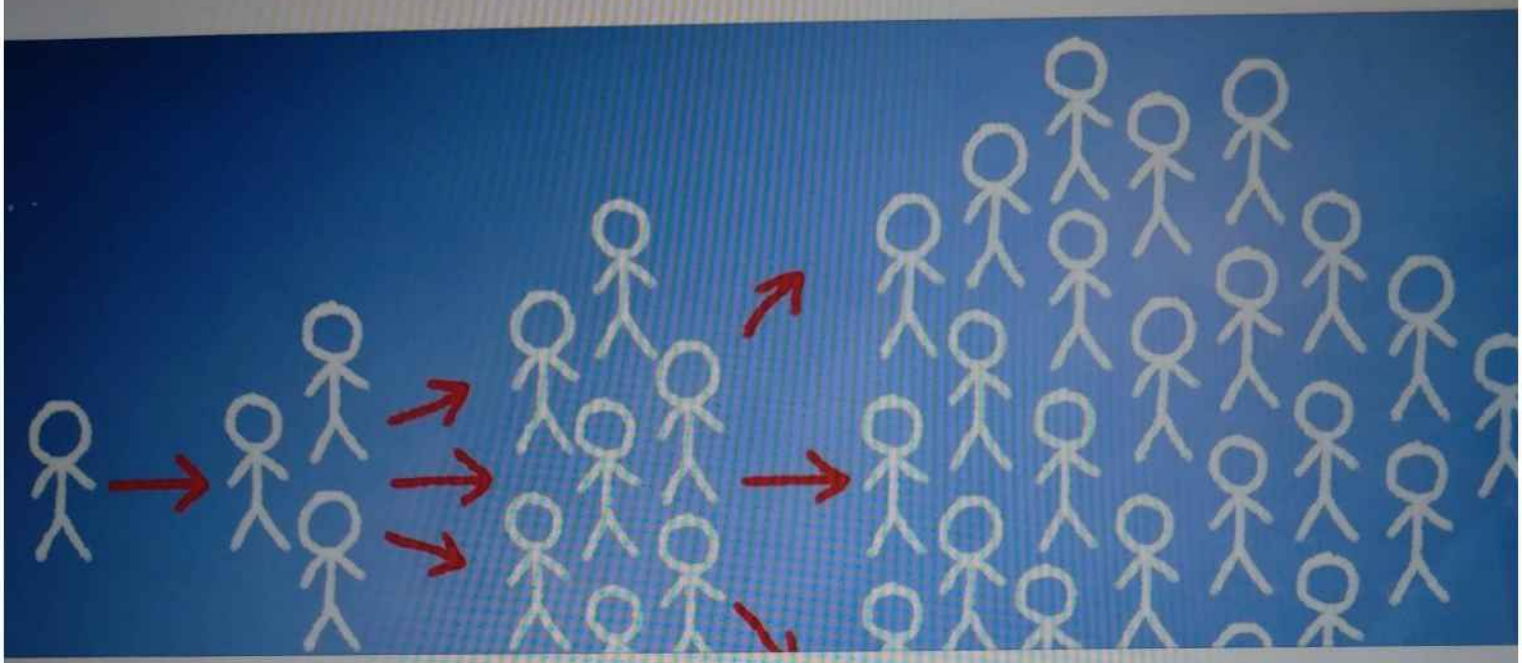
AMENDED Exhibit 234

PLAINTIFFS' OMNIBUS OPPOSITION TO DEFENDANTS' MOTIONS FOR SUMMARY JUDGMENT



We are Responsible for Viral Content

RAVI IYER · WEDNESDAY, DECEMBER 11, 2019 · READING TIME: 15 MINUTES

TLDR: Virality is a new thing that we introduced globally, and which we have a responsibility to improve entirely independent of free speech concerns. I suggest: improving external transparency of what goes viral, more precisely matching incentives to user value, collecting more negative signals, and creating more domain specific algorithms.

When I first joined Facebook, I was one of the roughly 50% Americans who was worried about how angry, disrespectful, and uncivil political conversations on social media are, as compared to other places where people discuss politics. While I still worry ab

Chats

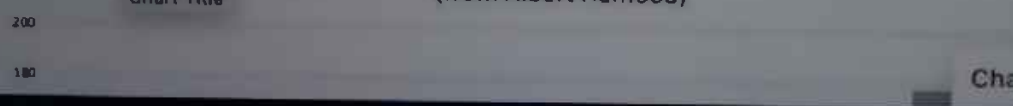
MacBook Pro

EXHIBIT
21
HAUGEN - 9/25/25

When I first joined Facebook, I was one of the roughly 50% Americans who was worried about how angry, disrespectful, and uncivil political conversations on social media are, as compared to other places where people discuss politics. While I still worry about that, I have also come to appreciate that there are also positive aspects to giving a voice to many and that it isn't necessarily our job to get in the way of what people want to say to their friends and family. Indeed, there are plenty of other ways to communicate to one's close friends and stopping people from spreading offensive jokes to their friends might just change the method of delivery rather than actually making any real difference. It also would bring about a whole other set of risks as we attempt to judge which jokes are or are not violating our standards, in contexts that we may not understand.

However, I believe it is entirely possible to have a greater appreciation for free speech while also simultaneously having a greater concern for our role in what goes viral. Unlike communication with close friends and family, virality is something new we have introduced to many ecosystems (so new that many spell-checkers don't recognize the word) and it occurs because we intentionally encourage it for business reasons. For example, users have told us the pages they would like to see content from, but we often override those explicit preferences because our predictions of what will get shared and engaged with disagree. In one recent analysis by Albert Hamood of pages that get views in a week, we can see that users get relatively little connected content (content from pages they chose to like) as opposed to unconnected content that is reshared by others, even as we know such content is less valued by users (see chart below from data from Albert's post).

Comparing Connected vs. Unconnected Pages with Views for a User in a Week
(from Albert Hamood)

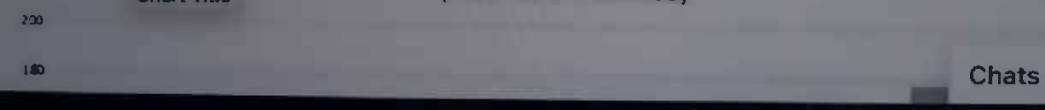


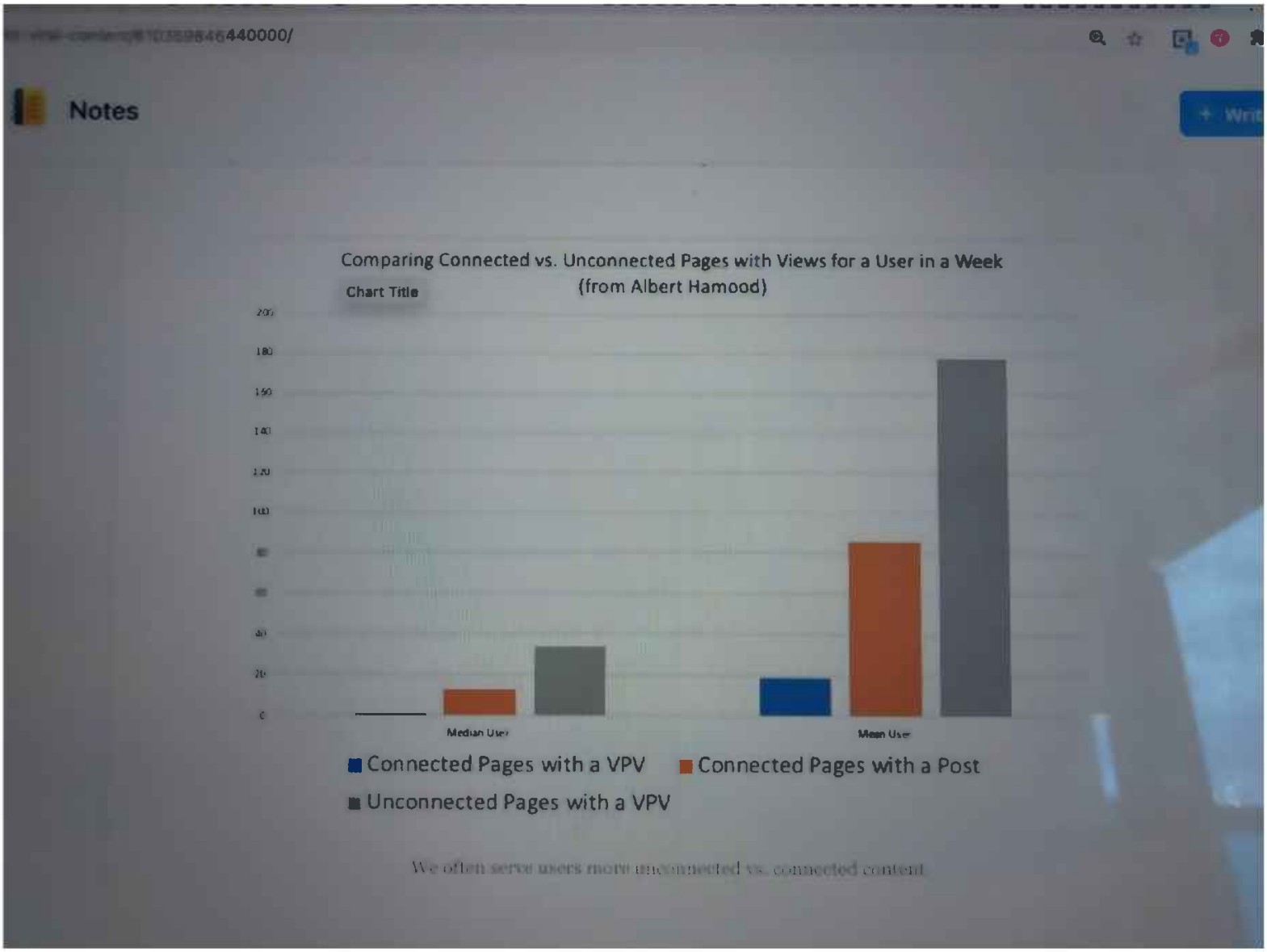
MacBook Pro

When I first joined Facebook, I was one of the roughly 50% Americans who was worried about how angry, disrespectful, and uncivil political conversations on social media are, as compared to other places where people discuss politics. While I still worry about that, I have also come to appreciate that there are also positive aspects to giving a voice to many and that it isn't necessarily our job to get in the way of what people want to say to their friends and family. Indeed, there are plenty of other ways to communicate to one's close friends and stopping people from spreading offensive jokes to their friends might just change the method of delivery rather than actually making any real difference. It also would bring about a whole other set of risks as we attempt to judge which jokes are or are not violating our standards, in contexts that we may not understand.

However, I believe it is entirely possible to have a greater appreciation for free speech while also simultaneously having a greater concern for our role in what goes viral. Unlike communication with close friends and family, virality is something new we have introduced to many ecosystems (so new that many spell-checkers don't recognize the word) and it occurs because we intentionally encourage it for business reasons. For example, users have told us the pages they would like to see content from, but we often override those explicit preferences because our predictions of what will get shared and engaged with disagree. In one recent analysis by Albert Hamood of pages that get views in a week, we can see that users get relatively little connected content (content from pages they chose to like) as opposed to unconnected content that is reshared by others, even as we know such content is less valued by users (see chart below from data from Albert's post).

Comparing Connected vs. Unconnected Pages with Views for a User in a Week
(from Albert Hamood)





Confidential

otes

Here are four reasons why virality is a very different issue from free speech, as implemented by the algorithms we have created. TLDR: Our algorithms are not neutral.

We often optimize for network value rather than just user value.

Part of the reason this occurs is that we don't just predict what you would like (what we internally call "local value"), we also predict what will get the most downstream engagement (sometimes called 'network value"). Sometimes these things diverge and we show things to users that we think they have a small chance of sharing, leading to comments between people who see it downstream, over things that have a greater chance of being explicitly liked by that user. For example, in this analysis by [redacted] we see that as things are more viral (getting to a user via many reshares), we generally don't see an increase in engagement. Our ranking systems have specific separate predictions for not just what you would engage with, but what we think you may pass along so that others may engage with. Unfortunately, research has shown how outrage and misinformation are more likely to be viral, and recent experiments that deprecate these models indicate that removing these models does positively impact metrics for misinformation and hate. Not all content that benefits from our choices here is bad, but as long as we continue to optimize for overall engagement and not solely what we believe individual users will value, we have an obligation to consider what the effect of optimizing for business outcomes has on the societies we engage in.

We often use ambiguous signals to decide what goes viral.

We prioritize content based on engagement, but engagement doesn't necessa

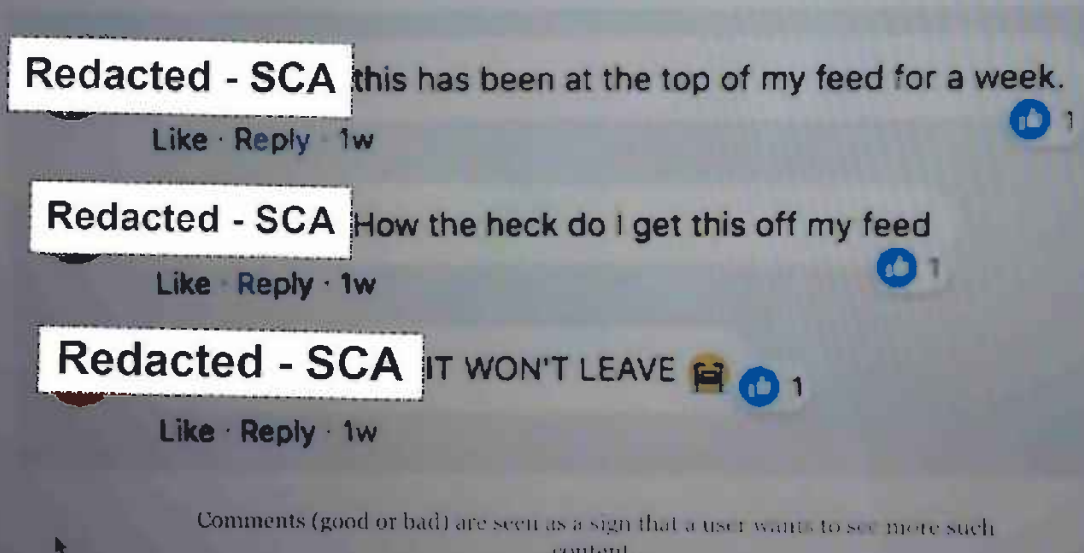
Chats

MacBook Pro

Search Bunny

We often use ambiguous signals to decide what goes viral.

We prioritize content based on engagement, but engagement doesn't necessarily mean that a user actually wants to see more of something. One of our biggest signals we use to provide more similar content is comments and as this Buzzfeed article points out (image from article below), a comment that you hate a thing can be seen as a positive signal leading content to get outsized distribution.



Comments (good or bad) are seen as a sign that a user wants to see more such content.

People game this in various ways, posting ever more outrageous things to get comments and reactions that our algorithms interpret as signs we should allow things to go viral. Publishers and campaigns tell us that they are incentivized to do this and we can see in our data that the above tactic of inciting users to negatively comment works to get more distril

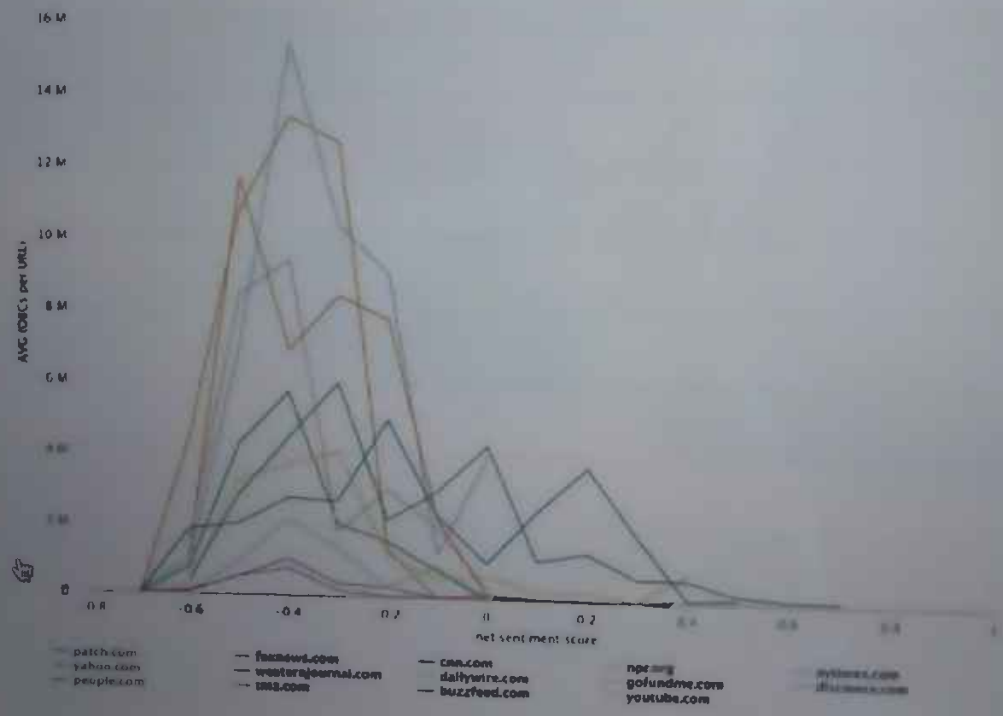
Chats

MacBook Pro

Help

lotes

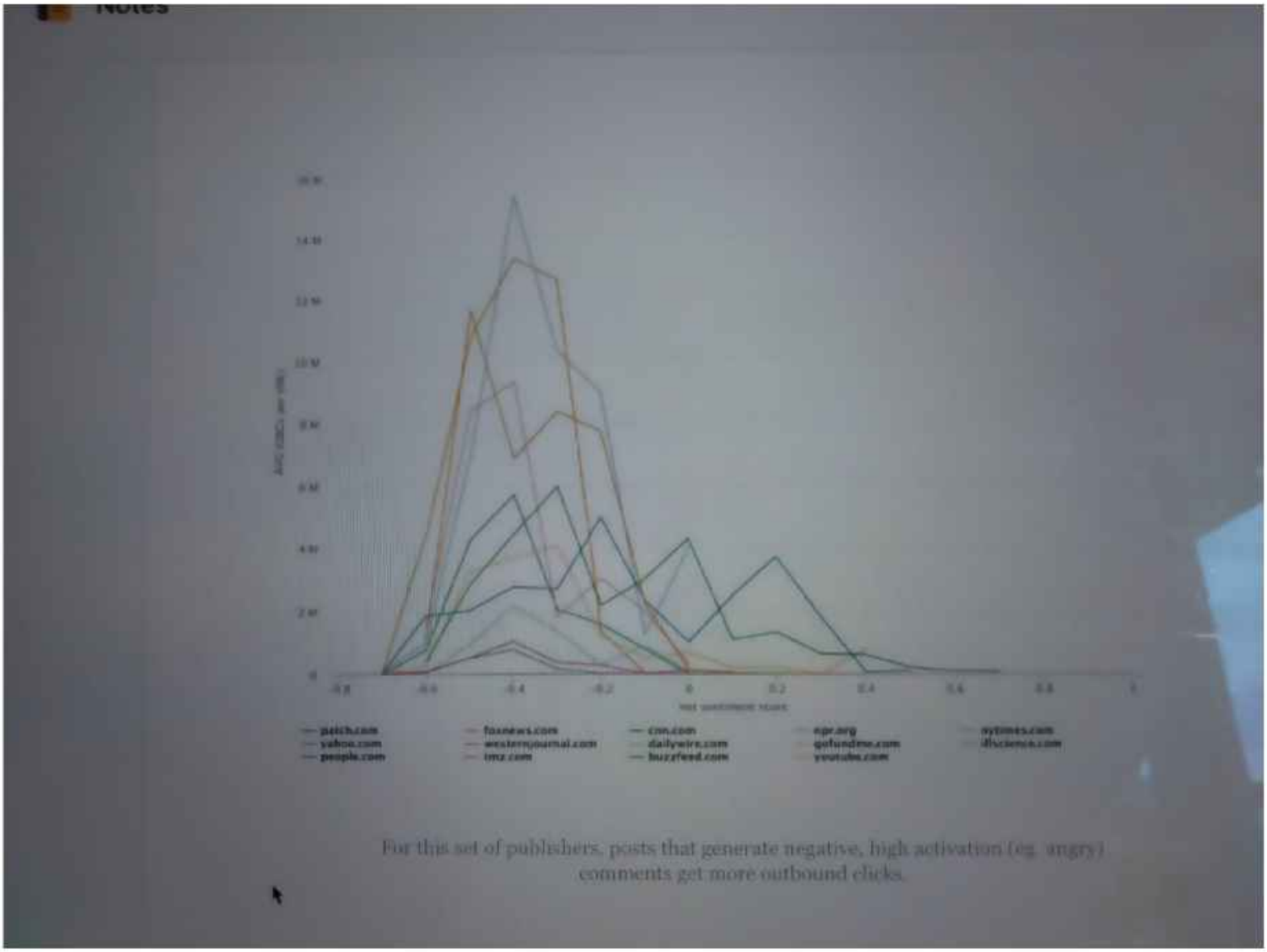
People game this in various ways, posting ever more outrageous things to get comments and reactions that our algorithms interpret as signs we should allow things to go viral. Publishers and campaigns tell us that they are incentivized to do this and we can see in our data that the above tactic of inciting users to negatively comment works to get more distribution (analysis by Delia Mocanu in graph below) .



For this set of publishers, posts that generate negative, high activation (eg. an[illegible] comments get more outbound click[illegible].

Chats

MacBook Pro



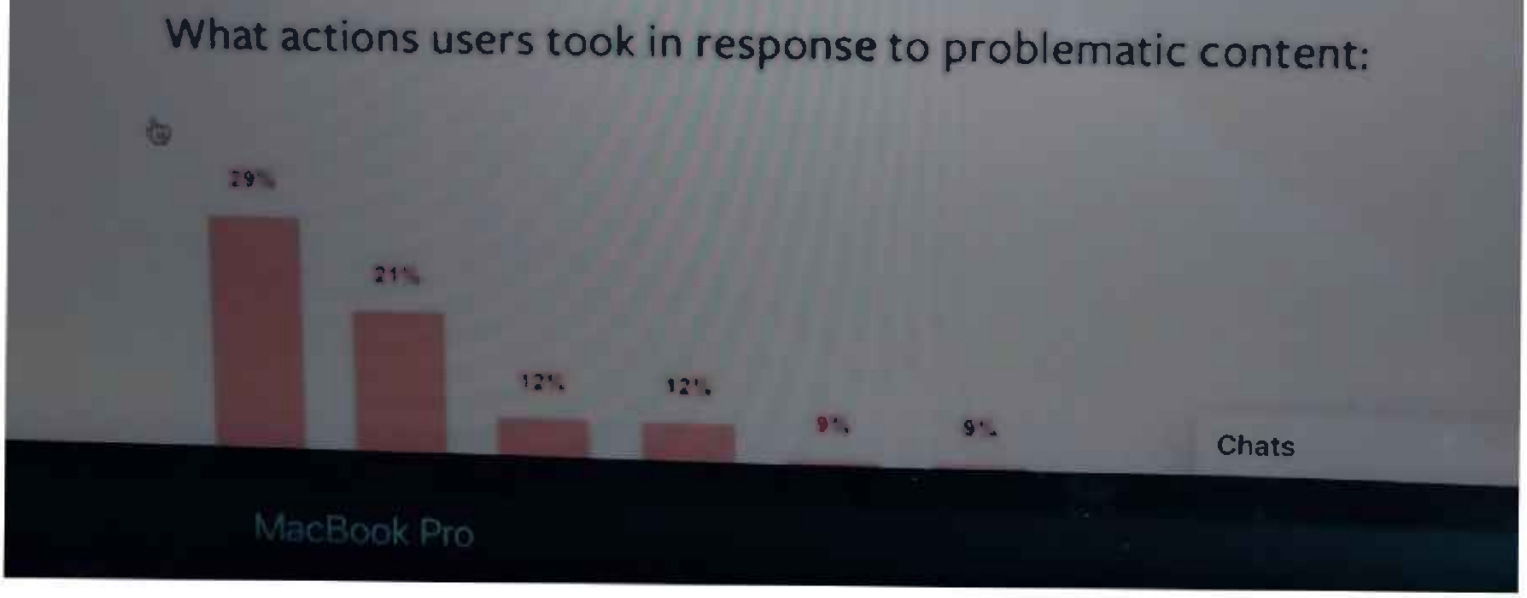
For this set of publishers, posts that generate negative, high activation (eg. angry) comments get more outbound clicks.

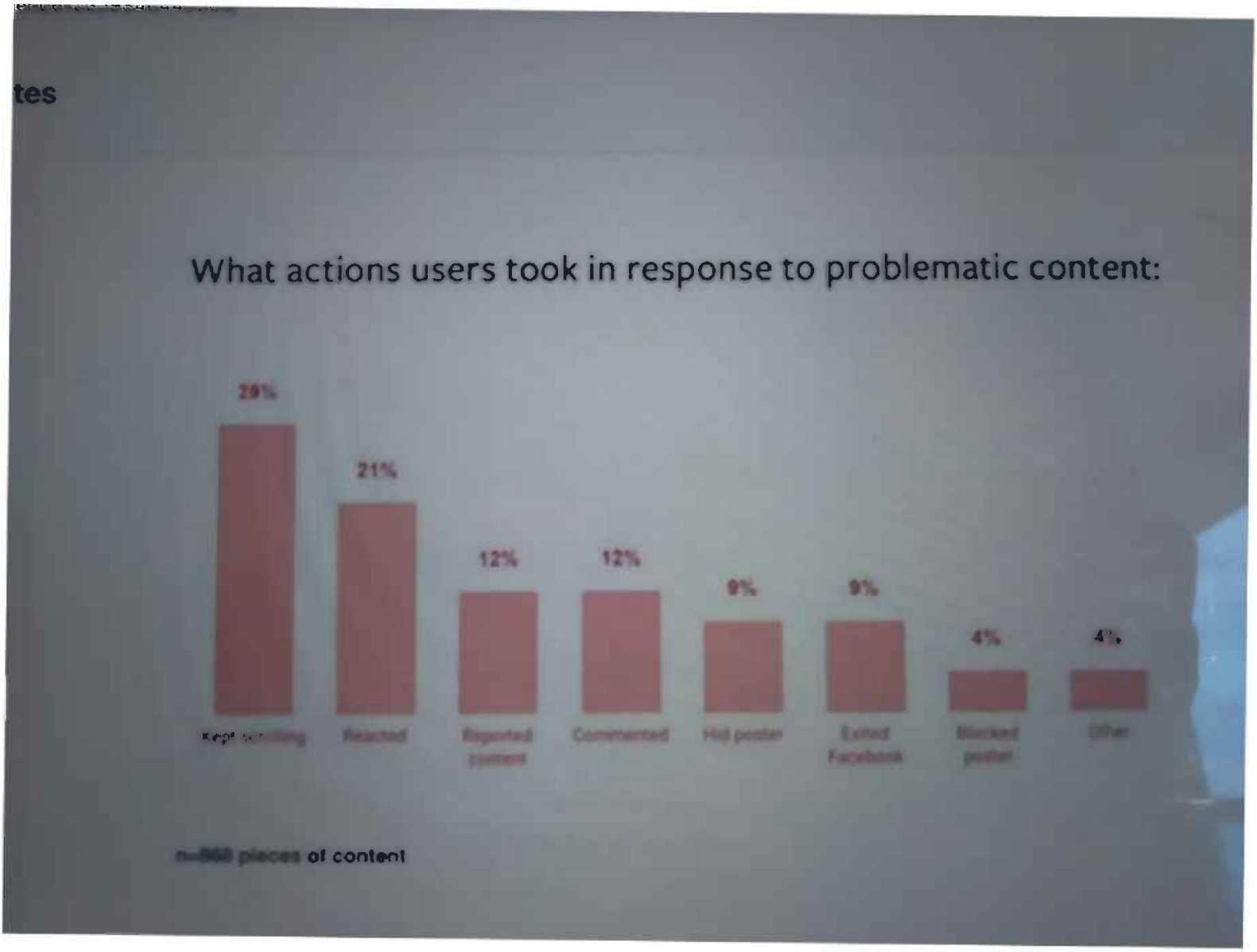
9846440000/

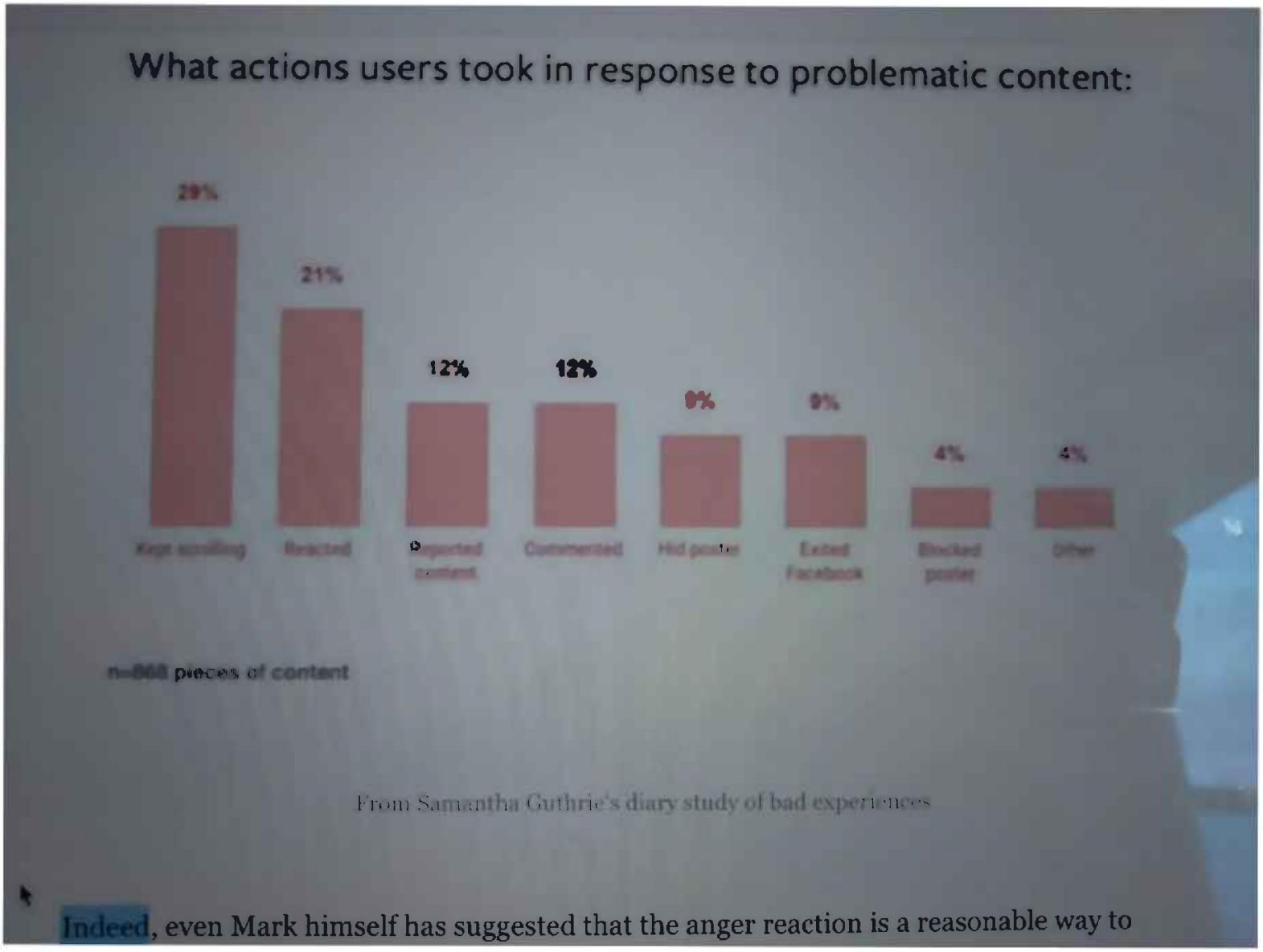
At a minimum, we should stop promoting content based on ambiguous signals, as it not only promotes bad experiences (in some cases), but also makes it easier for bad actors to game our algorithms by creating content (e.g. clickbait, sensationalism, outrage-bait) that our algorithms think are wildly popular, but are bad user experiences.

We make it difficult for users to give us clear negative signals.

A small group of people can make content go viral even if a larger group dislikes that content as there is no practical mechanism for the larger group to indicate their bad experience. What do people do when they have a bad experience? The Buzzfeed writer's experience above is not an anomaly. In one recent analysis of bad experiences by Samantha Guthrie, people were more likely to react (21%) or comment (12%) when faced with a bad experience than to do something that we would interpret as a negative signal, in part because providing a negative signal is difficult with the current user interface. Most people simply do nothing.



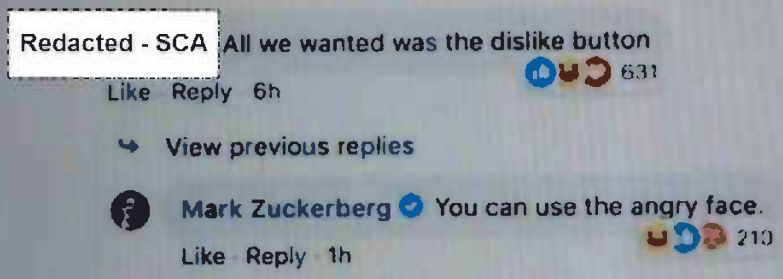

tes
What actions users took in response to problematic content:
29%
21%
12%
12%
9%
9%
4%
Reacted
Commented
Hid poster
pieces of content



Indeed, even Mark himself has suggested that the anger reaction is a reasonable way to

Confidential

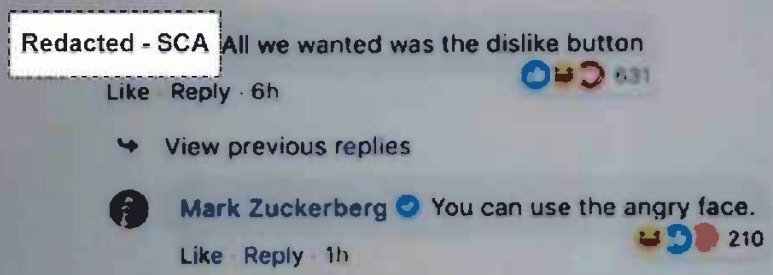
Indeed, even Mark himself has suggested that the anger reaction is a reasonable way to express that you don't like a piece of content (link here), even as we currently count it as 4x as important as a less ambiguous "like" for giving you more such content.



Mark acknowledges that reactions can be used to indicate dislike.

Recent data science analyses also confirm that Mark's suggestion to use an angry reaction as a dislike is one that users often take to heart as the angry reaction is positively associated with hiding/reporting/blocking content. A large body of research has found that negative experiences are far more impactful than positive experiences, so it would make sense to go out of our way to collect more negative experience signals, which would also have the advantage of being internationalizable, capturing borderline experiences, adapting to cultural context, and being more externally defensible. Evan Turner has pointed out that even though bad experiences are somewhat prevalent (estimates in his note), we capture but a fraction of those due to the fact that reporting/hiding are not top level actions. He also points out the wider benefits of getting more negative signals in more detail as well as some ideas for implementation (note here). Not providing scalable ways for people to signal bad experiences means that our algorithms will necessarily be skewed against preventing them as no AI system can predict something it doesn't have data on.

Indeed, even Mark himself has suggested that the anger reaction is a reasonable way to express that you don't like a piece of content (link here), even as we currently count it as 4x as important as a less ambiguous "like" for giving you more such content.



Mark acknowledges that reactions can be used to indicate dislike.

Recent data science analyses also confirm that Mark's suggestion to use an angry reaction as a dislike is one that users often take to heart as the angry reaction is positively associated with hiding/reporting/blocking content. A large body of research has found that negative experiences are far more impactful than positive experiences, so it would make sense to go out of our way to collect more negative experience signals, which would also have the advantage of being internationalizable, capturing borderline experiences, adapting to cultural context, and being more externally defensible. Evan Turner has pointed out that even though bad experiences are somewhat prevalent (estimates in his note), we capture but a fraction of those due to the fact that reporting/hiding are not top level actions. He also points out the wider benefits of getting more negative signals in more detail as well as some ideas for implementation (note here). Not providing scalable ways for people to signal bad experiences means that our algorithms will necessarily be skewed against preventing them as no AI system can predict something it doesn't have data on.