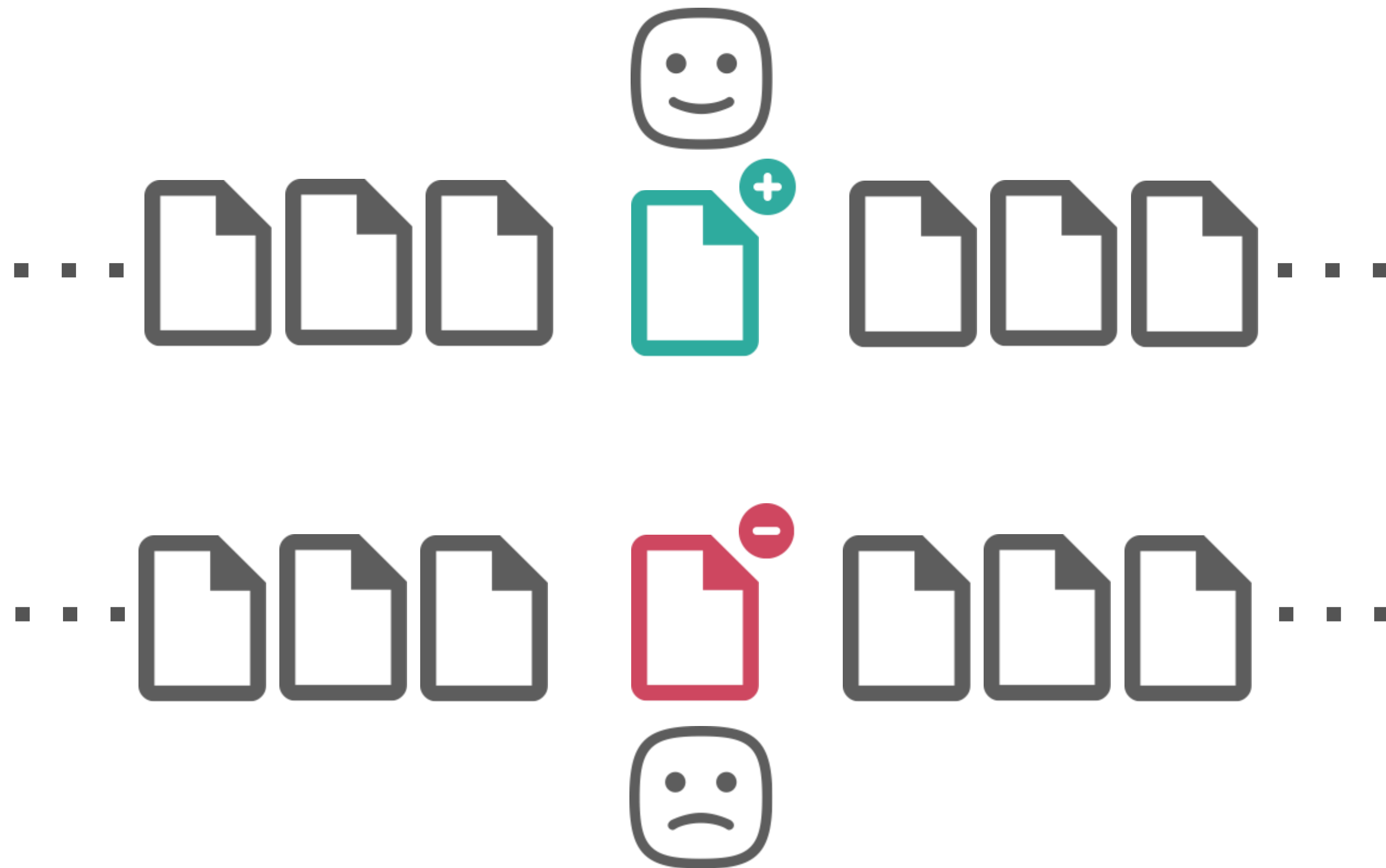
Also, bad impressions are quicker to form and more resistant to disconfirmation.

Brinko (1993), Baumeister, et al. (2001)

The effect of evaluations



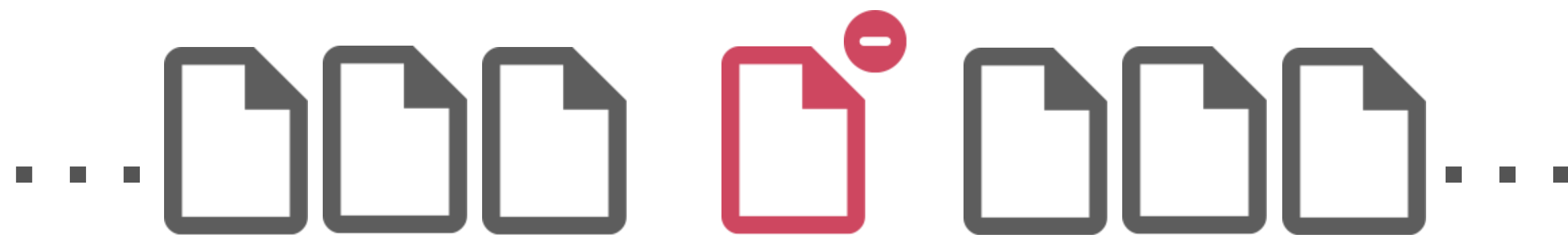
The effect of evaluations



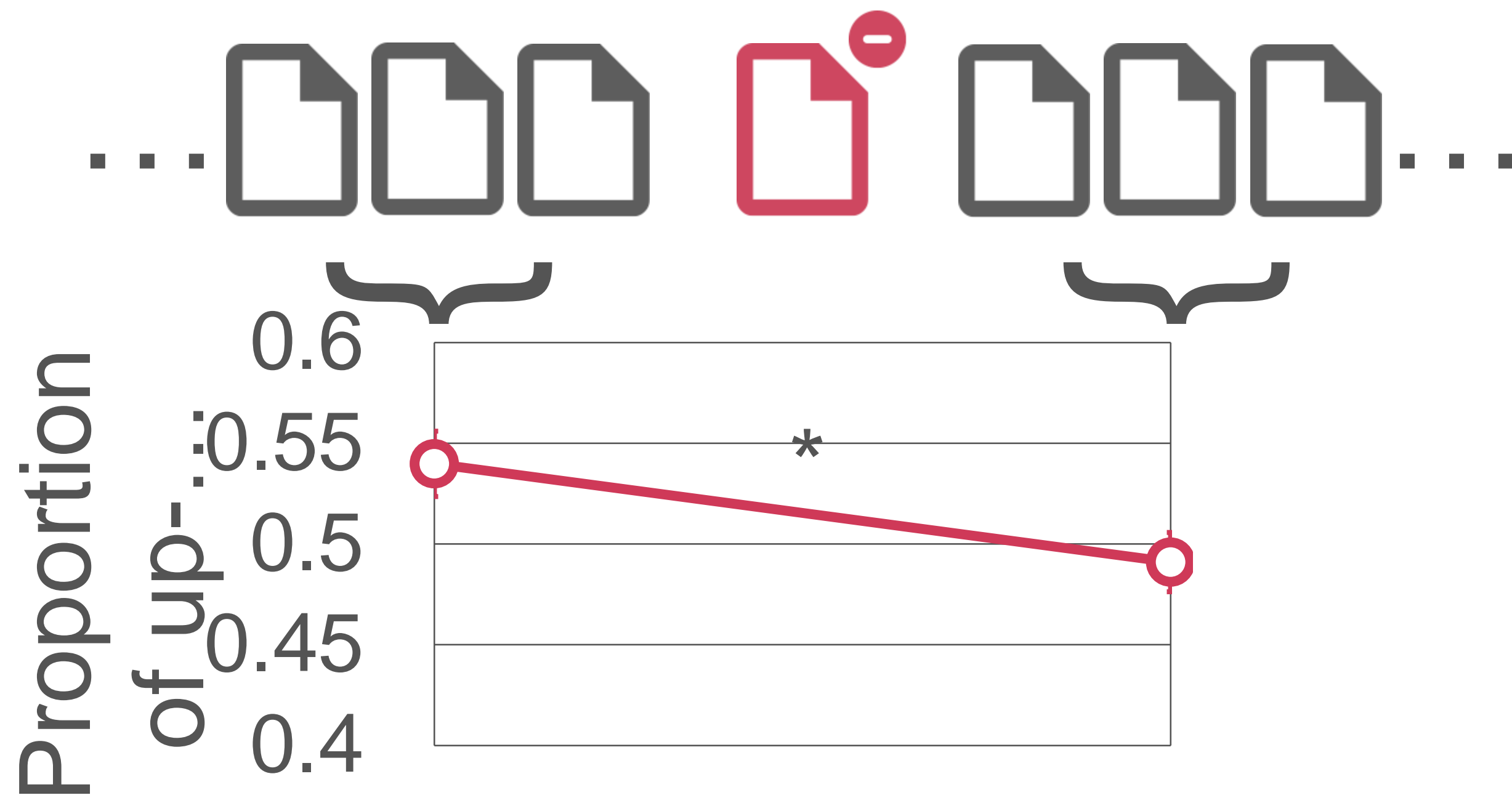
Are users evaluated better/worse
after a positive evaluation?



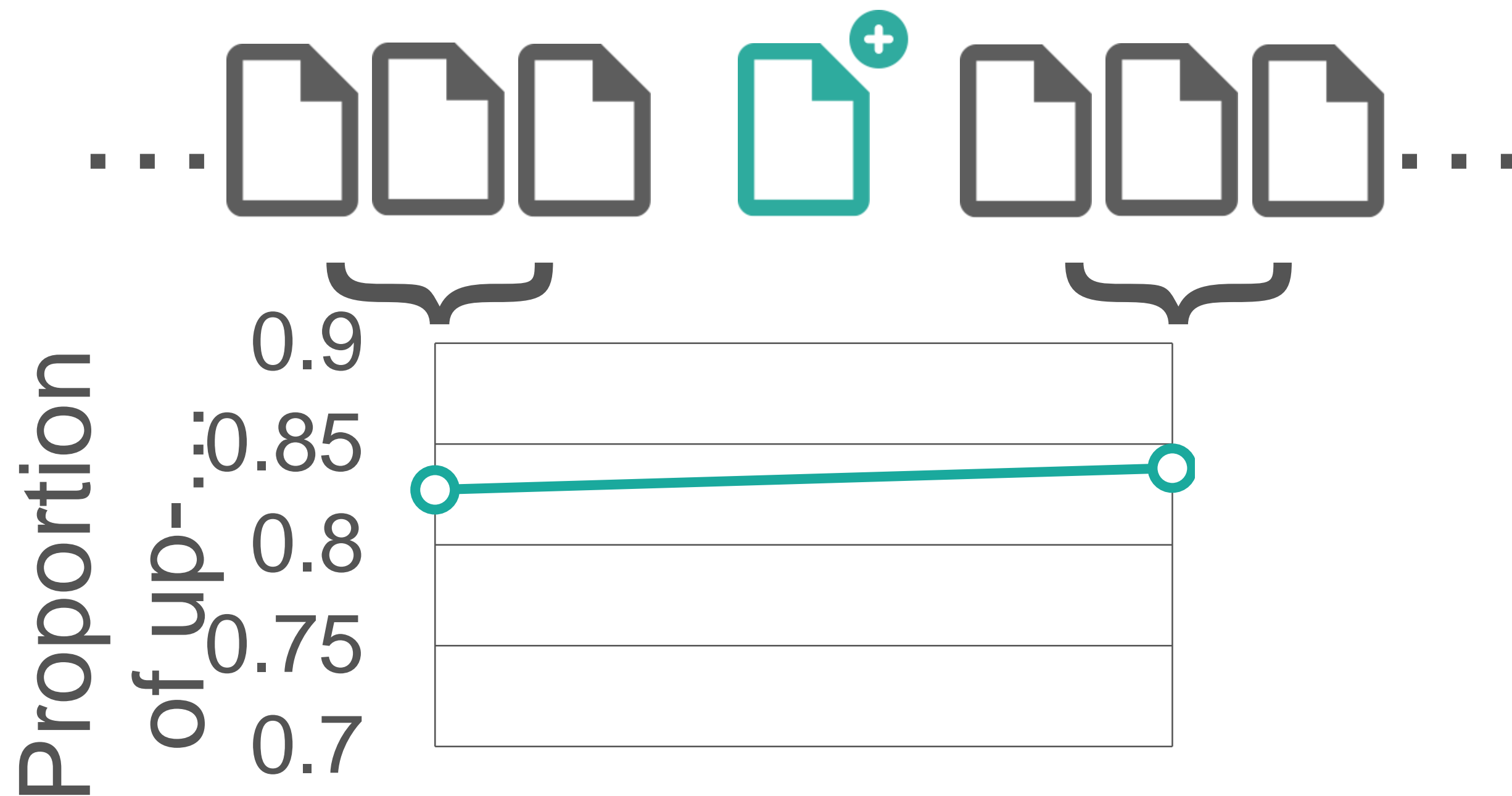
Are users evaluated better/worse
after a negative evaluation?



Negatively-evaluated users are
evaluated worse in the future



Positively-evaluated users are evaluated
no better in the future

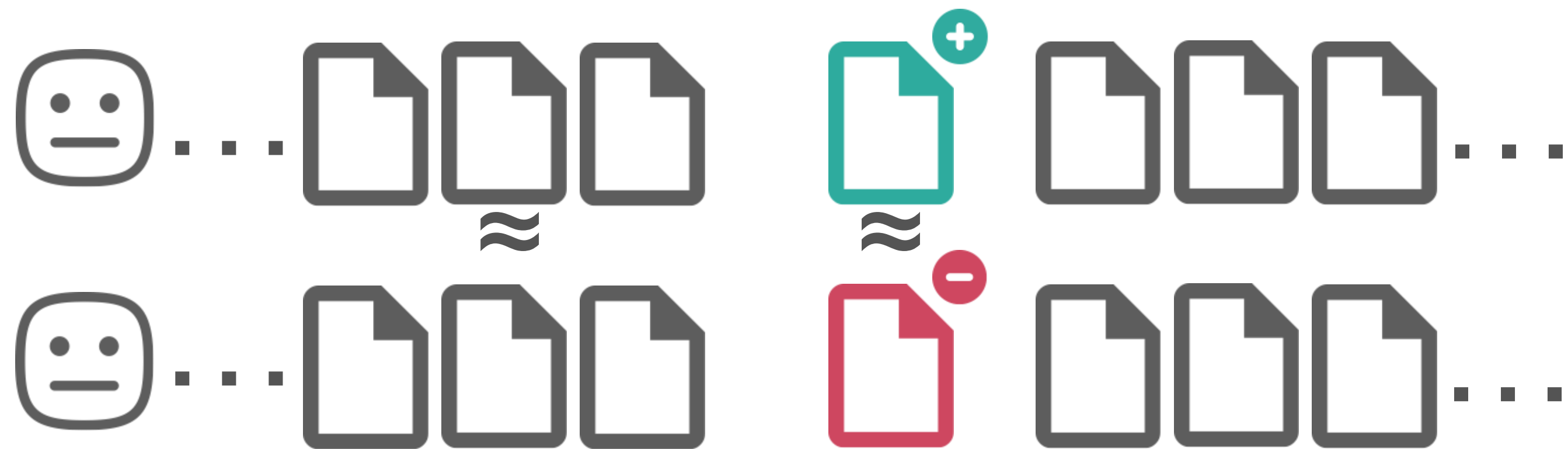


Feedback doesn't seem to improve
future behavior.

Challenge: how to compare different
users and posts?

Are downvoted users/posts inherently worse?

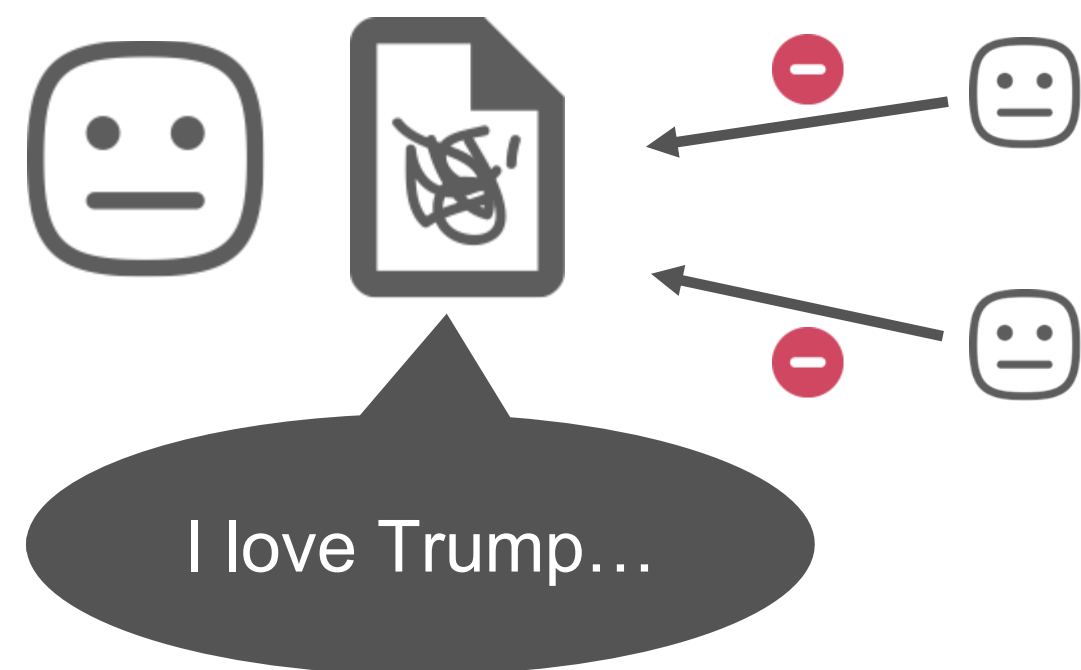
Solution: Propensity score matching



Rosenbaum & Rubin (1983)

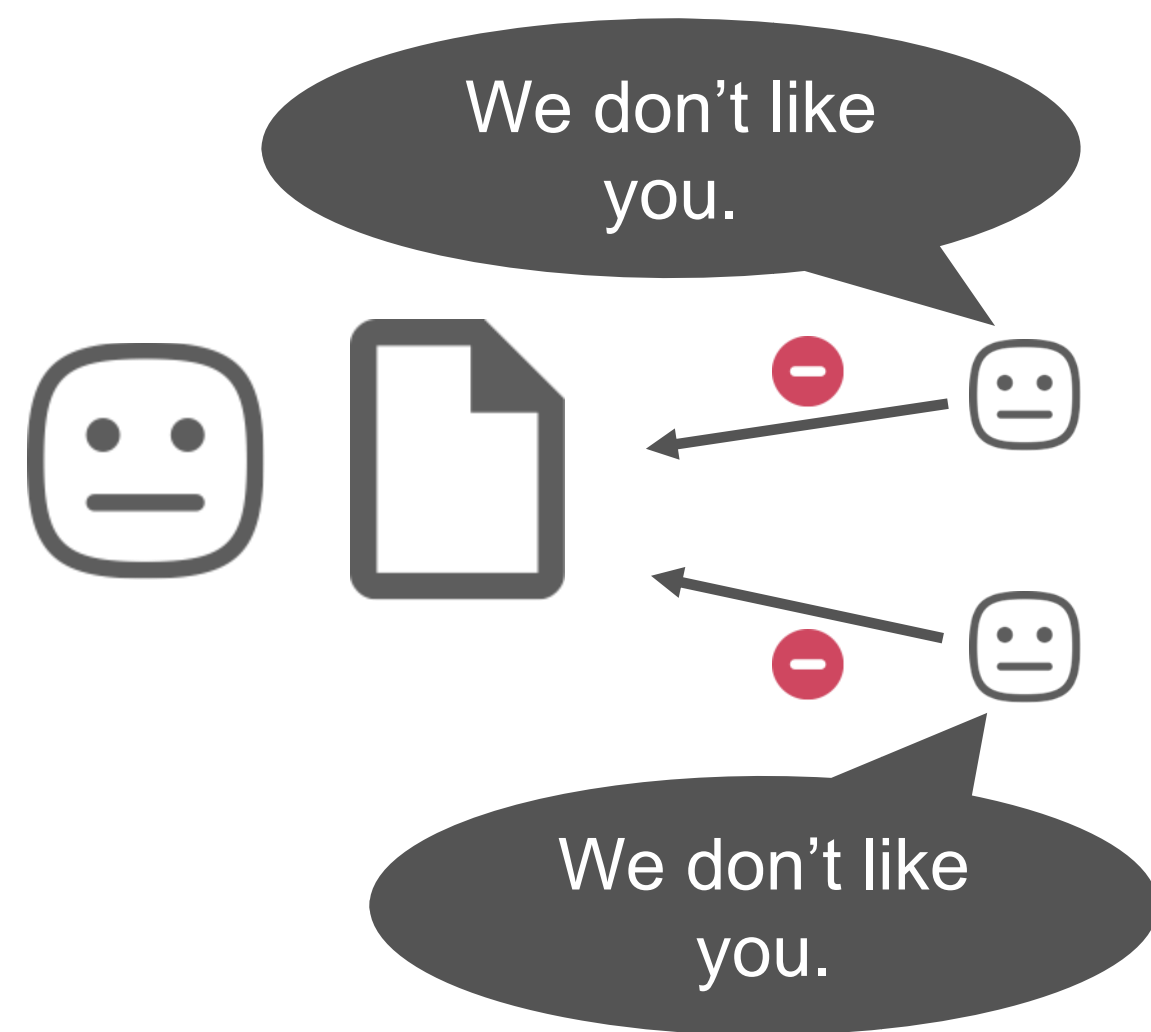
How much is an evaluation due to
textual or community effects?

How much is an evaluation due to textual or community effects?




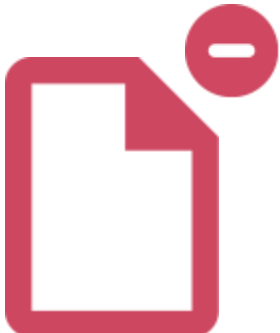


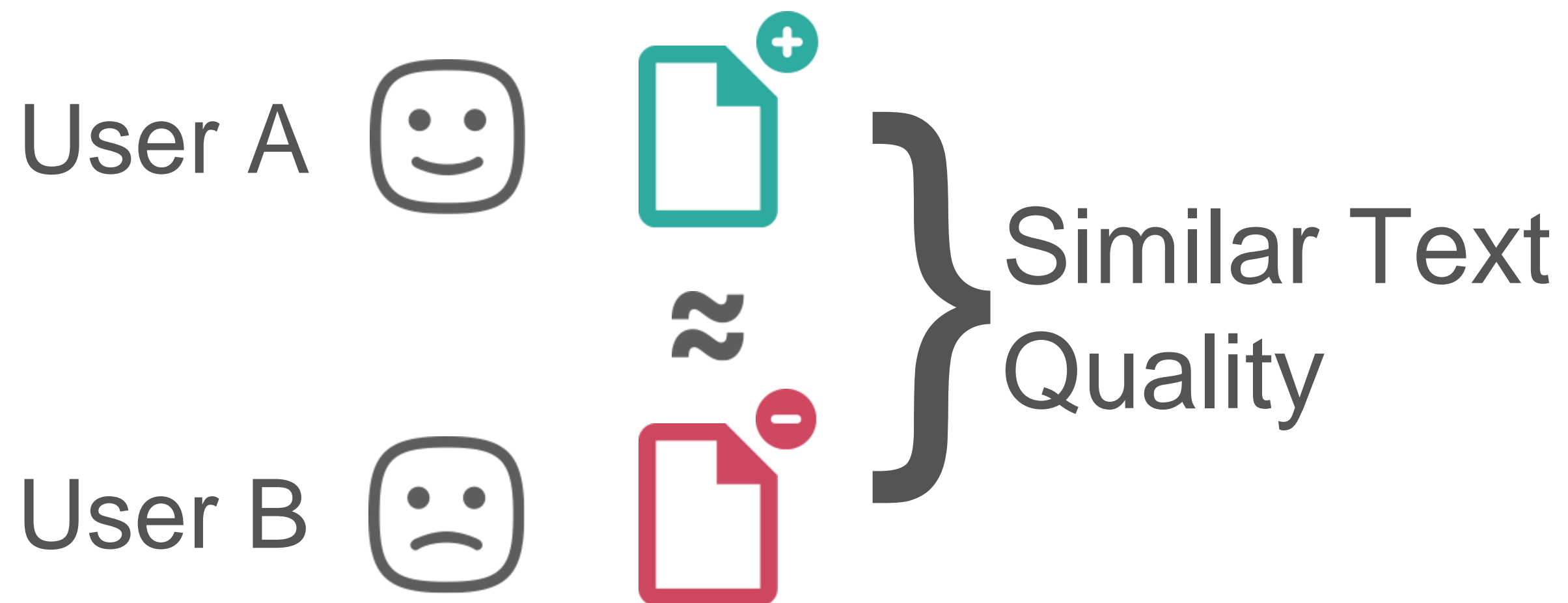
i.e. down-voting because of the post content

How much is an evaluation due to textual or community effects?



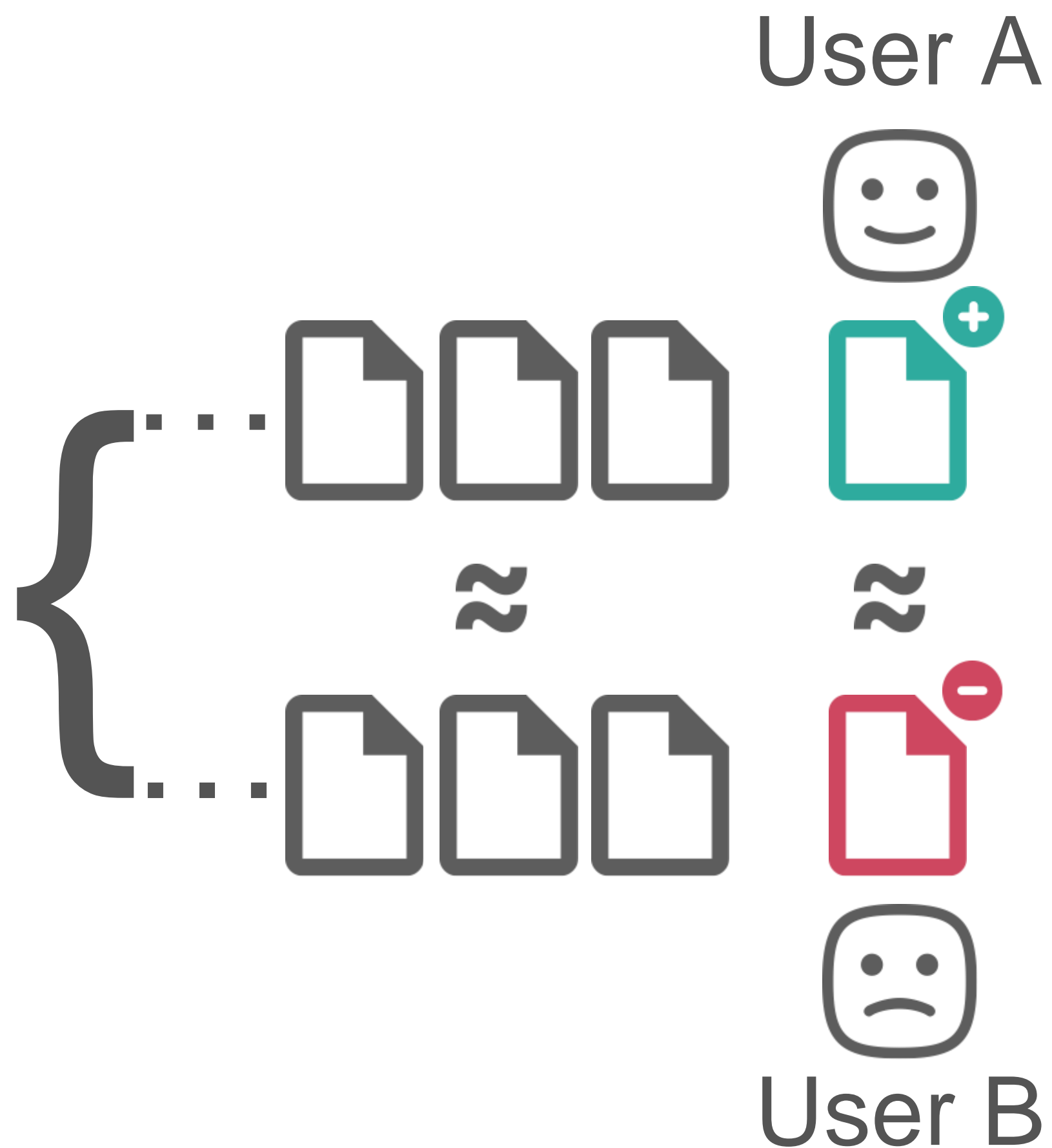
i.e. down-voting
because the community
dislikes the author

User A			Positively Evaluated
User B			Negatively Evaluated



- ✦ Text quality determined by training a machine learning model using text features, validated using the crowd.

Similar History
Number of posts,
overall proportion
of up-votes, etc.



User A



User B

What are the textual effects?

After a **positive evaluation**,
do users write better?



Or do they write worse?



What are the textual effects?

After a **negative evaluation**,
do users write better?



Or do they write worse?



What are the textual effects?

Negativity Bias

Post quality drops
significantly after a negative
evaluation, but not after a
positive evaluation.

($p < 0.05$ in all communities)

Computing community bias



Actual Evaluation $P/(P+N)$	0.9
-----------------------------	-----

Text Quality	0.8
--------------	-----

Community Bias	$0.9 - 0.8$ $= +0.1$
----------------	-------------------------

User A



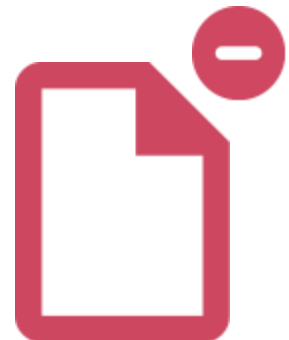
...



...

≈

≈



...



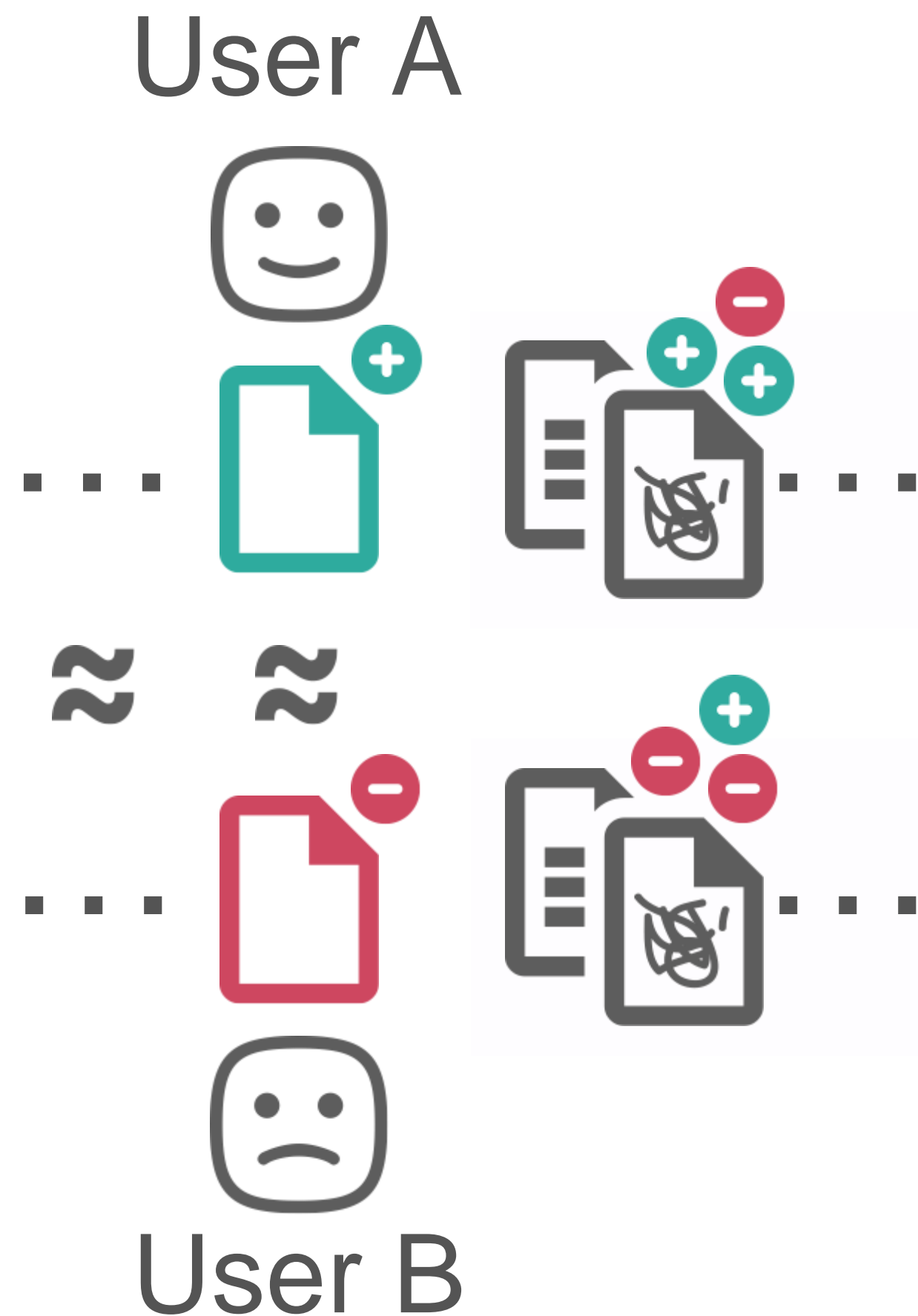
...



User B

} Compare
Community Bias
After

What are the community effects?



Does the community perceive a user worse after a negative, than a positive evaluation?

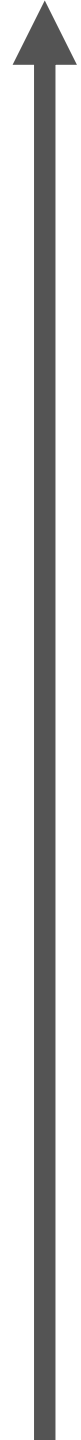
What are the community effects?

Halo Effect

Posts made after a negative evaluation were perceived worse than those made after a positive evaluation.

($p < 0.05$ in all communities)

More Positive

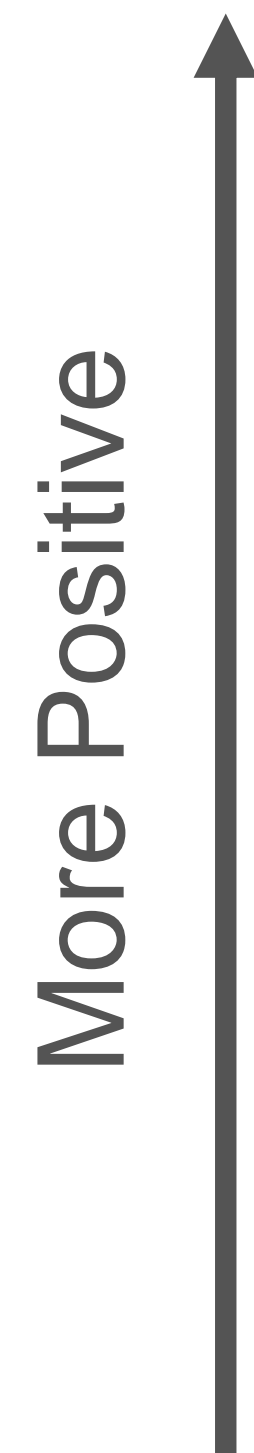


Positive Eval.



Negative Eval.





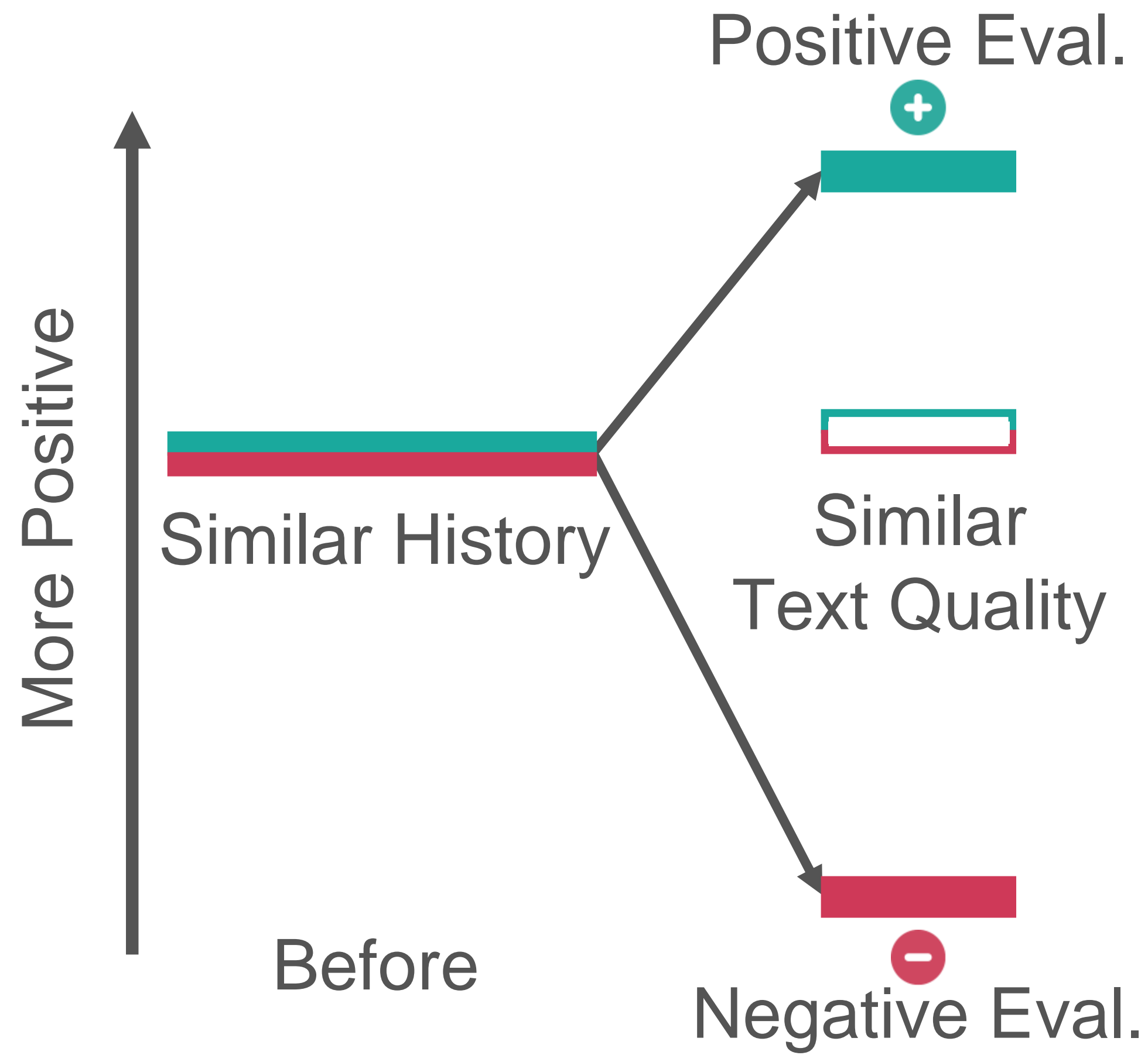
Positive Eval.

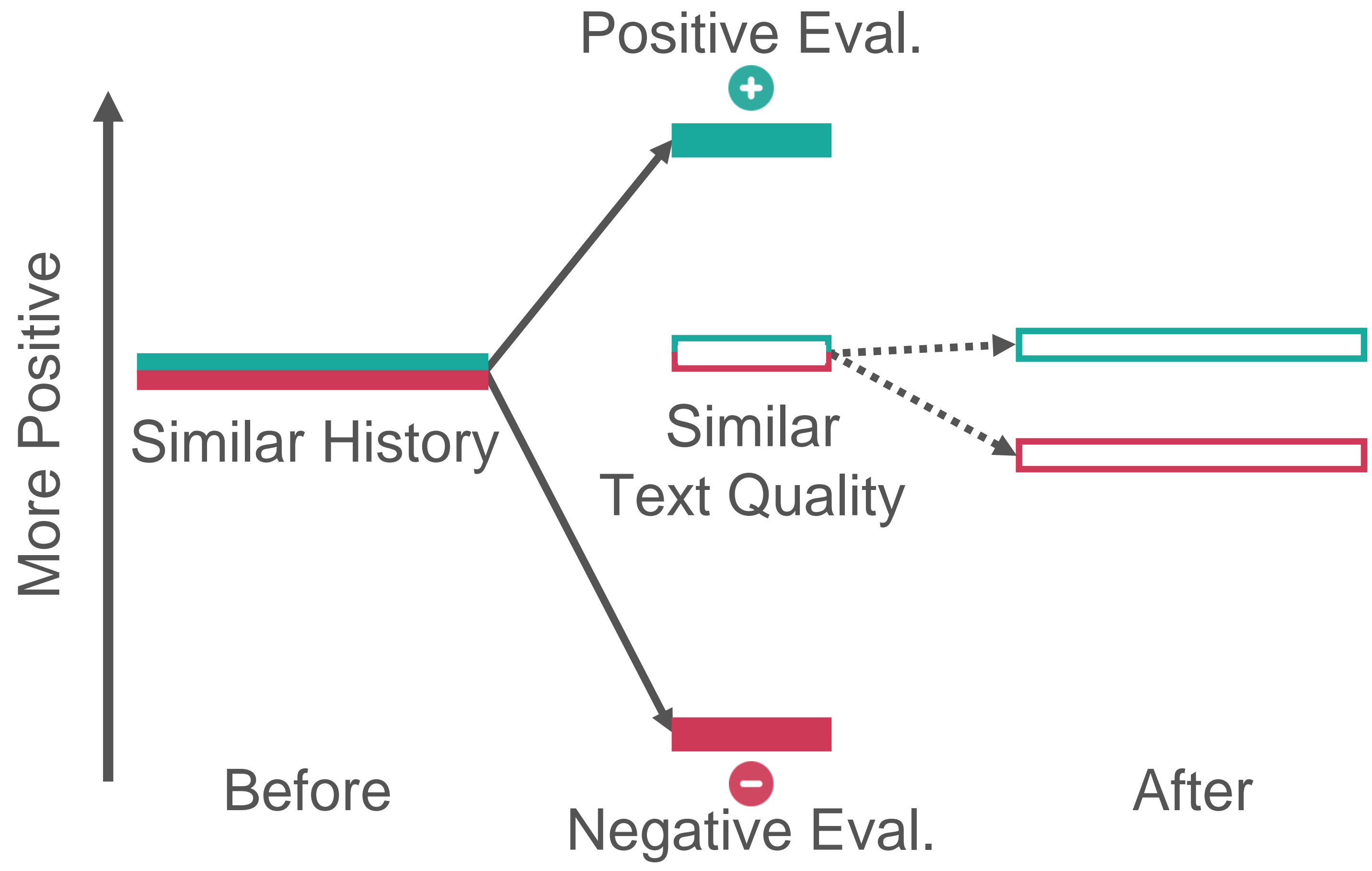


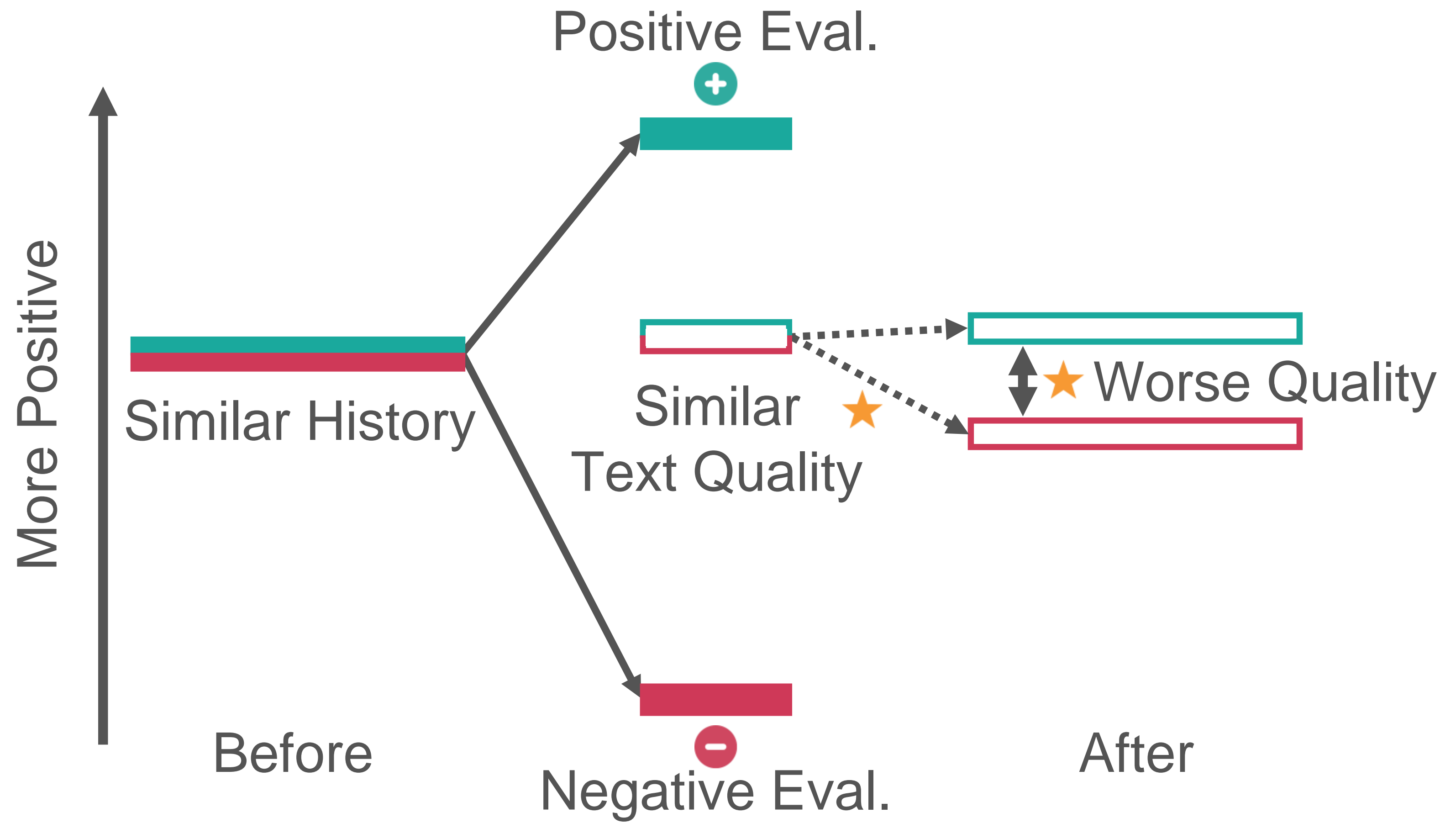
Similar
Text Quality

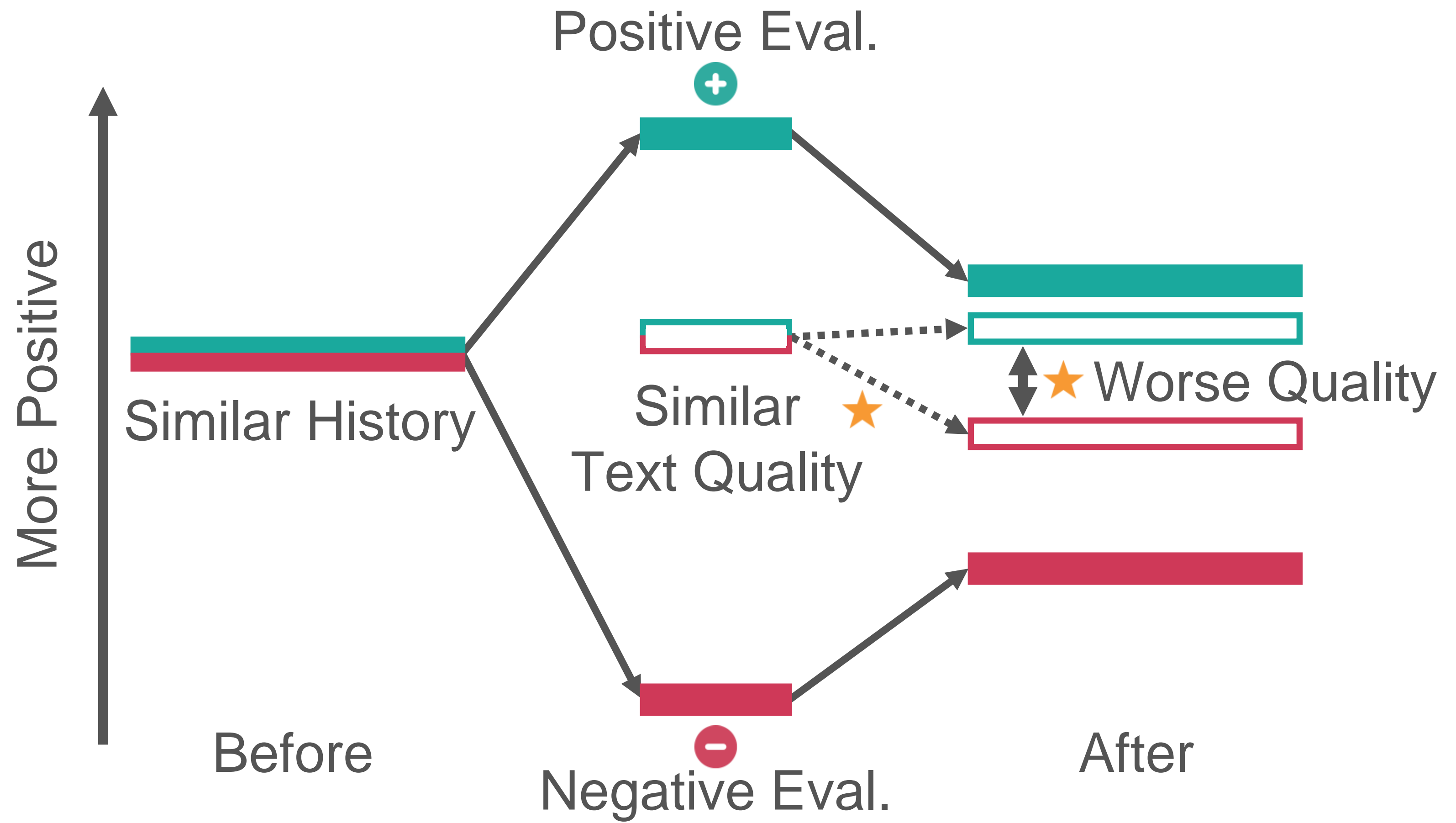


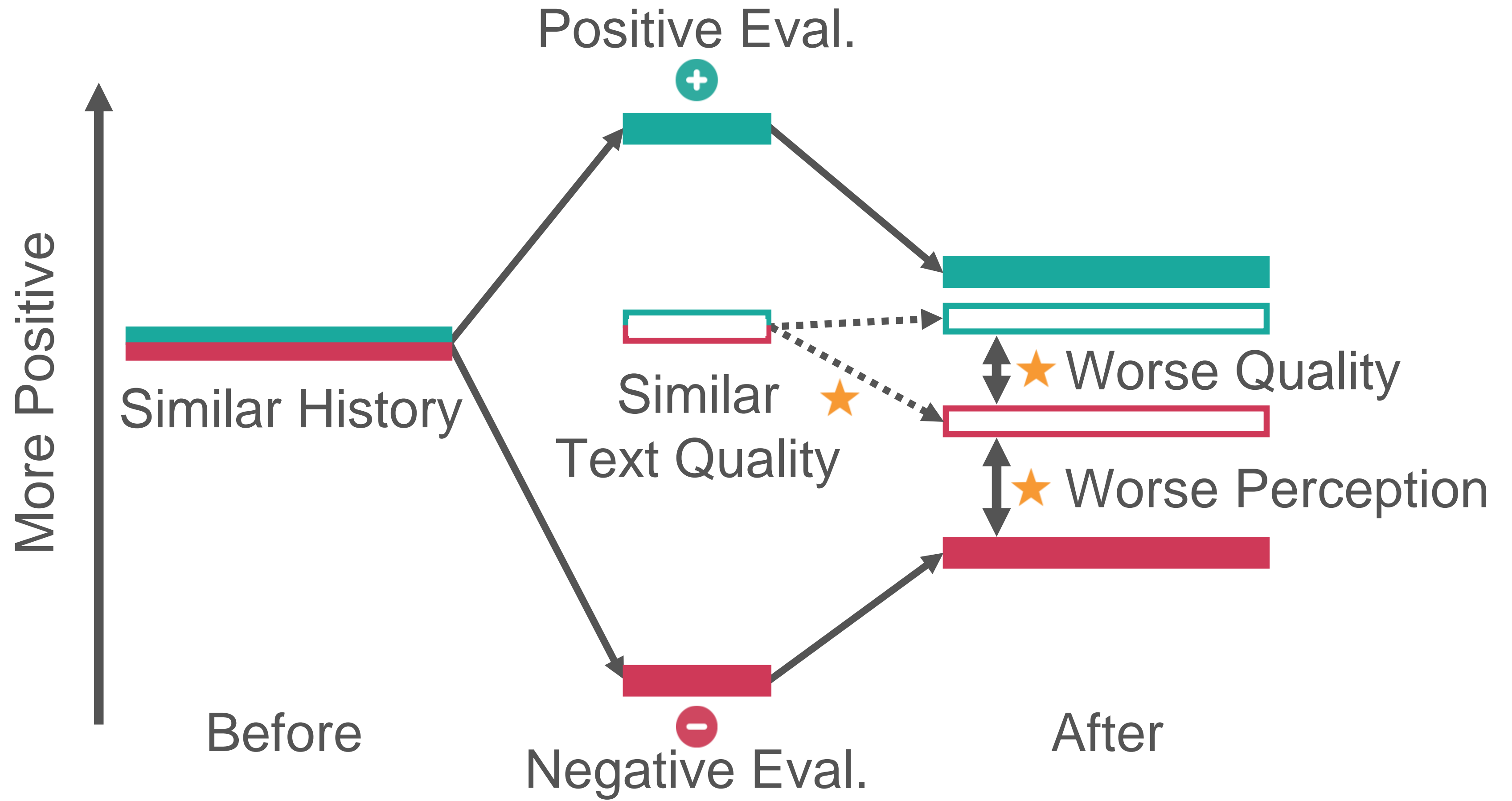
Negative Eval.





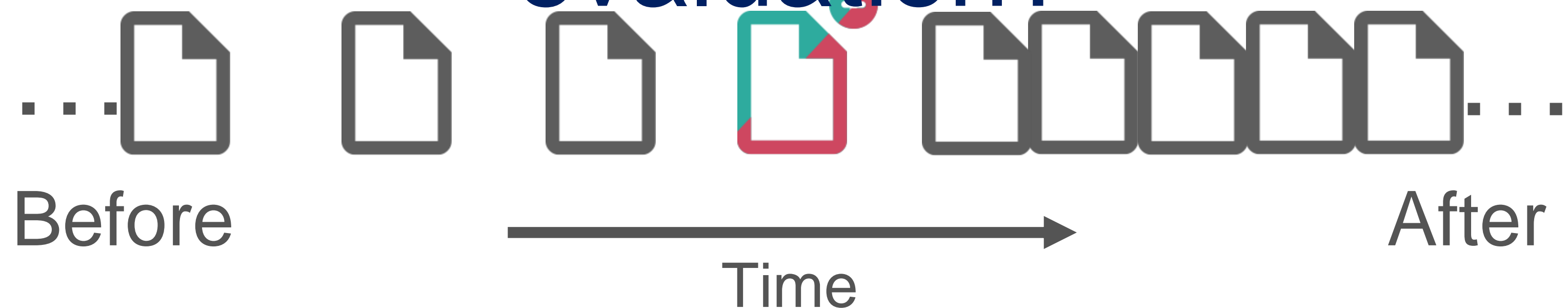




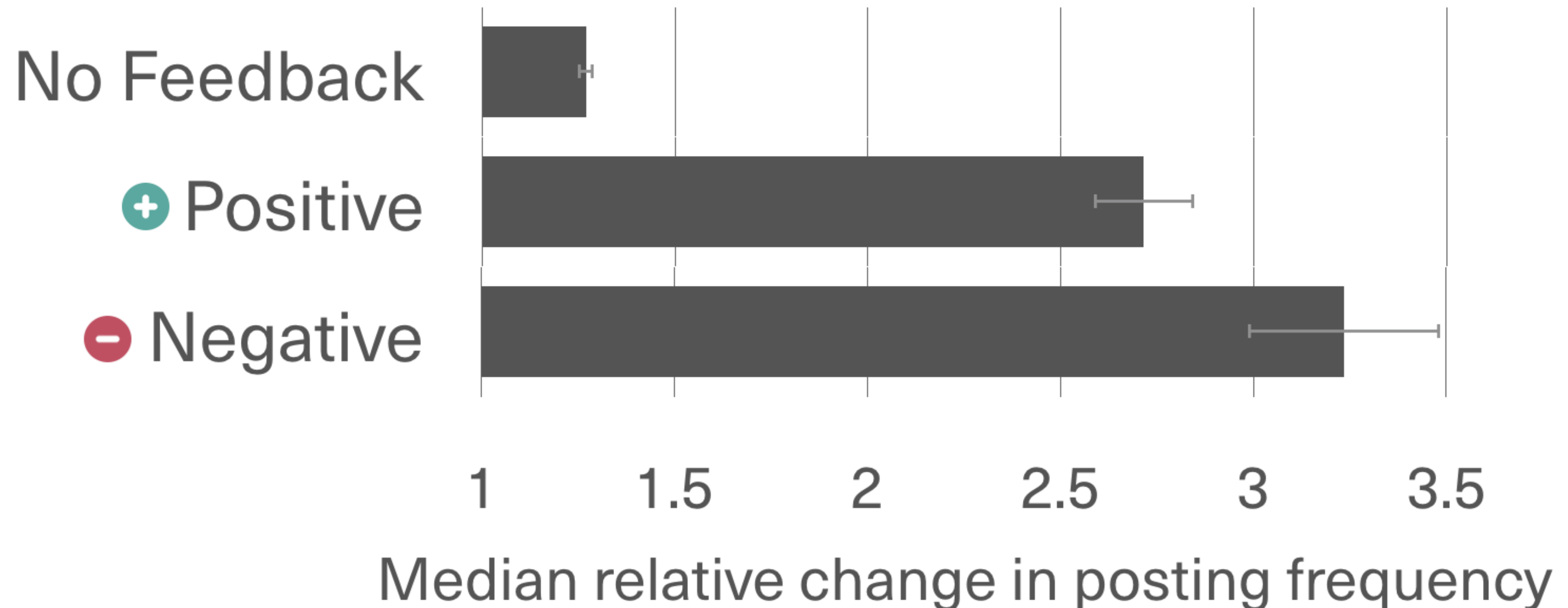


Does feedback
regulate posting activity?

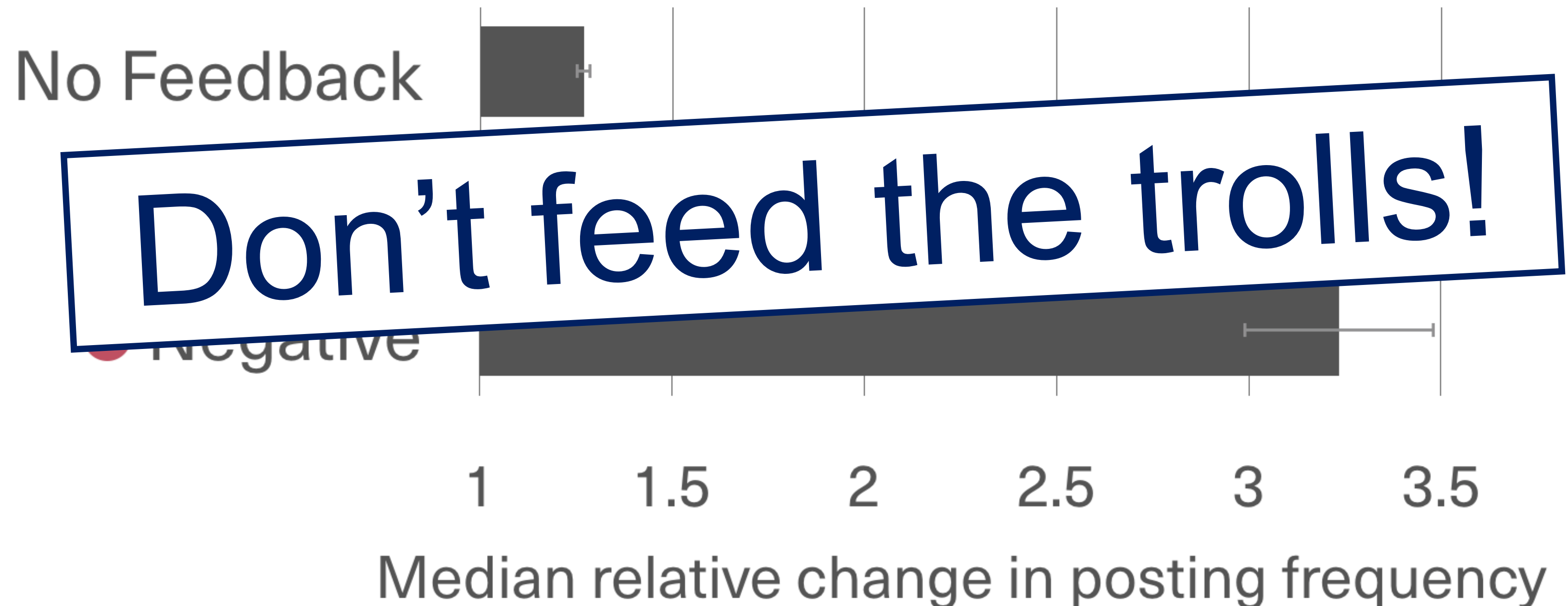
Do users post more frequently
after a positive/negative
evaluation?



Users who receive negative feedback post more frequently

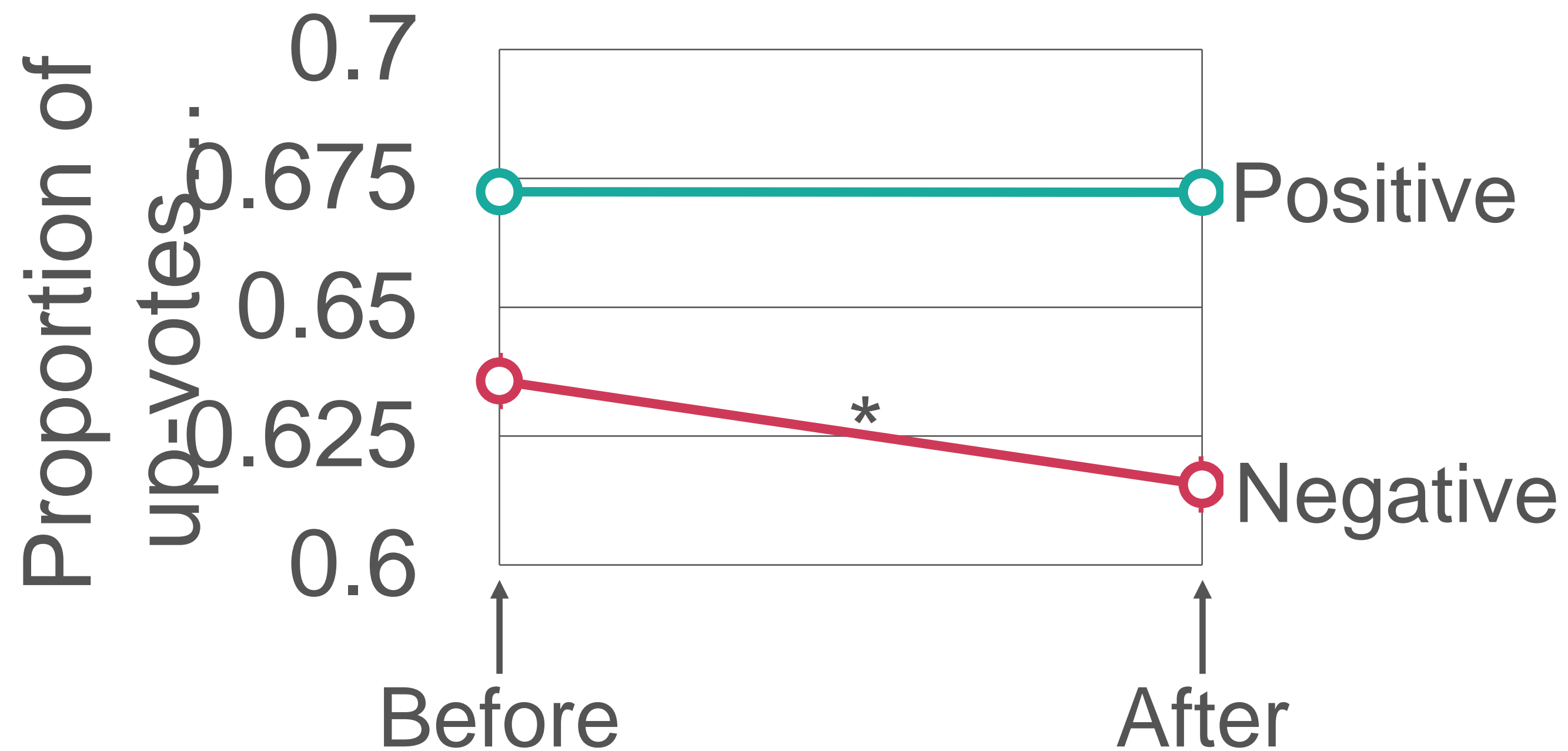


Users who receive negative feedback
post more frequently



Does feedback result in subsequent
backlash?

Negatively-evaluated users evaluate others worse in the future



Trolls write worse over time

Trolls start out writing worse, and
worsen more over time.

($p < 0.05$ in all communities)

Communities become less tolerant

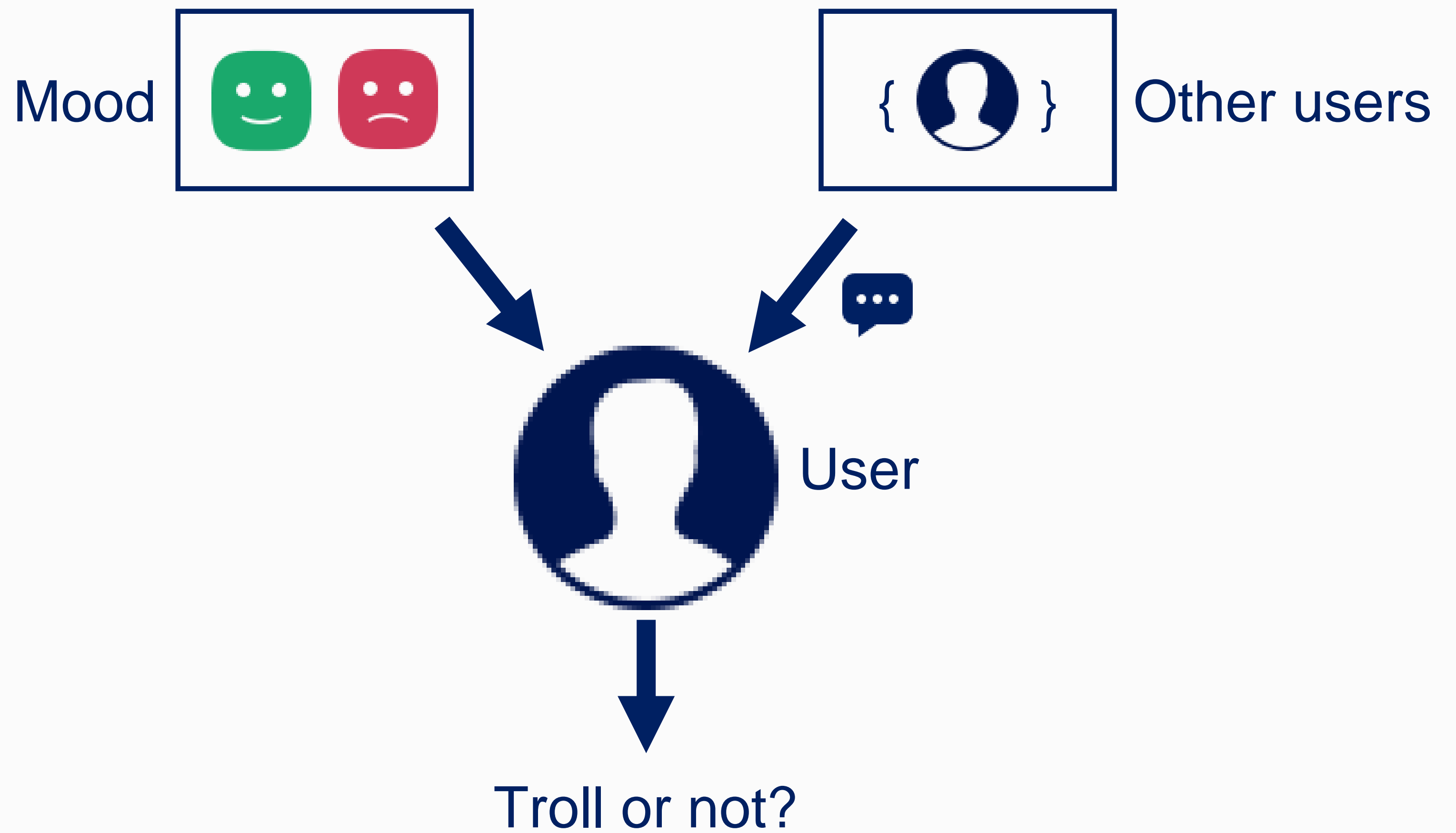
A troll's posts are more likely to
be deleted later in their life.

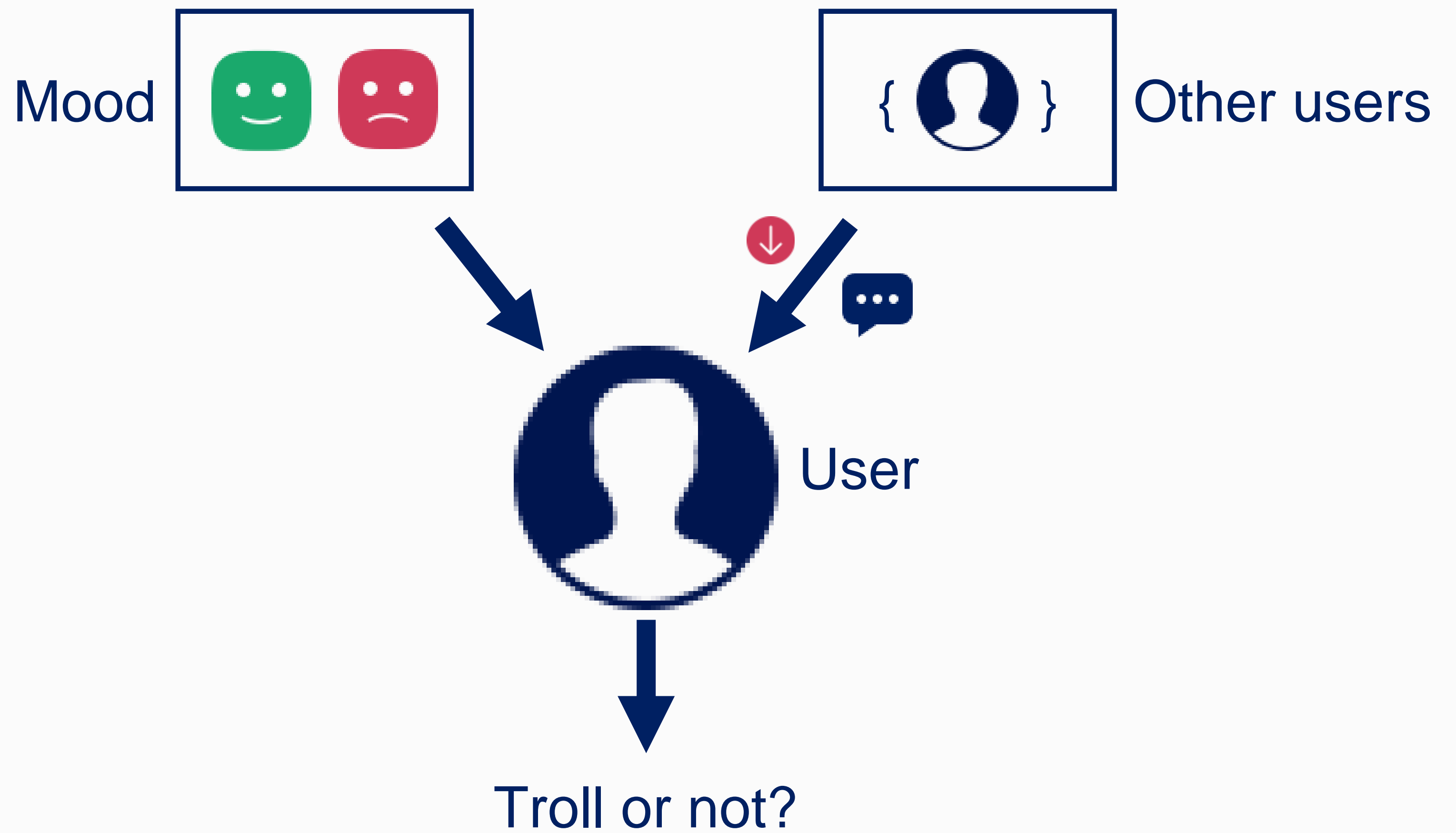
($p < 0.05$ in all communities)

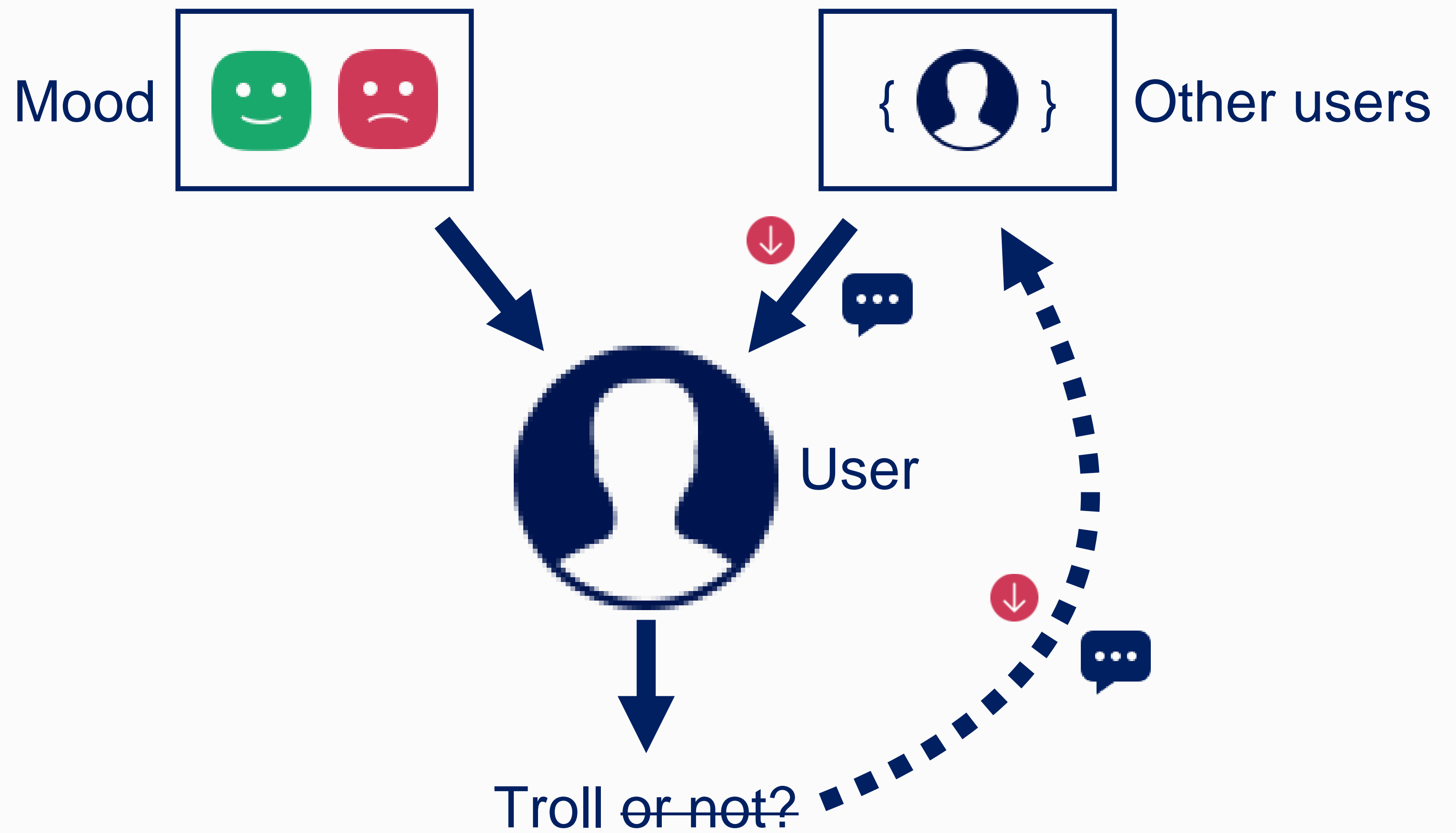
Communities exacerbate trolling

Unfairly deleting a user's posts
causes them to write worse later.

($p < 0.05$ in all communities)

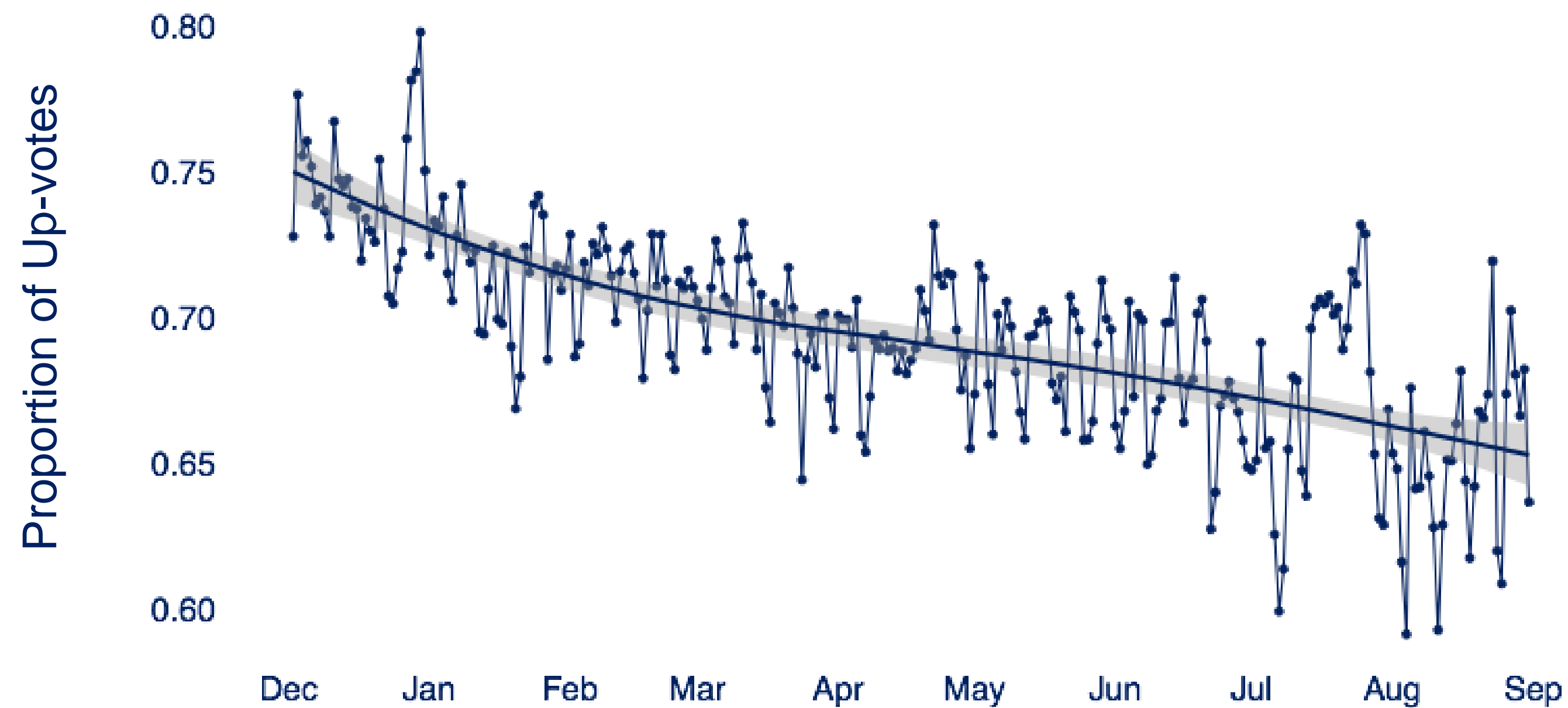




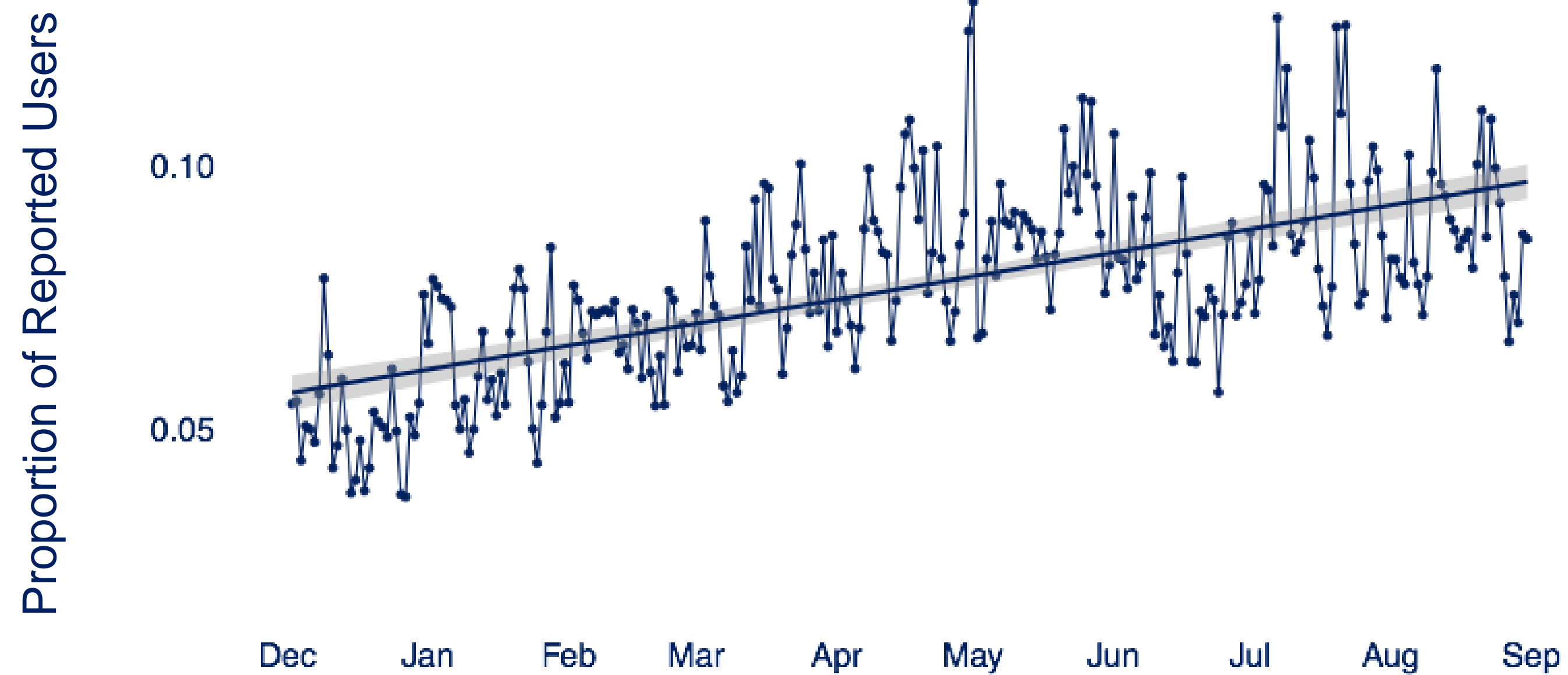


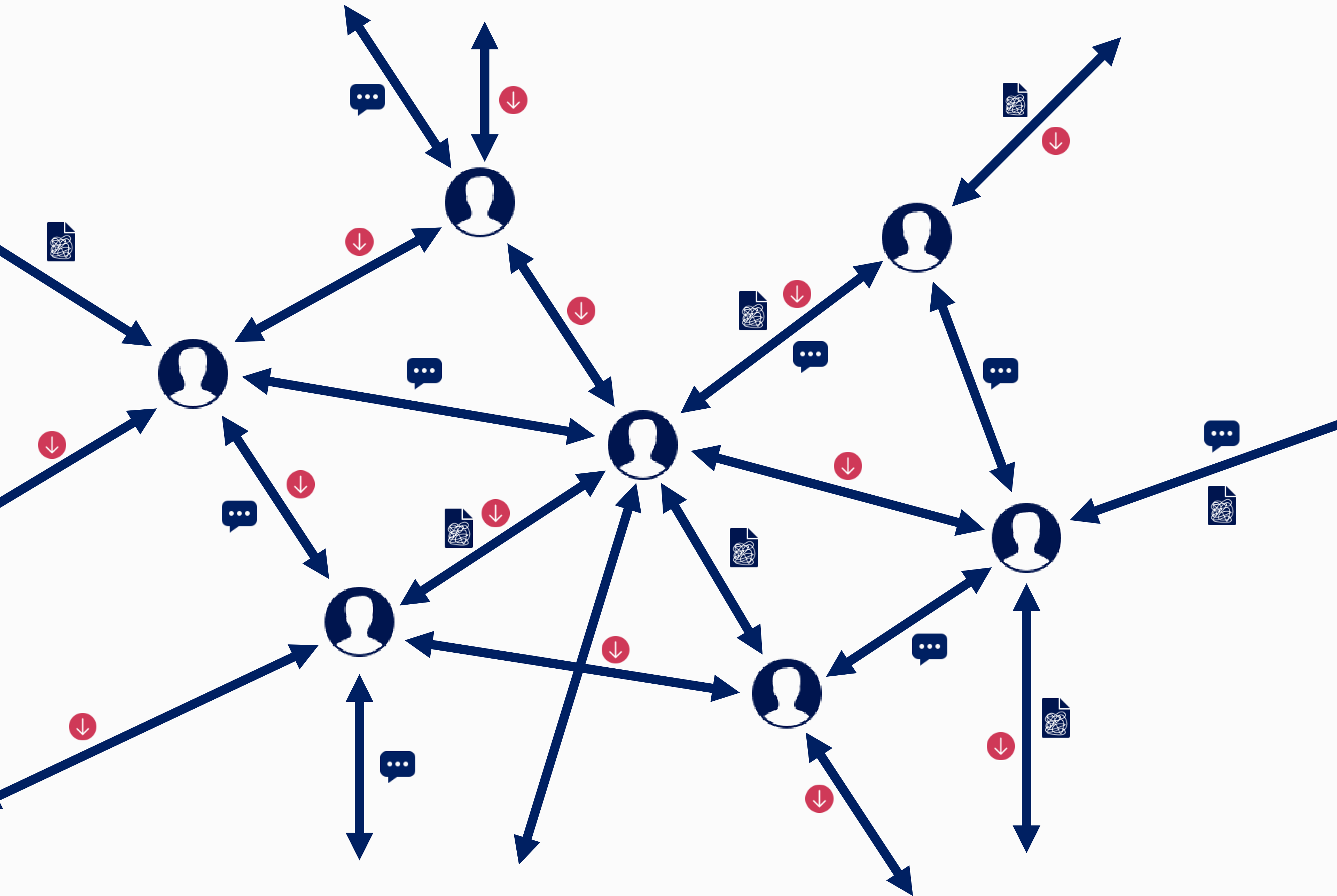
Is there a downward spiral?

Communities worsen over time



Communities worsen over time





What can we do?

Predictive Policing?

Can we predict whether a user
will get banned in the future?
(Using only first 10 posts)

Prediction Features



Post
of words
...



Activity
posts/day
...



Community
% upvoted
...

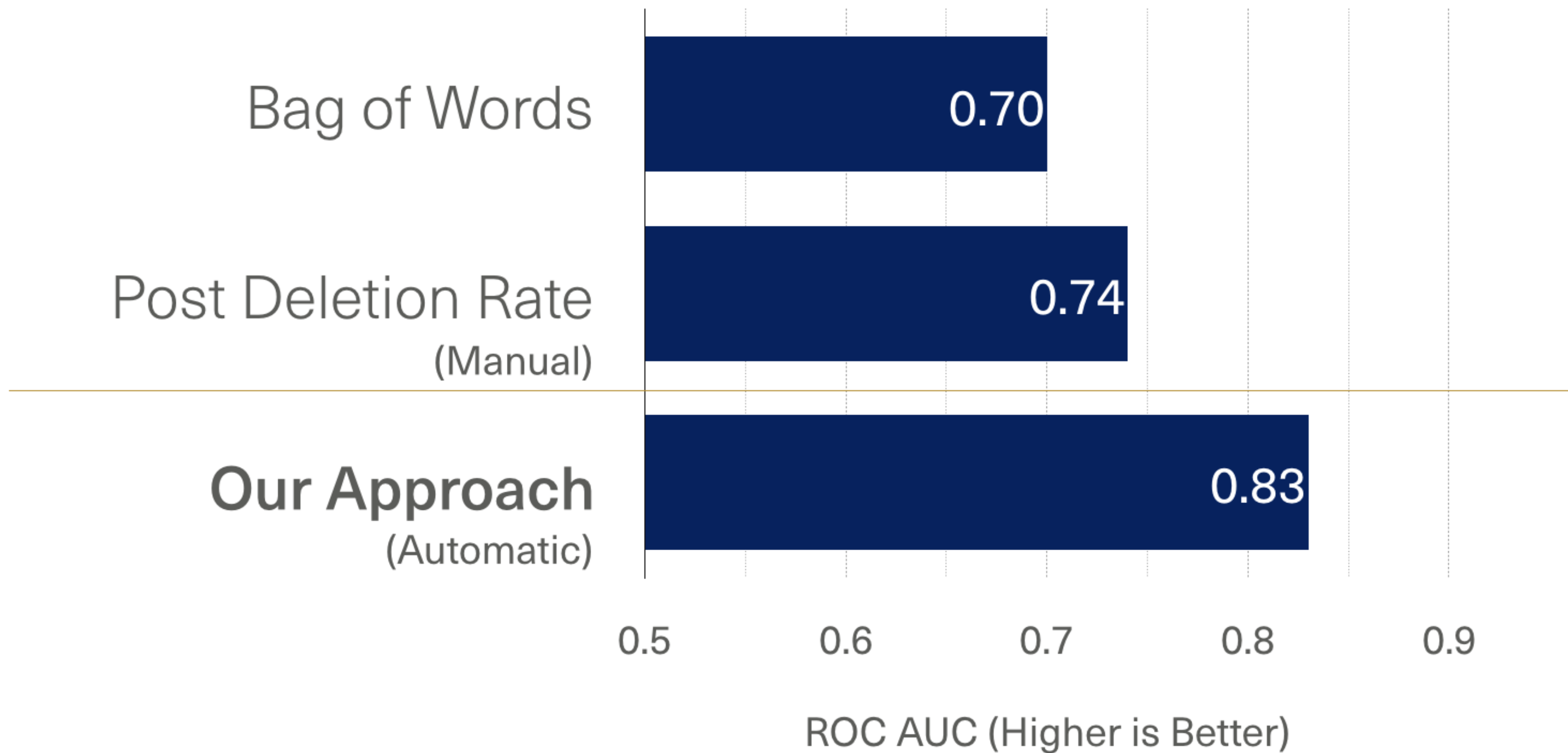


Moderator
Deletions
...

Automatic

Manual

Predictive Policing?



Predictive Policing?

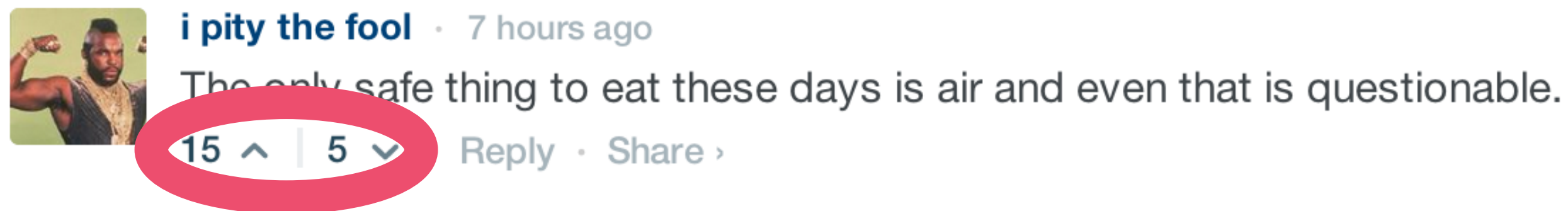
Our automatic approach
generalizes across communities.
(Cross-domain AUC = 0.68)

Predictive Policing?

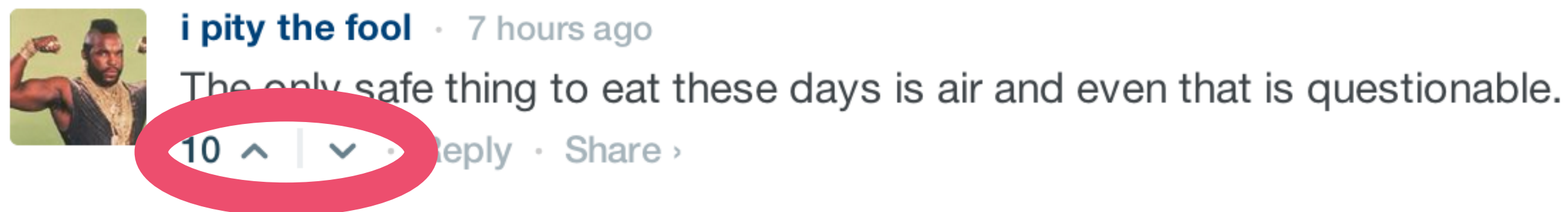
Our automatic approach
generalizes across communities.
(Uses interaction patterns, not language)

Discouraging Negative Feedback?

Before



After



Discouraging Negative Feedback?



2733



When I make comments too quickly together, and Reddit says "You're doing that too much. Please wait 5 minutes."

(i.imgur.com)

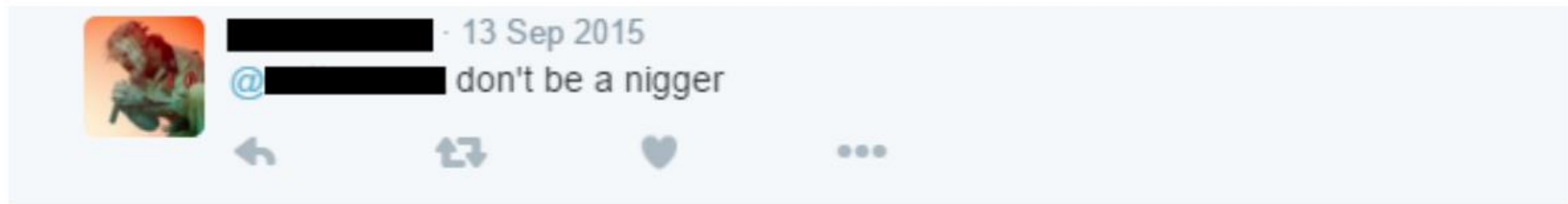
submitted 1 year ago by [seemedlikeagoodplan](#)

609 comments [share](#)



You're replying too quickly. Please wait 22 seconds before trying again.

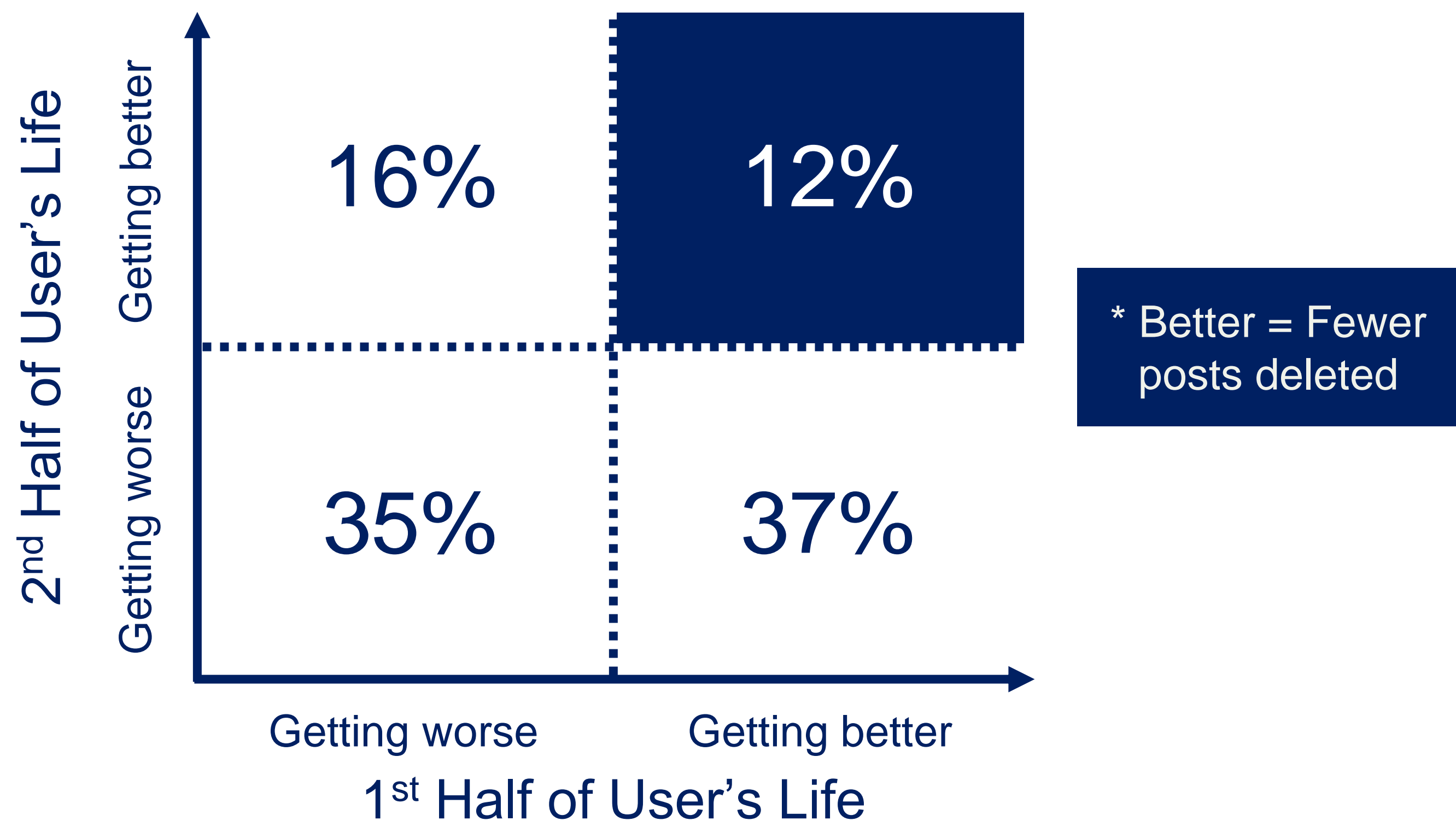
Bots to reduce harassment?



@ [redacted] Hey man, just remember that there are real people who are hurt when you harass them with that kind of language

“Subjects who were sanctioned by a high-status white male significantly reduced their use of a racist slur”

How can we help users that are trying to improve?



Summary

What we thought

Trolls are a
vocal minority

Trolling is innate

Trolls must be identified
manually

What we now know

Trolls can be
ordinary people

Trolling can spiral from
a single bad post

Trolls and troll posts can be
automatically identified

Trollish references

- Baker, P. (2001). Moral panic and alternative identity construction in Usenet. *J Comput-Mediat Comm.*
- Buckels, E. E.; Trapnell, P. D.; and Paulhus, D. L. (2014). Trolls just want to have fun. *Pers Individ Differ.*
- Cheng, J., Bernstein, M.S., Danescu-Niculescu-Mizil, C., & Leskovec, J. 2017. **Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions.** CSCW 2017. <http://bit.ly/anyonepaper>
- Cheng, J., Danescu-Niculescu-Mizil, C. & Leskovec, J. (2015). **Antisocial Behavior in Online Discussion Communities.** ICWSM 2015. <http://bit.ly/trolls-paper>
- Cheng, J., Danescu-Niculescu-Mizil, C. & Leskovec, J. (2014). **How Community Feedback Shapes User Behavior.** ICWSM 2014. <http://bit.ly/feedback-paper>
- Donath, J. S. (1999). Identity and deception in the virtual community. *Communities in cyberspace.*
- Herring, S.; Job-Sluder, K.; Scheckler, R.; and Barab, S. (2011). Searching for safety online: Managing "trolling" in a feminist forum. *The Information Society.*
- Jones, J. W., & Bogat, G. A. (1978). Air pollution and human aggression. *Psychological Reports*, 43(3), 721-722.
- Munger, K. (2016). Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment. *Political Behavior*, 1-21.
- Kirman, B.; Lineham, C.; and Lawson, S. (2012). Exploring mis- chief and mayhem in social computing or: how we learned to stop worrying and love the trolls. In *CHI EA*.
- Rotton, J., & Frey, J. (1985). Air pollution, weather, and violent crimes: concomitant time-series analysis of archival data. *Journal of personality and social psychology*, 49(5), 1207.
- Shachaf, P., and Hara, N. (2010). Beyond vandalism: Wikipedia trolls. *J Inf Sci.*
- Wilson, J. Q., & Kelling, G. L. (1982). Broken windows. *Critical issues in policing: Contemporary readings*, 395-407. Chicago
- Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. In Nebraska symposium on motivation. University of Nebraska press.