

Social Computing Capstone

Day 10: Community Moderation

CSE 481p | Spring 2023

Galen Weld

Graduate Student | University of Washington, Allen School of Computer Science & Engineering

Schedule for today's class

- Discussion of reading and lecture on community moderation (20 min)
- Groups pair up to do a prototype feedback session (60 min)

Community Moderation

What is content moderation?

Content moderation exists *everywhere* we have content!

We have always had content moderation. Back when we consumed content via newspapers, TV, and radio, content moderation meant having editors and regulators overseeing journalists and producers. There were fewer avenues for media consumption, and “gatekeepers” were fewer and had more power.

Today, the nature of content moderation has changed drastically alongside the explosion of social media.

Some types of content moderation:

Commercial content moderation: thousands of paid contractors who work for major platforms like Instagram and TikTok manually reviewing content

Algorithmic moderation: AI systems trained on previously removed comments and other signals automatically predict whether new comments should be removed

Third way: Community moderation

Members of the community, or moderators who run the community, handle reports and proactively remove comments

Examples: Reddit, Twitch, Discord, HackerNews

It's best practice for the moderator team to publish their rules, rather than let each moderator act unilaterally



MENU 

Moderator Guidelines for Healthy Communities

Effective April 17, 2017.

1 Engage in Good Faith

Healthy communities are those where participants engage in good faith, and with an assumption of good faith for their co-collaborators. It's not appropriate to attack your own users. Communities are active, in relation to their size and purpose, and where they are not, they are open to ideas and leadership that may make them more active.

Management of your own Community

- 2 Moderators are important to the Reddit ecosystem. In order to have some consistency:

Community Descriptions:

- 3 Please describe what your community is, so that all users can find what they are looking for on the site.

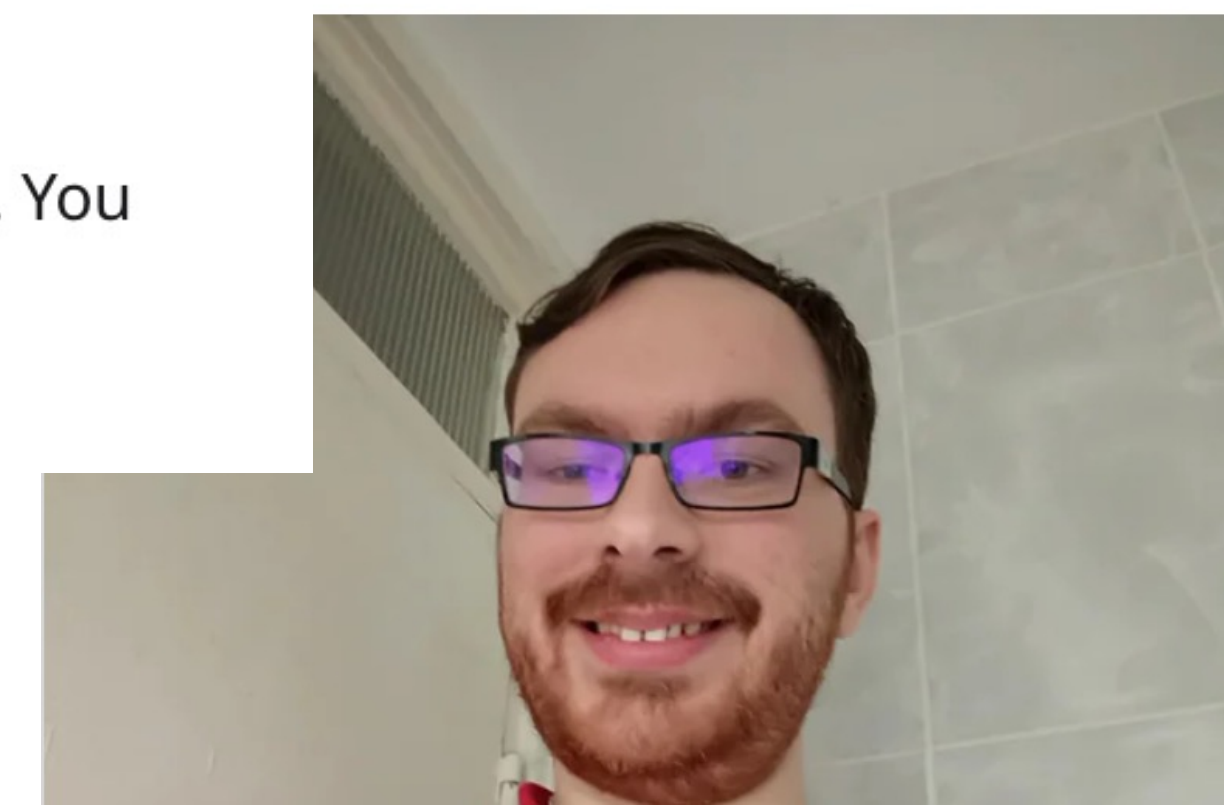
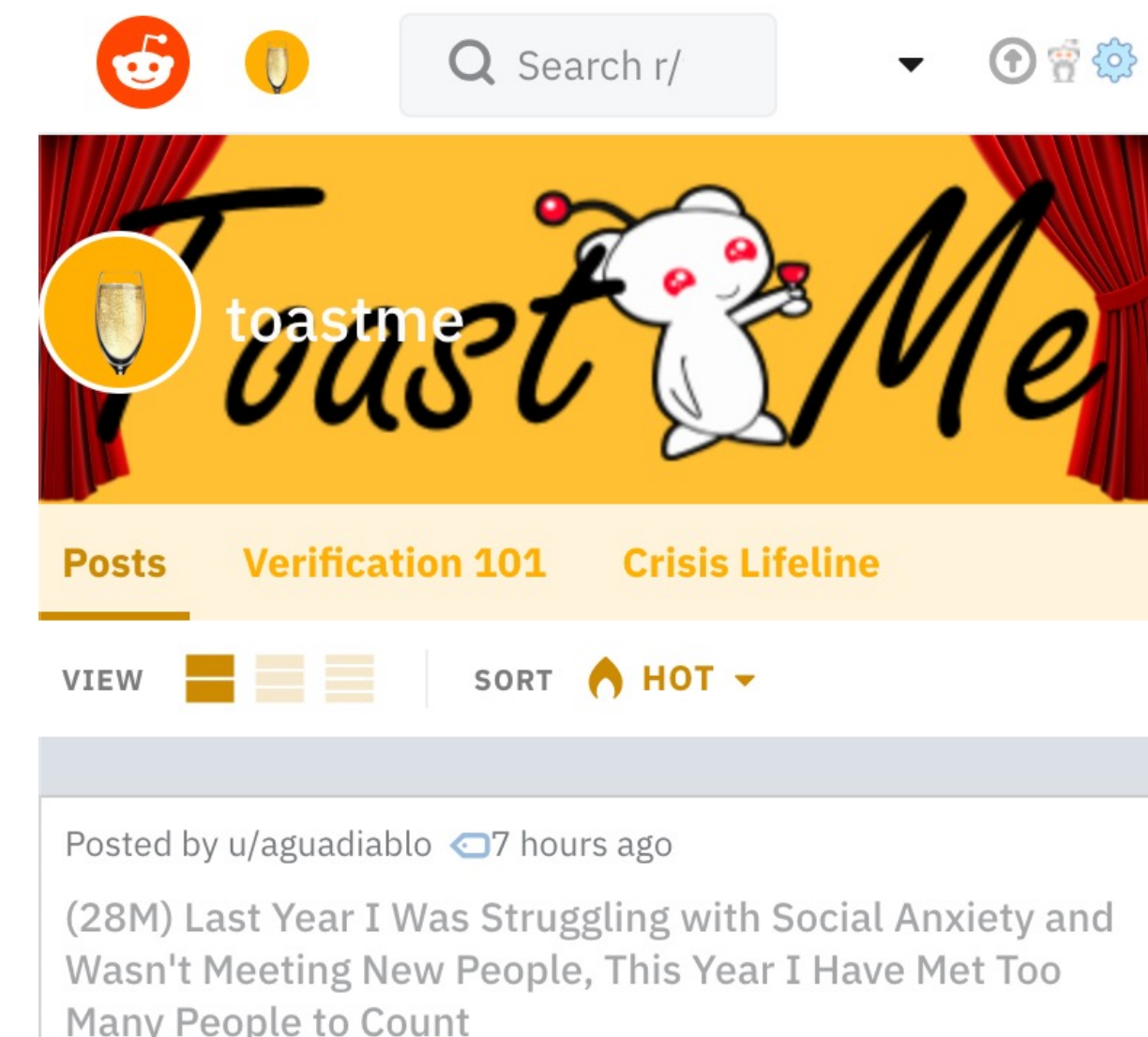
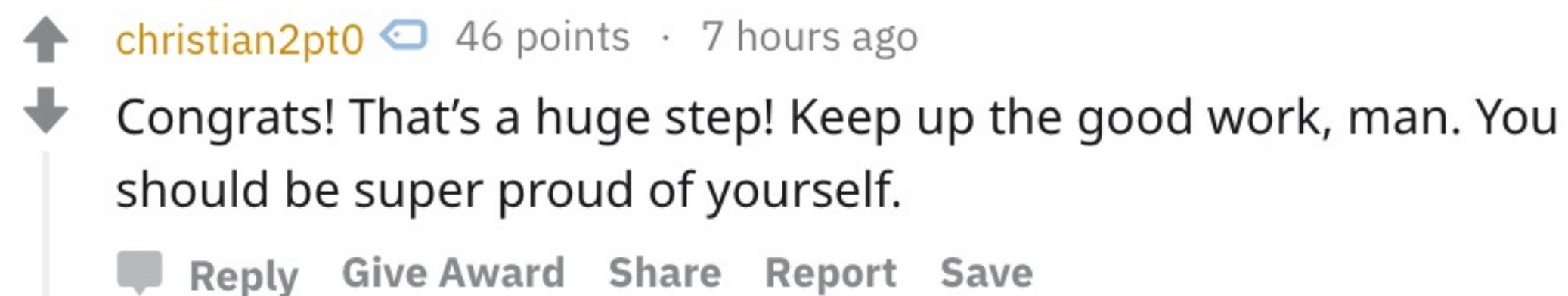
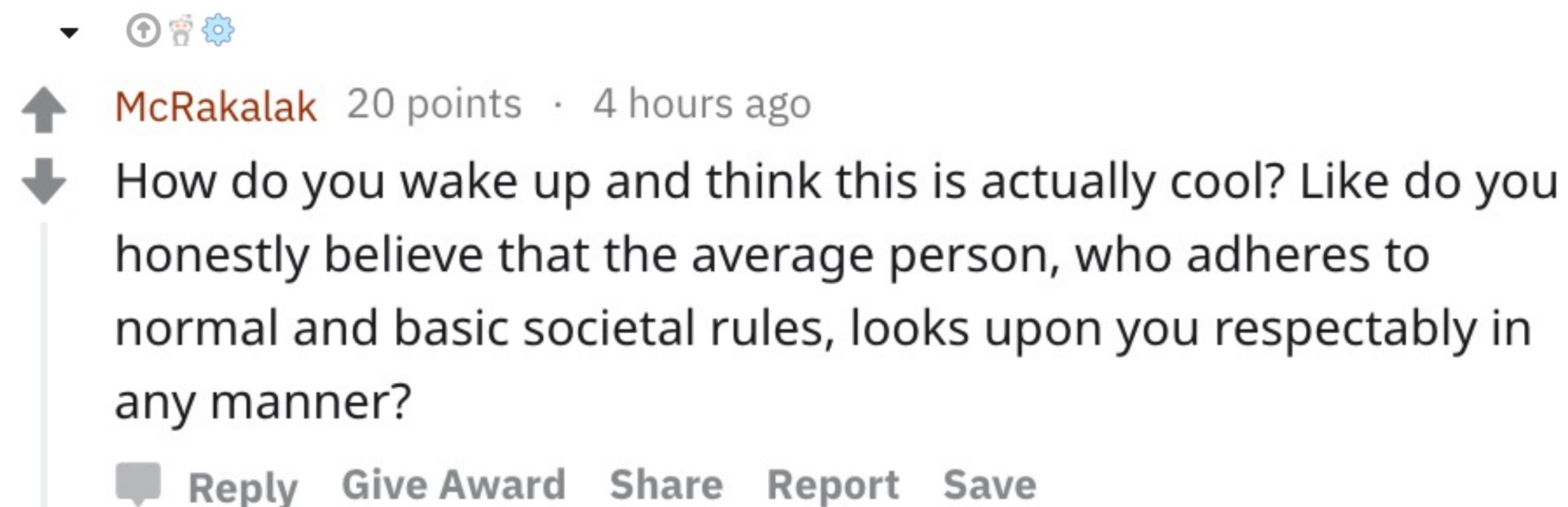
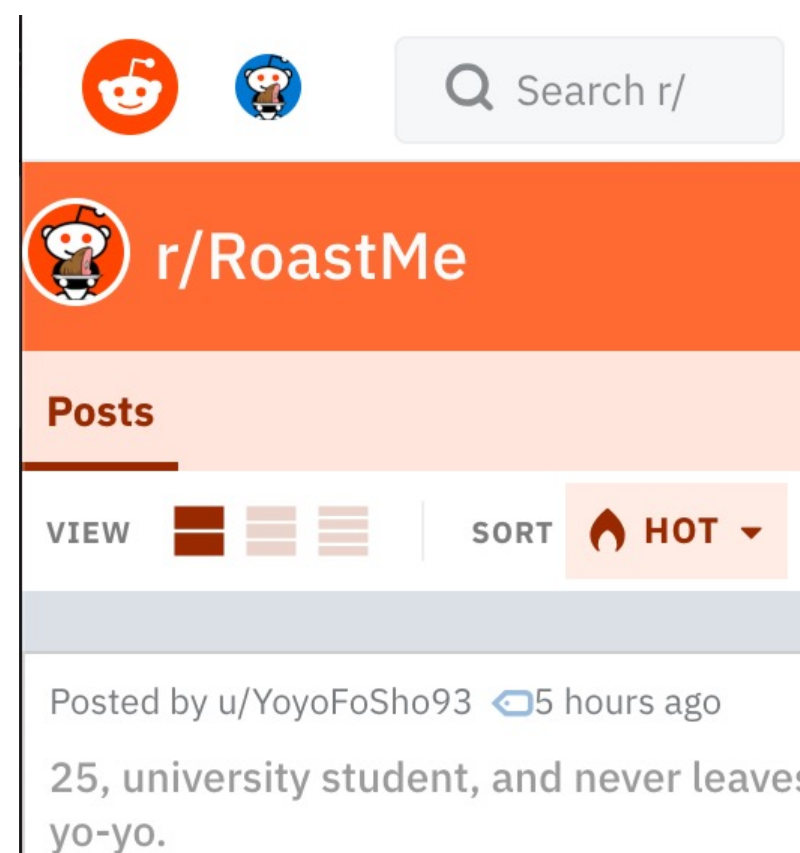
“I really enjoy being a **gardener** and cleaning out the bad weeds and bugs in subreddits that I’m passionate about. Getting rid of trolls and spam is a joy for me. When I’m finished for the day I can stand back and admire the clean and functioning subreddit, something a lot of people take for granted. I consider moderating a glorified **janitor’s** job, and there is a unique pride that janitors have.”

- /u/noeatnosleep, moderator on 60 subreddits

[<https://thebetterwebmovement.com/interview-with-reddit-moderator-unoeatnosleep>; Seering, Kaufman and Chancellor 2020; Matias 2019]

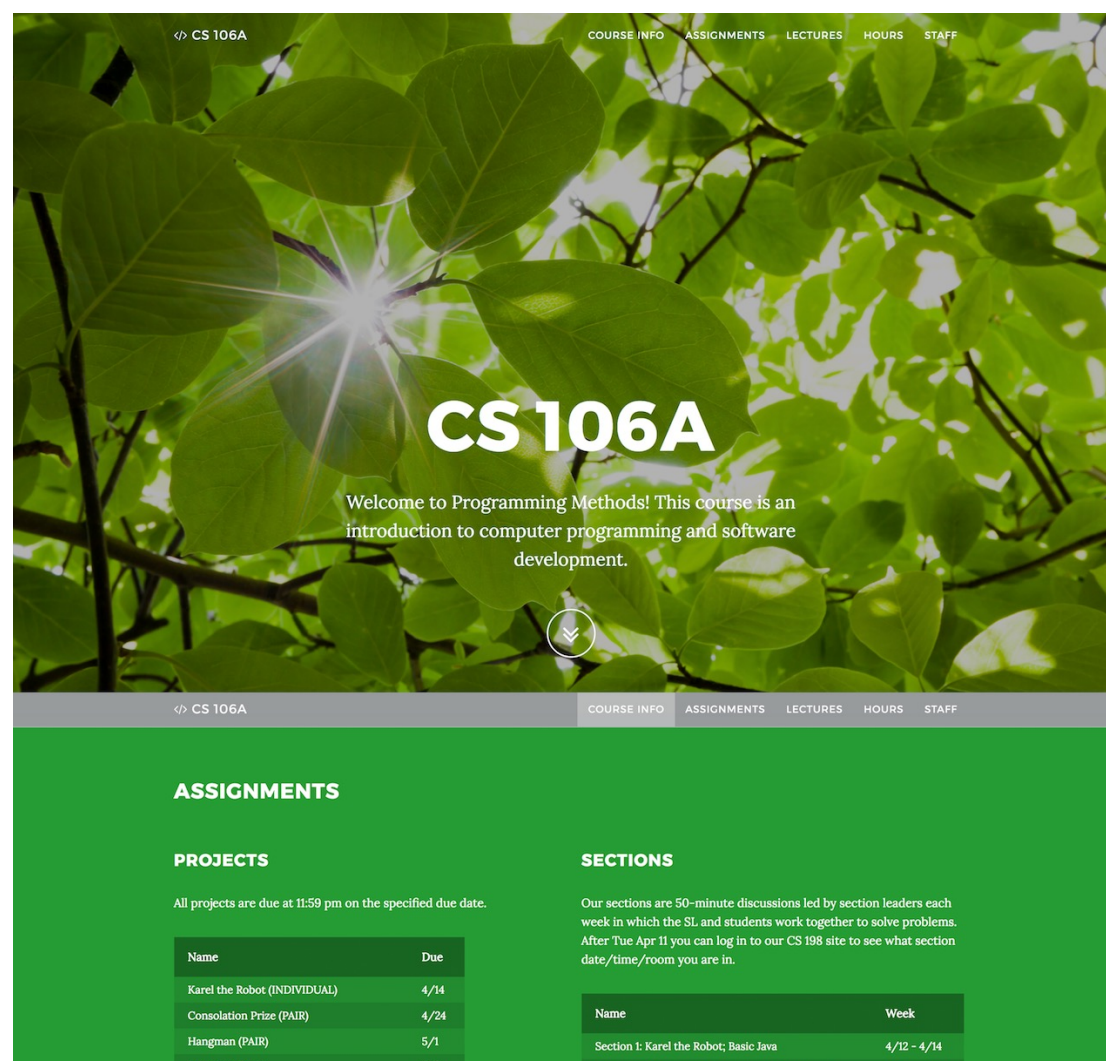
Why community moderation?

- Moderators can do a lot more than taking down content retroactively. They get to influence **community norms**.

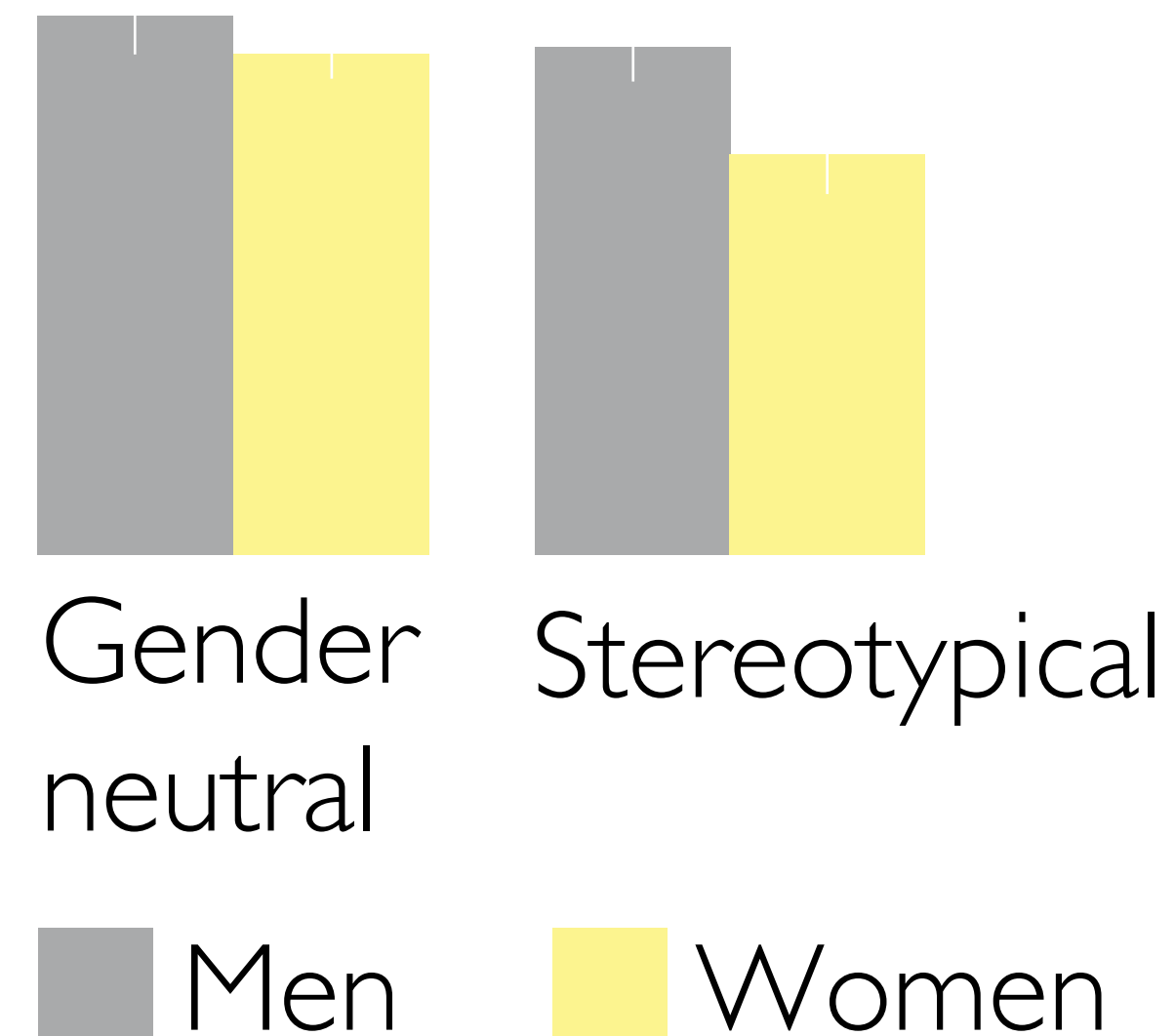


Norms are incredibly important!

- Norms are the informal rules that govern behavior in groups and societies [Stanford Encyclopedia of Philosophy]
- They can be intuited quickly.
- They can be influenced by design.



Intent to enroll



[Metaxa et al. 2018]

Descriptive Norms

- Norms can also be influenced by common behavior (descriptive norms).
- This is particularly the case for behavior by high status members of the community (moderators). Mods can model what is good behavior in a community.

Is it the norms or the people?

[Rajadesingan, Resnick and Budak 2020]

Are community norms influenced more by the people who choose to join them, or by what we see in the space?

Comparing people before and after they joined 56 political subreddits with different levels of toxicity: **it's the norms.** People match toxicity levels with their first post in the community, differing from their prior behavior in other political subreddits.






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



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  |  · [Reply](#)



User9054 · 3 hours ago
Hillary is a cunt. I am voting with my dick for Putin. /s

(Real comments on the article)

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)

Why community moderation?

- Moderators can also make their own rules. Different communities can have different rules (inviting pluralism!).
- Rules can evolve into entire systems of governance.
 - EN Wikipedia has 100s of policy pages and >1000 admins. It has courts, mediation, bureaucrats, a Supreme Court...
 - Other language Wikipedias can be very...different

Shock an aw: US teenager wrote huge slice of Scots Wikipedia

Nineteen-year-old says he is 'devastated' after being accused of cultural vandalism

 Scots, the language of Robert Burns, has been enjoying a resurgence. Photograph: S Vincent/Alamy

The Scots **Wikipedia** entry on the Canada goose - or "Canadae guiss" - was at first honest about its provenance. A tag warned: "The 'Scots' that wis uised in this airticle wis written bi a body that's mither tongue isna Scots. Please impruive this airticle gin ye can."

But, as the author grew in confidence, so he removed the caveat, and continued on his Scots-writing spree.

Now an American teenager - who does not speak Scots, the language of Robert Burns - has been revealed as responsible for almost half of the entries on the Scots language version of Wikipedia.

What's not so great about community moderation?

Invisible labor is a term drawn from studies of women's unpaid work in managing a household, emphasizing that what the women do is labor in the traditional sense, but is not recognized or compensated as such.

[Star and Strauss 1999]

Why is the labor invisible?

Because oftentimes people see just the **results** of the moderation, not the work involved behind the scenes.

The invisible nature of this labor makes moderation feel thankless, and the content that mods face can prompt **burnout**. A second component of this is **emotional** labor, or labor in which you must manage and perform emotions, which can also add to the burnout.

In addition, community moderators get little support from platforms despite enriching them.

MODERATING ON DISCORD

Discord is a place where anyone can build and manage a community dedicated to the things they love, whether that is a favorite game, creating amazing art, or simply hanging out with friends and making new ones. Moderators are at the forefront of creating spaces where people feel safe and can find belonging. Moderators are a key part of making communities great and a place where people want to gather.

Moderation takes hard work, and a commitment to learning more about how to make communities better. A key part of learning how to moderate comes simply from building and managing a community, but also from sharing knowledge learned by others sharing their insights on how to moderate communities more effectively.

We built the Discord Moderator Academy as a comprehensive resource so that anyone, from first-time moderators to experienced veterans of massive online communities, can find resources to learn about moderation, community management, and more.



Community moderation

Strengths:

Leverages intrinsic motivation

Local experts are more likely to have context to make hard calls

Mods have more levers, time, and social standing to influence norms

A Plurality of different spaces with different norms and rules

Weaknesses:

Mods don't feel they get the recognition they deserve + burnout + unpaid labor

Not necessarily consistent across a platform

Without broader oversight, mods can grow problematic communities

Lo-Fi Prototype Feedback Time! (60 min)

- Get into your groups and pair up with another group.
- Take 20-30 min for the other group to playtest your prototype(s). Don't forget to follow your playtest plan, use your script, and ask the questions you have for the other team during the playtest.
- During the feedback session, **observe and take notes** on what people do, and also take notes on what people say (**ask people to speak aloud their thoughts!**).

Next Tuesday

- Reminder that Part C of the G3 report is **due on Tuesday**. You can get additional feedback and play testing from people outside of class if you want!
- We have two readings for Tuesday on the topic of commercial content moderation. Amy and Kevin will be back from CHI! We will also have a guest visitor for the first part of class named Lindsay Blackwell - <http://www.lindsayblackwell.net/>
- Lindsay Blackwell is a recognized expert in content moderation and social media governance. She is Head of Trust & Safety for both Yik Yak and Sidechat, where she developed a unique content moderation ecosystem designed to better support the people who are most vulnerable to online harm. She previously served as lead researcher on Facebook's Hate Speech and Violence & Incitement teams, and pursued similar work as a senior researcher on Twitter's Content Health team.