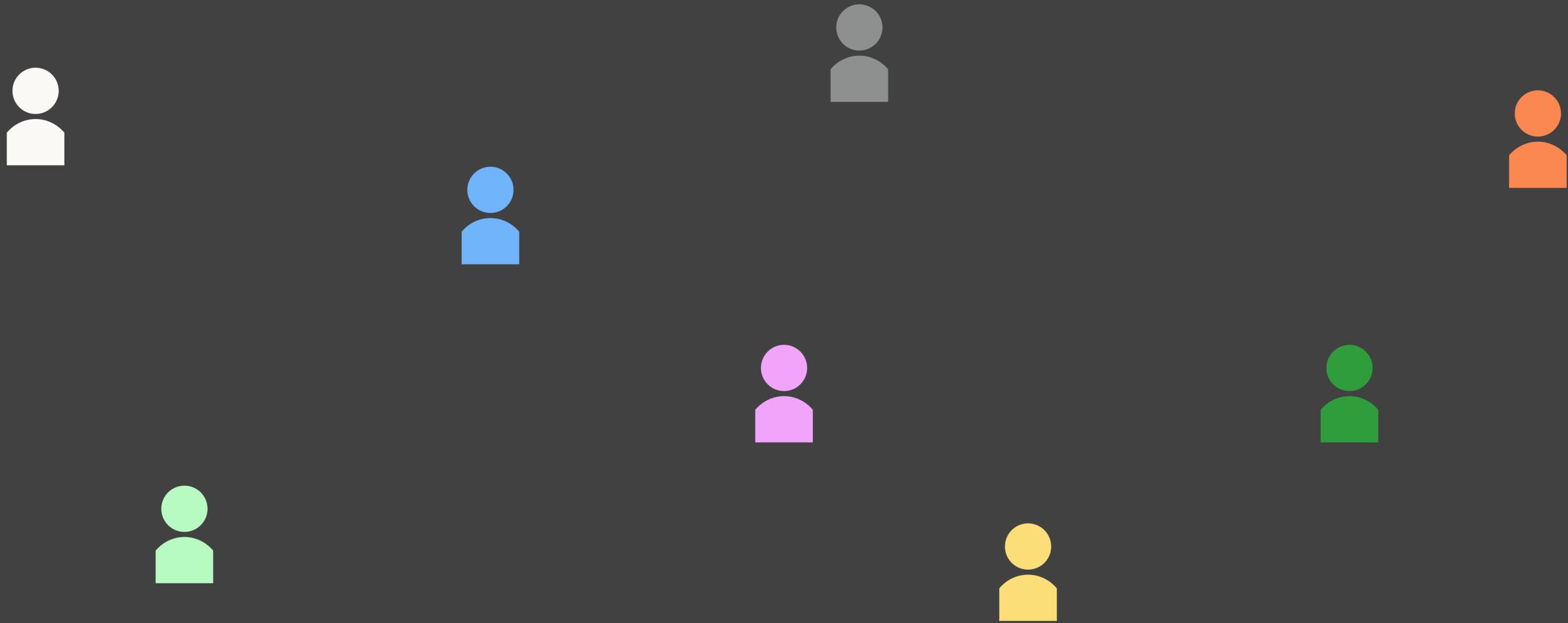




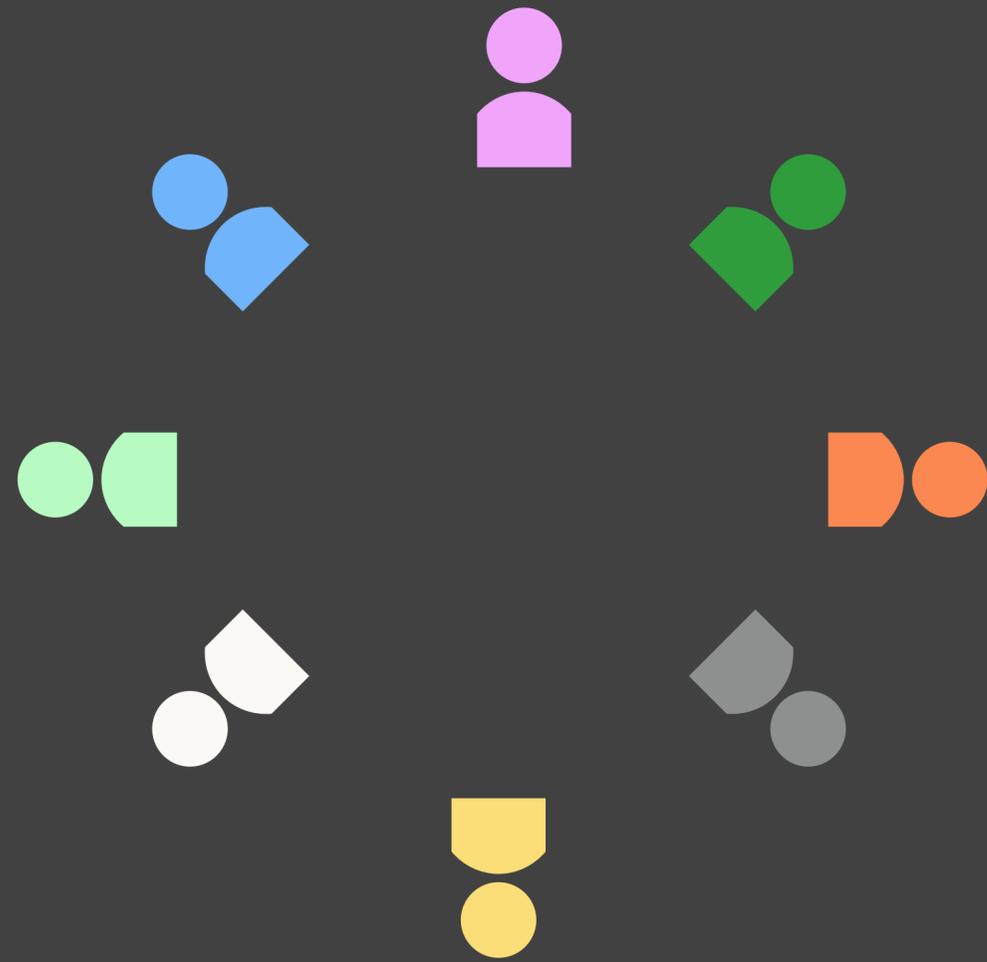
# Anti-Social Computing

CS 278 | Stanford University | Michael Bernstein

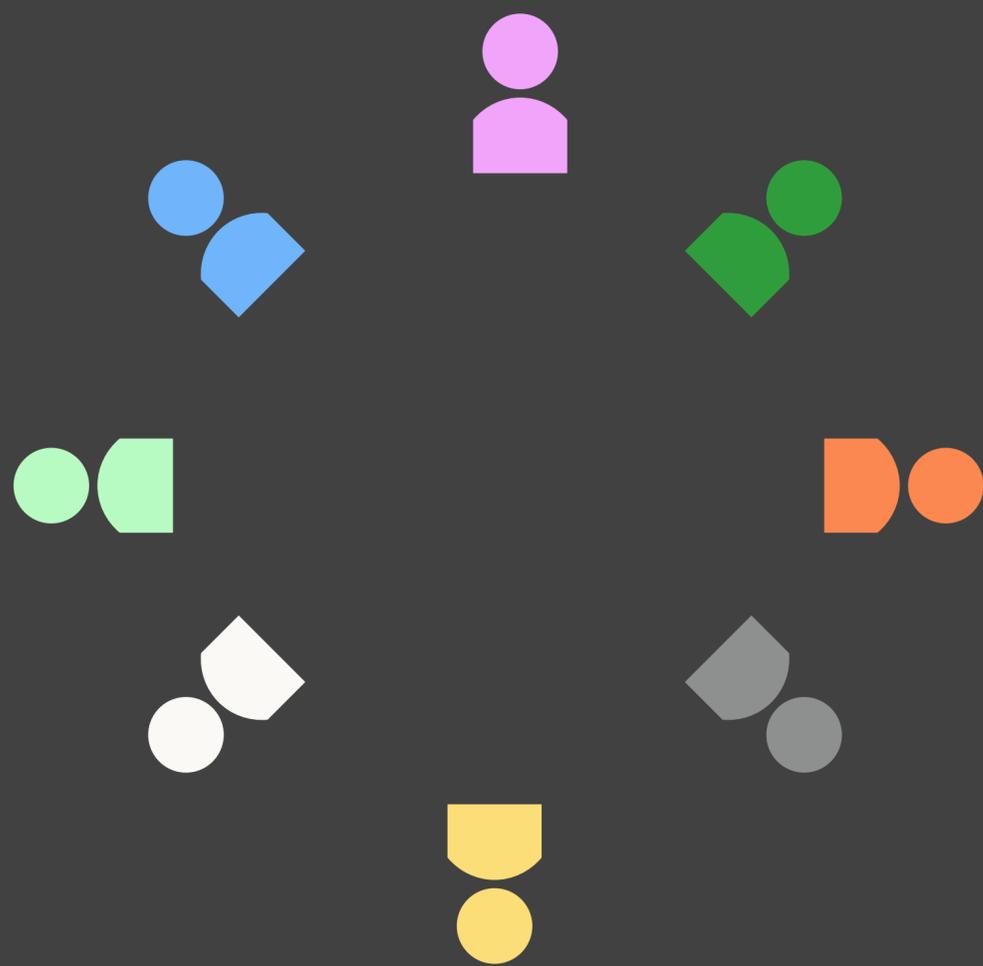
**We will dedicate time  
near the end of class  
today to forming teams.**



*Once upon a time, we were disconnected and decentralized.*



*So, we formed social computing systems and online communities to connect us to each other.*



*We met new friends, created net culture, and shared our ideas with the world. Life was happy.*



*But then, anti-social behavior arose. It grew and grew until it threatened to destroy the social computing systems.*



*The people were trolled and flamed. Their communities fractured. What could help?*

# Today: anti-social computing

How can a community manage anti-social behavior?



Topics:

- Premeditated anti-social behavior

- Non-premeditated anti-social behavior

- The darkest of the dark: beyond trolling

# Premeditated anti-social behavior

a.k.a., trolling 

# **How Trolls Are Ruining the Internet**

## **When Will the Internet Be Safe for Women?**

### **FAKE NEWS IS ABOUT TO GET EVEN SCARIER THAN YOU EVER DREAMED**

Examples via Justin Cheng [Time 2016;The Atlantic 2016;Vanity Fair 2017]

A pie chart with two segments: a larger light gray segment on the left and a smaller reddish-pink segment on the right. The reddish-pink segment represents 47% of the total.

47% of online users  
have been harassed

Examples via Justin Cheng [Data & Society 2017]

# **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**

# What is trolling?

Intentional disruption of an online community [Schwarz, NYT 2008]

Behavior that falls outside the acceptable bounds of the community [Binns 2012; Hardaker 2010]

People who habitually engage in trolling are known as **trolls**, as in the grumpy monsters who hide under bridges.

# Is trolling worse online?

It's certainly more well-publicized. There are two reasons we might run into it more online than offline:

- (1) Scale: a single troll can impact many communities, or a single highly visible community, in ways that their reach would otherwise be limited.
- (2) People troll more online than they do offline.

Are these true?

# Online disinhibition effect

[Suler 2004]

When we interact online, we say and do things that we would not do offline and in-person. We self-disclose more, and we act out more.

This is known as the **online disinhibition effect**: we have less inhibition when online.

Online disinhibition would imply that we do troll more online than offline.

# Dealing with disinhibition

Reasons include:

Anonymity: dissociation from my real identity, so fewer consequences

Few social cues: no facial expressions, reactions, etc.; a socially opaque system

Asynchronicity: conversations never cool off

So, design interventions might be:

Re-individuate by associating actions with an identity that I care about

Re-introduce social cues: e.g., when I reply to a mean comment, they often soften up

Take it offline: don't try to manage fights online if possible

# Blocking

If a few people are responsible for the majority of the deleterious content, then blocking them should silence most of the negative behavior.

However, blocking just becomes whack-a-mole if it's easy for participants to create another account. So, this only works if identities are expensive to create.

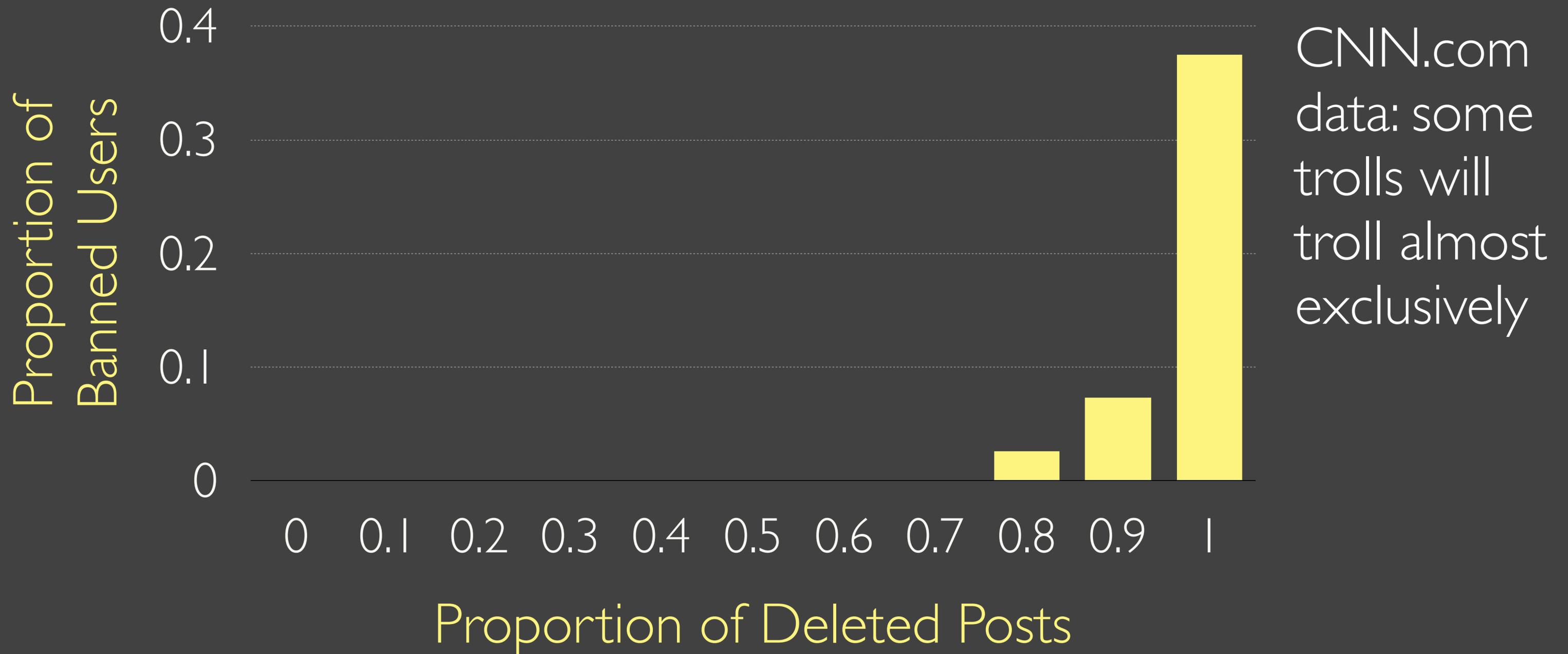
# Why do trolls troll?

One argument: trolls are born that way.

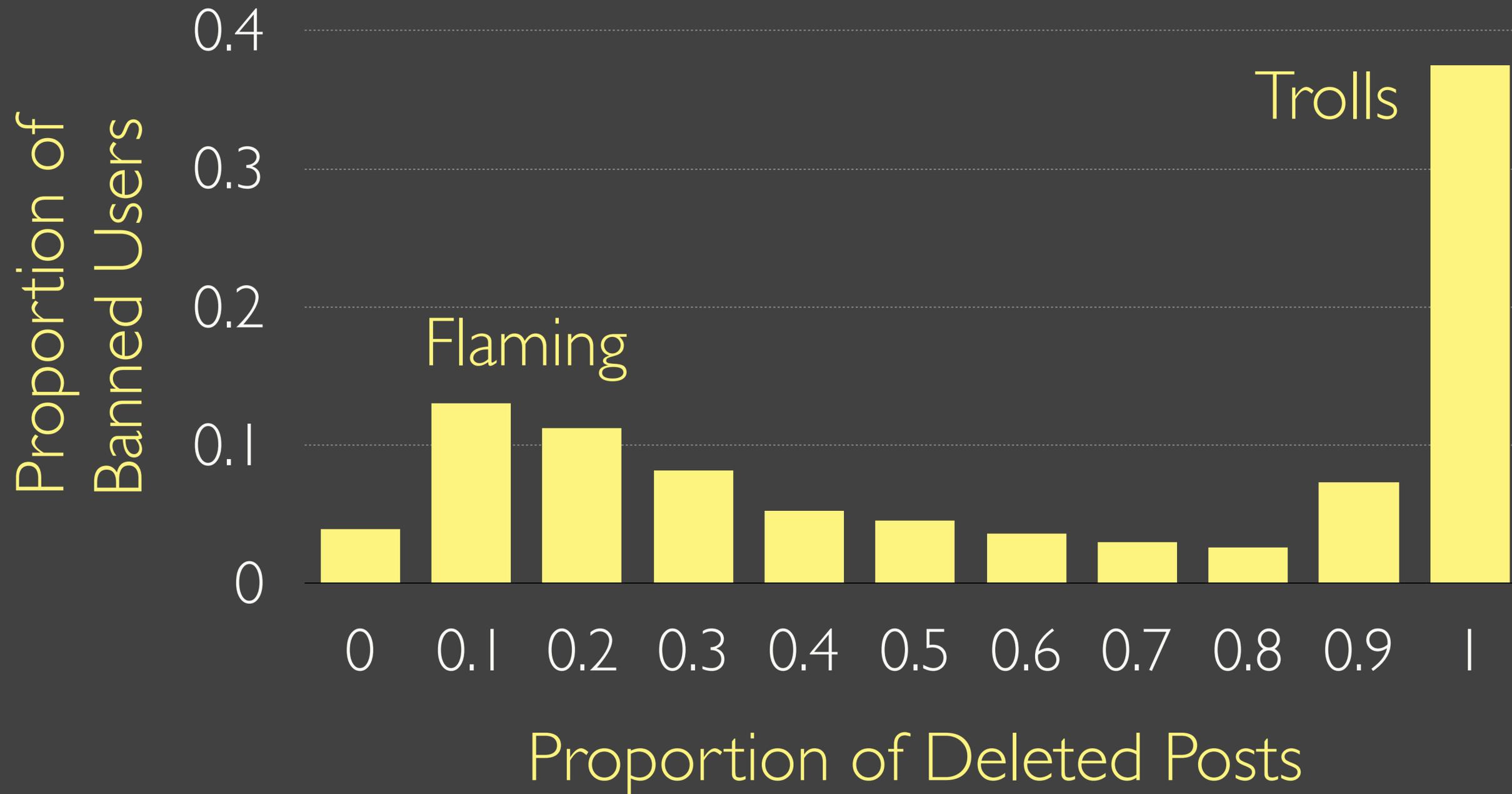
Inveterate trolls do, on average, register strong personality dispositions such as high self-report scores in the Dark Tetrad of personality traits: **sadism**, psychopathy, and Machiavellianism [Buckels, Trapnell, and Paulhus 2014]

Reasons given range from boredom [Varjas et al. 2010], to doing it for fun [Shachaf and Hara 2010], to venting [Lee and Kim 2015].

# How much do trolls troll?



# How much do trolls troll?



CNN.com data: some trolls will troll almost exclusively... and some only rarely.

Why?

# Unpremeditated anti-social behavior

a.k.a., flaming 

# What is flaming?

Flaming: uninhibited hostile behavior directed at another person or group [Kayany 1998, Kiesler 1986]

Common examples: swearing, calling names, ridiculing, insulting

While trolling usually refers to someone intentionally riling people up, flaming usually refers to someone who lost self-control.

# Trolling and disinhibition

Online disinhibition would also predict that we flame more online than in person. We are more likely to act out, others are more likely to act out in response, and so on: a flame war.

But does the environment suppress or amplify the amount of trolling that people engage in?

# Recall: environment matters

## 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

(Real comments)

## 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 | · [Reply](#)



**User9054** · 3 hours ago

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

Positive comments

Result: 35% troll comments

[Cheng et al. 2017]

Negative comments

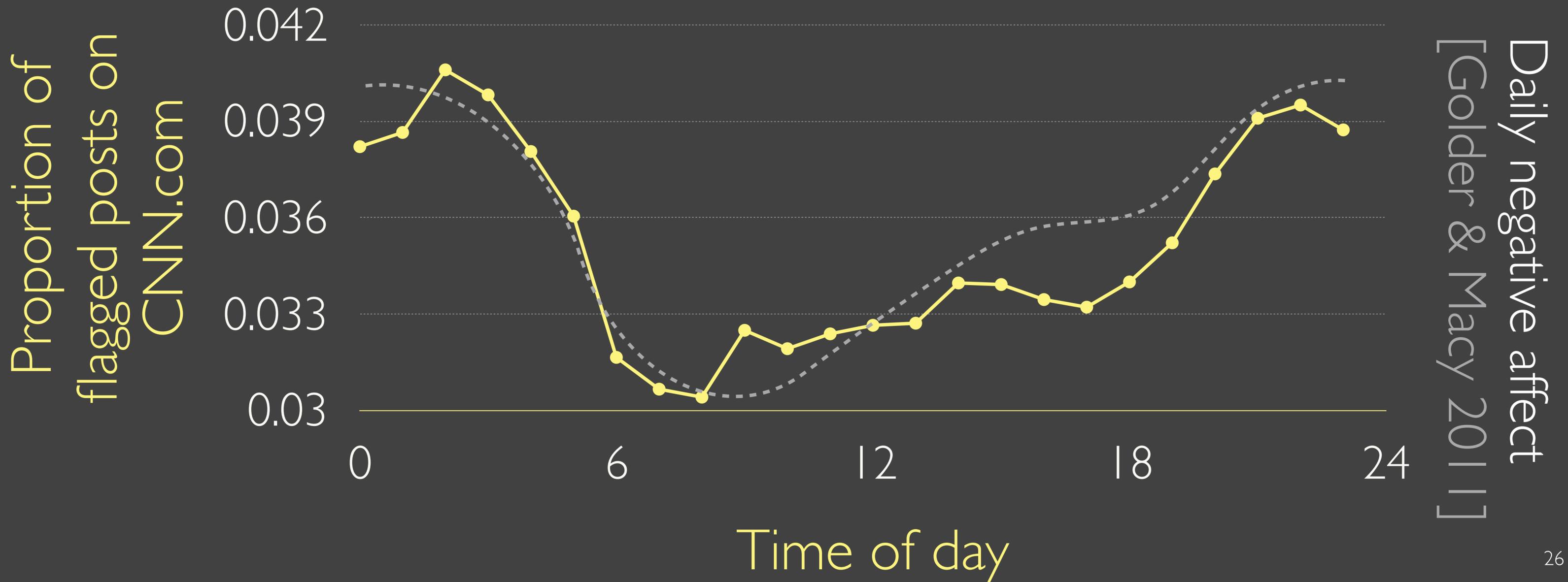
Result: 47% troll comments  
(Relative increase of one third compared to the 35% baseline)

# Mood

The environment is something the designer has control of. But people also bring their own affective state to a social computing system.

Being in a bad mood reduces self-regulation [Leith and Baumeister 1996] and results in less favorable impressions of others [Forgas and Bower 1987].

# Antisocial behavior tracks human diurnal mood patterns



# Mood influences behavior [Cheng et al. 2017]

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**

Placed in a good mood by doing well on an easy test

Result: 35% troll comments

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**

Placed in a negative mood by doing poorly on a difficult test

Result: 49% troll comments

(Same effect as seeing troll comments!)

Positive Norm  
Negative Norm

# Positive Mood

Unscramble the following letters to form an English word:

"P A P H Y"

Type in your answer.



**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.

35% troll comments

# Negative Mood

Unscramble the following letters to form an English word:

"D E A N Y O N"

Type in your answer.



**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.

49% troll comments

## The effects compound.

Unscramble the following letters to form an English word:

"P A P H Y"

Type in your answer.



**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.

47% troll comments

Unscramble the following letters to form an English word:

"D E A N Y O N"

Type in your answer.



**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.

68% troll comments

# Reducing flaming

Assuming the environment and norms aren't changeable, then one possible mechanism is to manage mood. Moods pass, so consider a cool-off period before posting a flame post.



So now the only reason to vote is gender. What a perfect way of proving you are intelligent. LOL.

Your reply will be automatically posted in 5 minutes.

[Undo?](#)

Example from Justin Cheng

# Face-saving

[Kiesler et al. 2012]

People self-regulate if they can do so without having to admit that they deliberately violated norms. MIT's warning email to students:

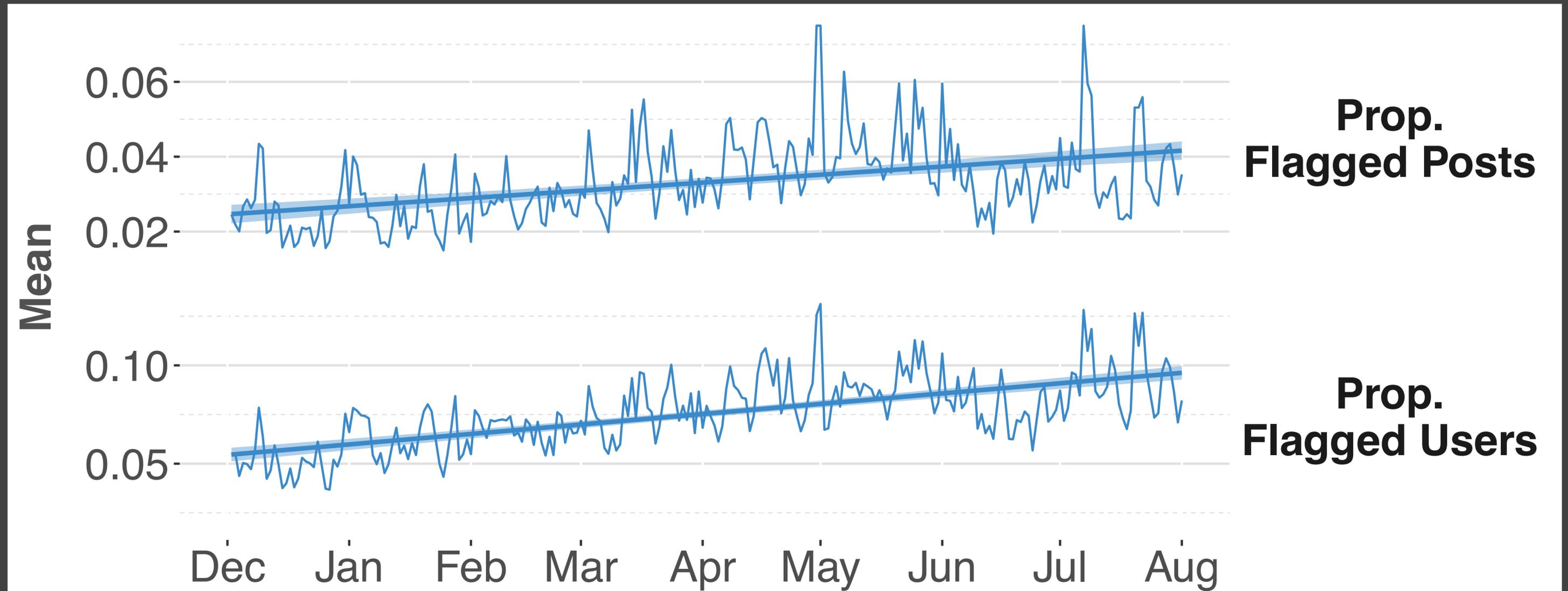
Someone using your account did [whatever the offense is]. Account holders are responsible for the use of their accounts. If you were unaware that your account was being used in this way, it may have been compromised. User Accounts can help you change your password and re-secure your account.

Many would change their password and the practice would stop, even if MIT knew from eyewitnesses that they had done it.

Calling them out instead prompted people to assert the behavior as within their rights and continue doing it to challenge authority.

**System-wide effects**

# Does trolling worsen over time?

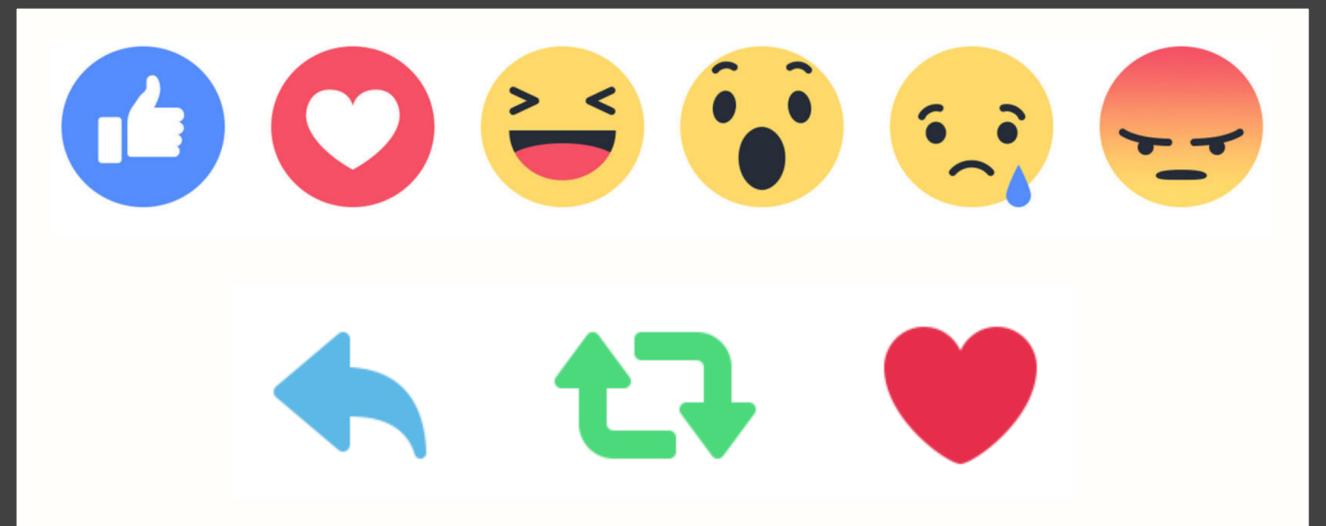


CNN.com

[Cheng et al. 2017]

# What does the design encourage?

Michael's take: systems that reward short-term engagement are likely to produce snark and flame, since these activities are the most likely ones to raise an affective response.



[Image from Niloufar Salehi]

# Internet mobs

Accusations can cascade, and without due process, the wrong people can be harmed.

Ex: Reddit misidentified Sunil Tripathi as the Boston Marathon Bomber

Misinformation spreads: the Boston Bomber rumors were corrected, but the corrections spread too slowly. [Starbird et al. 2014]



The image shows a screenshot of the BBC News website. At the top, the BBC logo is on the left, followed by navigation links for News, Sport, Weather, and More. A search bar is on the right. Below the navigation is a red banner with the word 'NEWS' in white. Underneath the banner is a secondary navigation bar with links for Home, Video, World, UK, Business, Tech, Science, and Stories. The main headline is 'Boston bombing: How internet detectives got it very wrong' in bold black text. Below the headline, it says 'By Dave Lee' and 'Technology reporter, BBC News'. The date '19 April 2013' is displayed at the bottom of the article header.



**Final project**

# Create a social computing system

Teams of three

Create and launch a new social computing system of **substantial design complexity**: it must represent a new social computing design that is novel, interesting, and thoughtful.

# Different routes to success

Socially challenging  
novel design; social bricolage;  
substantial behavior and dynamics

Choose based on your  
team's strengths. Be sure  
to articulate your goal.



# Stages

Proposal May 17

Form teams, develop a design direction, create a prototyping plan

In-class presentation June 5

One minute to share the project with the rest of the class

Project does not need to be complete

Final report June 11

2,500 words

# Time to talk teams

Closing part of the lecture afterwards. Final section **trigger warning**: rape, cyberbullying, doxxing, partner abuse. Please feel free to leave if you prefer not to discuss these topics.

**The darkest of the dark**

# Anti-social behavior gets (even more) personal

We typically think of anti-social behavior as hurling insults at each other, but this behavior escalates.

In this final section, we'll survey some of the most troubling behaviors and what we know about them. You may already have heard of some of these behaviors.

# A Rape in Cyberspace

[Dibbell 1993 in The Village Voice]

In LambdaMOO, a text-based online space (a la a textual MMO), a character named Mr. Bungle developed a piece of software that allowed him to command other characters to perform actions.

He then forced two other avatars to perform sexual acts on him and on each other, and to violate their own bodies.

Mr. Bungle was eventually banned from the server after a community meeting, but the damage had been done.

# Cyberbullying

Adolescents can execute their bullying of each other via social media, for example in public or private messages to each other. The effects are particularly problematic because they follow you home, not just at school. [Li 2006]

Women and LGBTQ are more likely to be victims, and perpetrators are more likely to be male. [Aboujaoude et al. 2015]

Being cyberbullies increases suicidal ideation and the probability of attempting suicide. [Hinduja and Patchin 2010]

# Doxing

Doxing refers to releasing the personal information (e.g., address, name, photo) of an individual online against their will. The term originated from the idea of sharing “docs”, or documents, of someone.

Once shared, the information cannot easily be taken back.

# Intimate partner abuse

[Freed et al. 2018]

Abusers in intimate partner violence utilize technology to intimidate, monitor, impersonate, and harass. Often, the victims are married to their abusers and share social networks and physical space.

- Owning the device or paying for the family plan, threatening to remove it

- Installing or authorizing software to track the victim (e.g., Find My Friends)

- Forcing victims to disclose social media passwords, monitoring messages

Generally, security approaches are not designed to combat attackers who know the victim intimately.

# What do we do?

There is no permanent solution here. New behaviors arise over time.

Step one: ensure that there are serious consequences for this kind of behavior — possibly legal ones.

Step two: find confidential and trusted means for users to report abuse. Develop a trusted adjudication system.

# Summary

Anti-social behavior is a fact of life in social computing systems. Trolling is purposeful; flaming may be due to a momentary lack of self-control.

The environment and mood can influence a user's propensity to engage in anti-social behavior: but (nearly) anybody, given the wrong circumstances, can become a troll.

Changing the environment, allowing mood to pass, and allowing face-saving can help reduce anti-social behavior.

Dark behavior exists: be prepared to respond.

# Social Computing

CS 278 | Stanford University | Michael Bernstein

Creative Commons images thanks to Kamau Akabueze, Eric Parker, Chris Goldberg, Dick Vos, Wikimedia, MaxPixel.net, Mescon, and Andrew Taylor.

Slide content shareable under a Creative Commons Attribution-NonCommercial 4.0 International License.