

Types of gaps between policy and bad experiences

Policy can be effective, but ...

1. incident doesn't meet the policy bar though clearly bad
2. incident doesn't fit current policy definitions
3. incident is hard to detect (by human or automated review)

Policy isn't the right tool when ...

4. "bad" is impossible to determine with any consistency
5. enforcement actions (hard or soft) cannot address the problem

Examples of bad experiences often in policy gaps

Mass harassment

Very intense, low reach. Disproportionately affects creators. Portion isn't CS policy violating but still felt as intense.

Non-credible or non-violent threats

Doesn't meet CS violating policy bar, but felt intensely with moderate reach.

Thinspiration

Easy to find content adjacent to eating disorder part of SSI and to broader social comparison and body image issue-triggers.

TL; DR

	Finding	Implication
1	Hate speech, divisive civic content, and graphic violence are frequently and intensely experienced, and have been shown to have a negative effect on sentiment and legitimacy, particularly with repeated exposures over time.	<ul style="list-style-type: none">● Prioritize Offensive Speech in near-term efforts to improve sentiment.● Examine ways we can identify and target high-violation ecosystems where people experience repeated exposures.
2	Borderline content can be seen as equally or more harmful than violating content and decreases sentiment and engagement. In most cases, users want Facebook to hide or remove it. 70.7% US users believe Facebook should be doing more to address harmful content	We should ensure we have an adequate understanding of which borderline content users find most offensive so that we can prioritize and refine interventions and actions.
3	Post content is not the only problem-- toxic and divisive comments commonly appear on benign posts . Reshares, Links and Status Updates are more likely to be rated as a Bad Experience compared to photos and videos	Include comments as a target for classifications and actions
4	Not every “bad experience” is unwanted. Some respondents describe “needing to see” content they considered a bad experience, such as violence and racism.	As we design systems that classify and demote content, and make tradeoffs across user value, engagement, and legitimacy, we should be mindful that content that seems bad, upsetting or anger-inducing may be positively regarded by the viewer as meaningful and important.
5	Users want Facebook to act. They hold us responsible for negative experience, and most think Facebook should automatically remove severe integrity-related content and hide less severe content. They perceive exposure to integrity harms as worse than false positive actions on benign posts.	Continue investing in current efforts to reduce exposure to violating and borderline content.

TL; DR *cont.*

	Finding	Implication
6	User experiences, preferences and perceptions vary. Reaction to content varies by gender, ethnicity, culture and other factors; sentiment of Low-exposure users is more affected by integrity harms; those with low digital literacy are more likely to see violating content; some may even deliberately seek out harmful content.	<ul style="list-style-type: none">● Personalization could be relevant for multiple interventions, not just soft demotion. E.g., selective display of warning screens and/or tombstones.● Offering controls can also be a way to capture relevant signals for integrity-related AI models
7	The #1 legitimacy detractor is the perceptions that FB is not doing enough to mitigate bad experiences on the platform. Legitimacy is also challenged by lack of transparency & understanding of ranking & enforcement . Content controls such as 'sensitive content preferences' serve a double role - not only do they reduce exposure, they help the user feel they understand what's under the covers.	Continue working to reduce bad experiences. .When evaluating the effectiveness of interventions, assess both impact on prevalence and impact on legitimacy.

Note: [see appendix](#) for information on observable attributes related to the likelihood a FB user has/had a bad experience

Bad experiences are common and frequent

2 in 5 users

say they've had a bad experience while using Facebook **THIS WEEK**



AXIS

1 minute after user opens app

the moment **when bad experiences are likely to happen** to users →

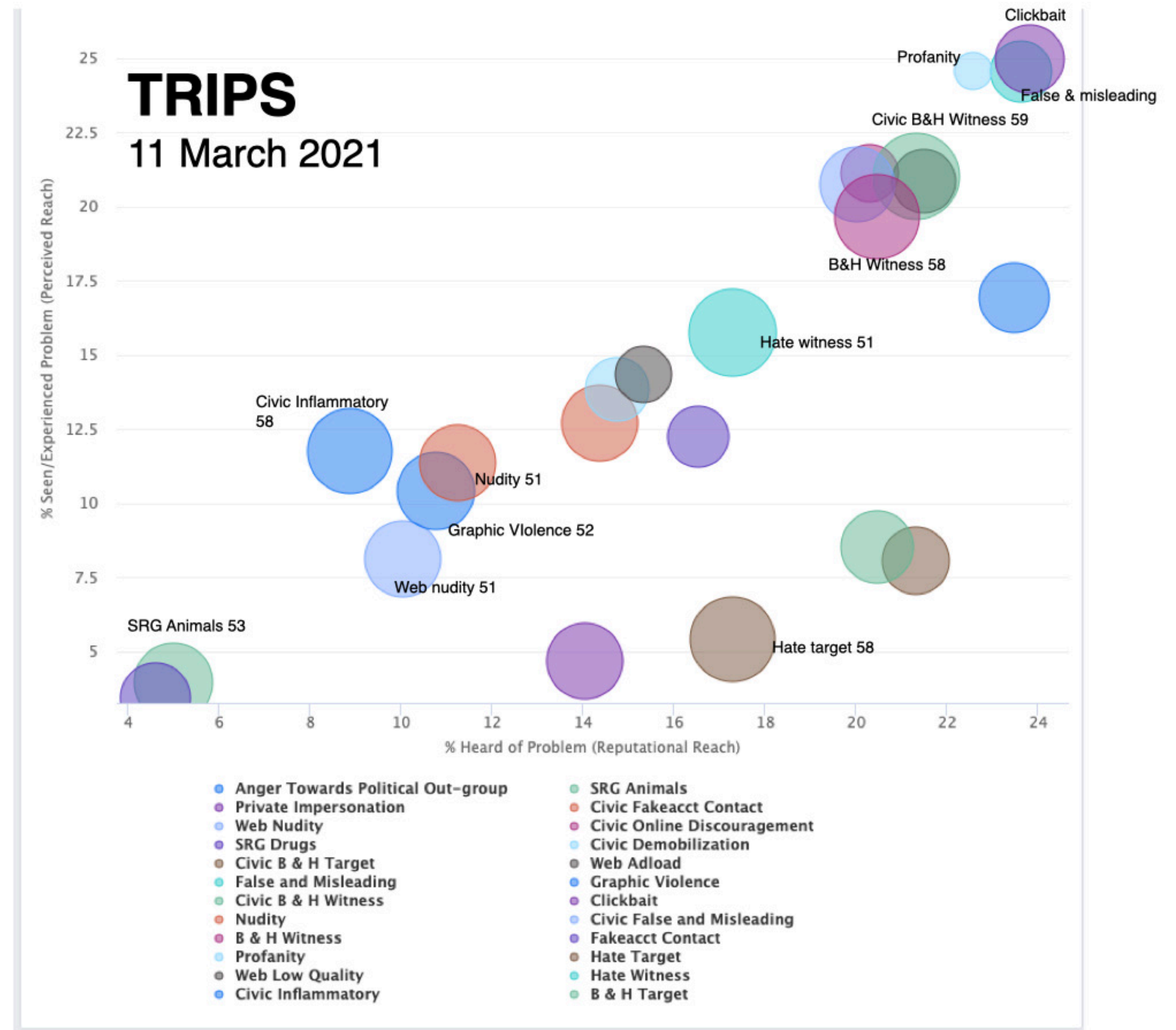
US diary study participants

Clickbait, misinfo, and profanity are the most commonly perceived harms, but toxic, hateful content is more intense

Most intense experiences:

- Civic B&H Witness
- Civic inflammatory
- B&H witness
- Hate Witness
- SRG Animals
- Graphic violence
- *Nudity / Web Nudity*

○ *Note: exposure to nudity has been shown to increase FSS*



Bad experiences are diverse

Bad Experiences (as operationalized in TRIPS) can include →

- Hate speech & discrimination
- Profanity
- Graphic violence
- Private impersonation
- Nudity
- Obscene website
- Drug sales
- Bullying & harassment
- False/misleading
- Fake accounts
- Clickbait
- Low quality Link
- Animal sales
- Ads farms

Post content is not the only problem--**toxic and divisive comments** commonly appear on **benign posts**.

And, content that leads to bad experiences does **not always violate Community Standards**. Borderline content is also a significant contributor to bad experiences.

In fact, borderline content is **1.5x closer** on average to violating than benign content in perceived harmfulness.

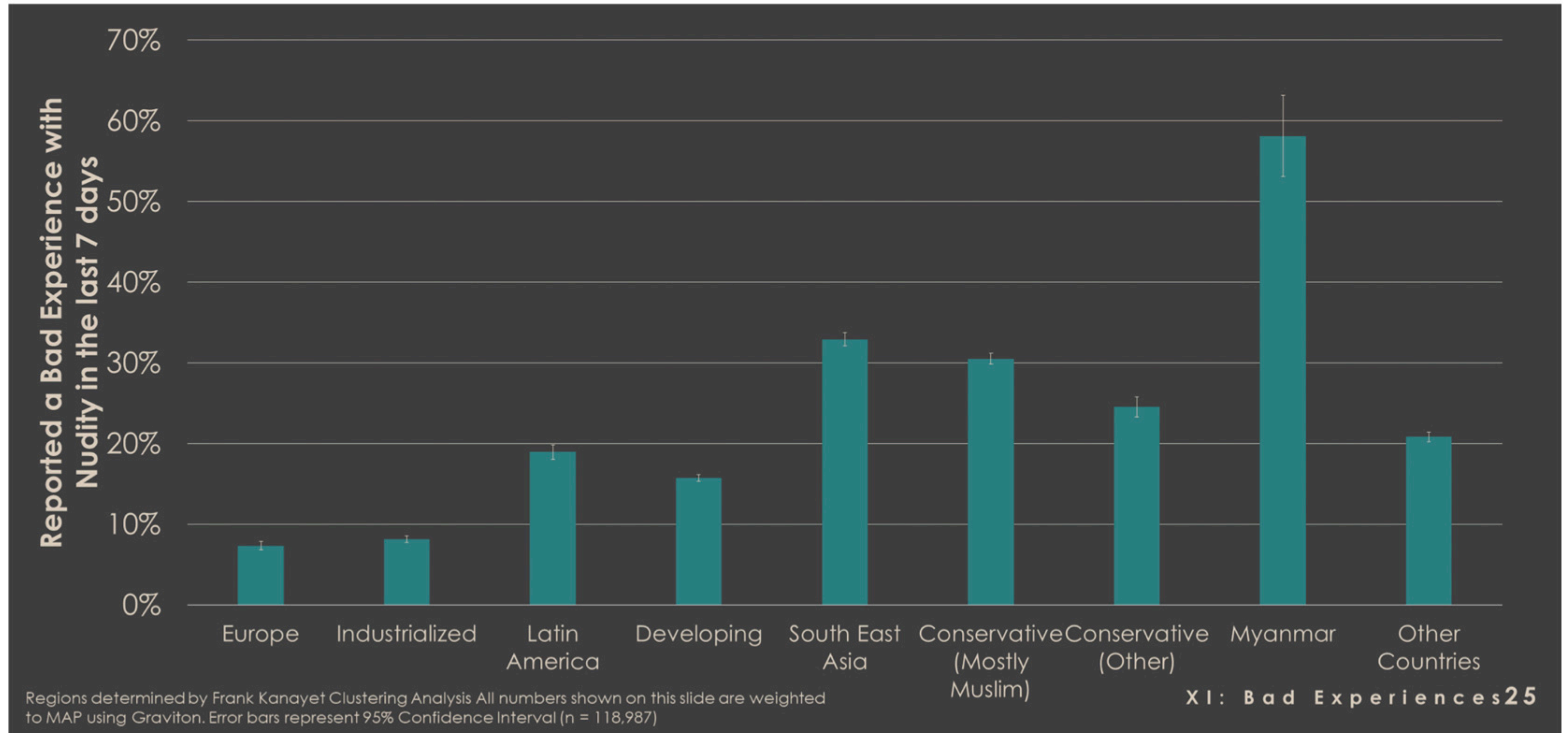
Users in this study rated borderline content* **as harmful** as violating content.



*4 types of borderline content were tested in this study: misinformation, toxic, demonizing, and hateful content.

source: [Borderline Content User Survey \(Aug 2018\)](#)

We don't all share the same set of values and beliefs. Bad experiences with harms like nudity can vary by market



Some respondents describe “needing to see” content they considered a bad experience, such as violence and racism.

- For example, they need to see content containing police brutality or military violence against civilians to better understand and contextualize the world around them.

Bad Experience

≠

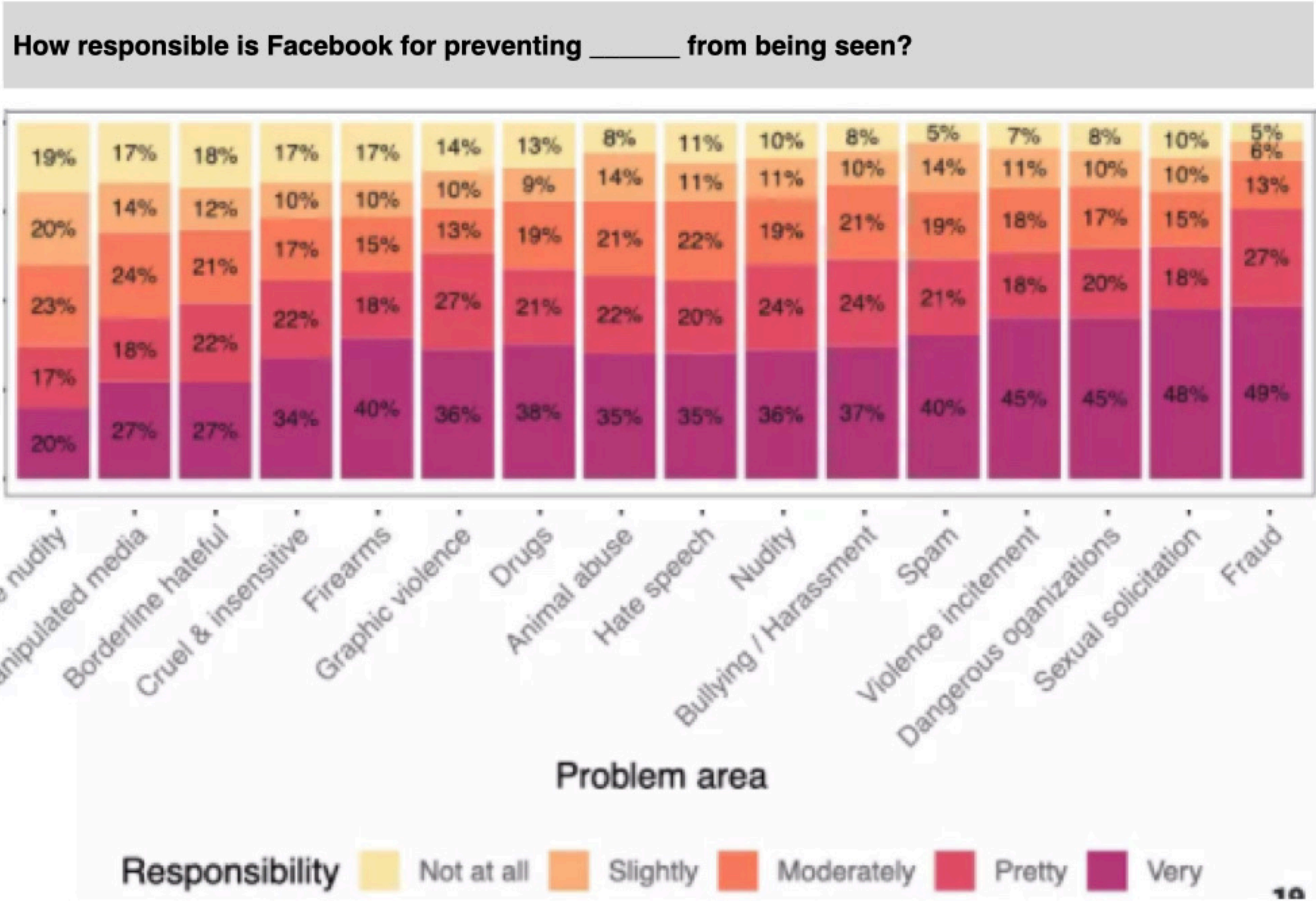
"Do Not Want to See"

Why should we care about bad experiences?

Bad experiences & borderline content are related to decreased sentiment and engagement

- Increased exposure to borderline content (FUSS Red/Yellow), was **linked to a decrease in DAP** ([Liu 2018](#)).
- Exposure to borderline civic hate leads to **negative emotions, disengagement from Facebook**, and decreased engagement with offline civic and political actions ([Travaglianti and Sacramone-Lutz 2019](#))
- Among US survey respondents, recent actual experiences with **hate speech or graphic violence** on Facebook was correlated with **lower average satisfaction** with News Feed ([Powell 2020 #1](#), [Powell 2020 #2](#)). These users also reported that that the posts in their feed were **less likely to be worth their time**, and that they had **fewer meaningful interactions** on feed.
- **Sentiment** is most negatively associated with integrity harm exposure among **low-exposure users**
 - Possible implication: set user-level prevalence reduction goals rather than total VPV goals to prioritize these users?
 - Possible implication: Warning screens could be shown more liberally to low-exposure users?
- When shown examples of borderline content (*misinfo, toxic, hateful*), the majority of US survey respondents said: they did not want to see it, felt that Facebook should hide or remove it, and **reported that they would spend less time** on Facebook after seeing it in their Feed ([Bodford, et al 2018](#)).

Users hold Facebook responsible for addressing both violating and borderline content



Majority of respondents felt Facebook is pretty or very responsible for preventing 13 out of 16 problems ([Powell 2020](#))

"I think that I would rather just see something where it was covered up and ask me, "Hey, are you okay with seeing this content?" And then that is my right."

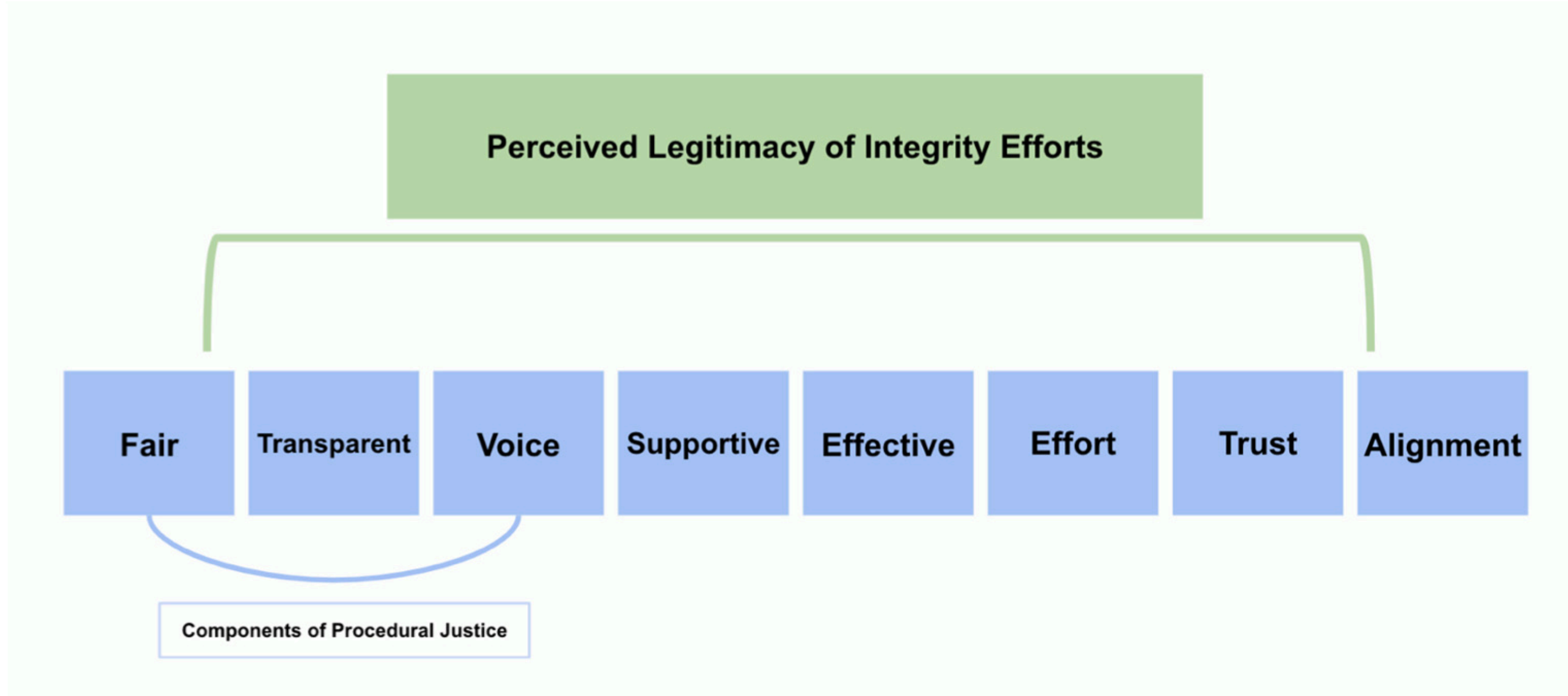
- Female participant on seeing animal abuse in her Feed.

Seeing borderline content on Facebook makes participants feel like Facebook **does not care** about them. They say Facebook is....



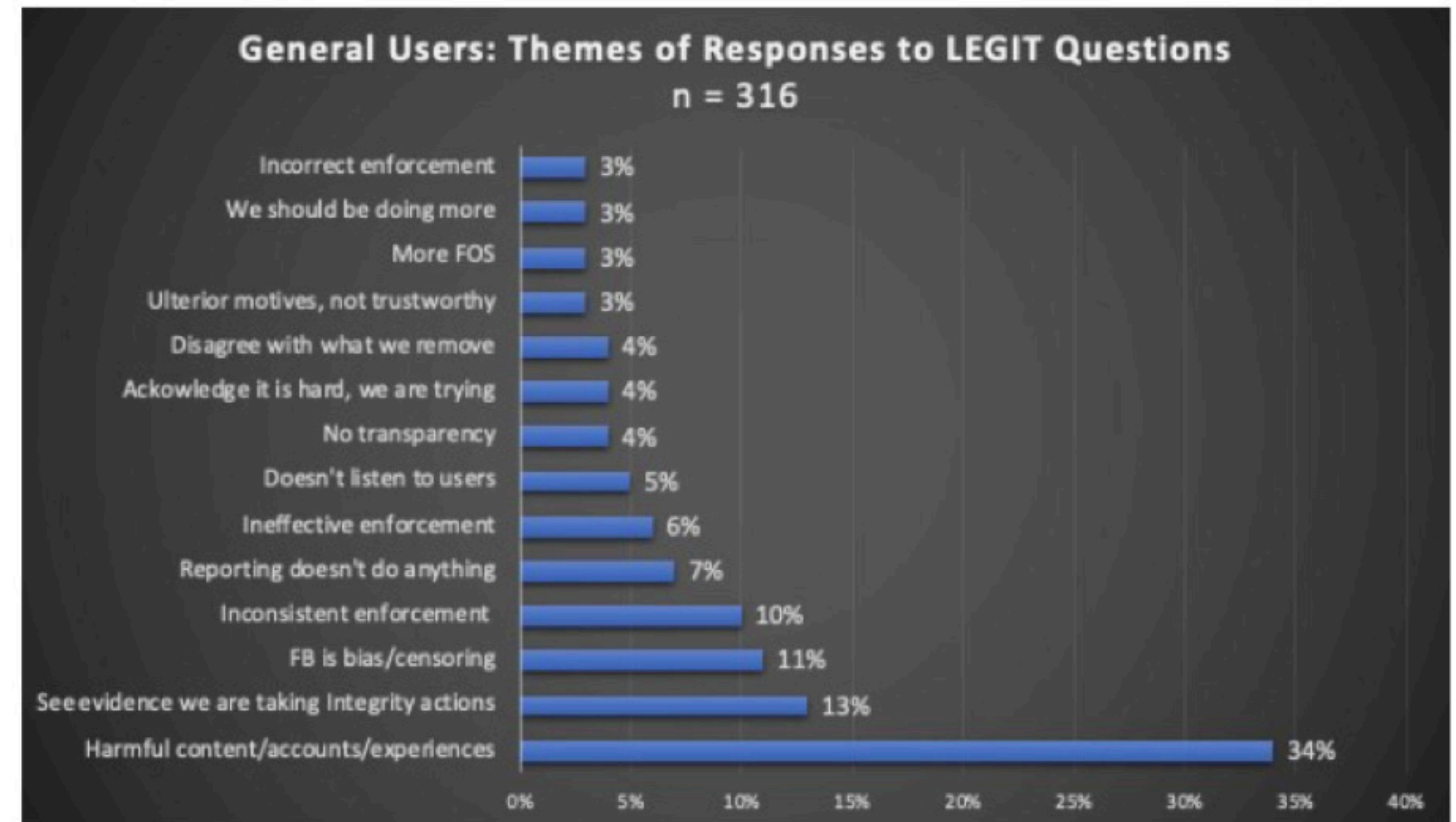
What is legitimacy and how do we measure it?

If our integrity efforts are legitimate, it means that people and external stakeholders believe that our integrity work is **effective** at reducing harm and that our enforcement is **defensible** and **fair**.



Biggest legitimacy detractor:
Harmful content/accounts/experiences

- People who report seeing graphic violence or hate speech felt Facebook was significantly less effective at reducing harm on the platform
- Those who recently saw graphic violence also had more negative perceptions of how hard Facebook is trying
- We must be also careful that as we reduce prevalence, users perceive that we enforce **fairly and consistently**



People don't trust that Facebook is actually reducing bad content on the platform--or that we are motivated to do so

Dear Facebook, I can't take it anymore. You keep feeding me and billions of others a whole bunch of lies and crap that is designed to influence people and spread conspiracies.

You get paid very well to do this and because of that have not taken the necessary measures to prevent it. You welcome it.

Lack of transparency & understanding of **ranking** and **enforcement** cause suspicion and lead to perceptions of bias

1. People don't have a good understanding of how ranking works.

"I mean, I don't know if they're actively pushing stuff down based off what I don't want to see, but I do think they're pushing stuff higher on my feed based off stuff I interact with. Like I mentioned, so maybe that just ends up pushing stuff down, but I'm not a hundred percent sure." - Male, 18 - 24

1. People don't have a good understanding of how Facebook enforces its rules, and enforcement can seem inconsistent and arbitrary.

*"I think that would be a good idea to put it in black and white, what they consider a hate speech and then follow their own rules. **Don't favor one side or another. Hate is hate.**" - Female, 55-64*

1. Given the lack of understanding of our rules and ranking, the greatest value of our controls may be the transparency they provide into Facebook's processes.

Negative experiences with our enforcement actions hurt legitimacy

- ▶ Facebook both **over-enforced** on non-violating speech, and **under-enforced** against clear violations of policy, leaving a Black user with low trust in our systems.

▶ *I have friends who, say things on Facebook, Who are saying things that aren't hate speech, but they'll get their accounts suspended. But then there are people who will say things like, Oh, these monkeys are over here protesting, or they'll say the N word or things like that. And they'll be perfectly fine never get in trouble or anything, basically let than go by. But my friends can't even say that they're tired of police shooting people and they get suspended from their Facebook. - Female, 18 - 24*

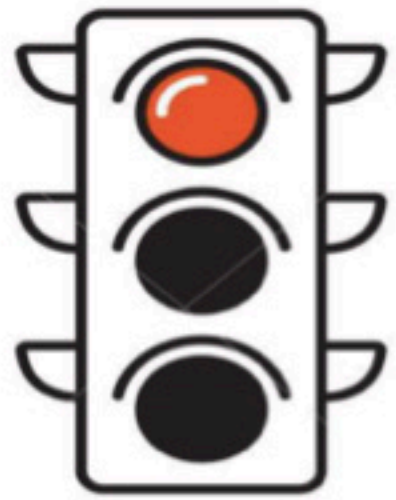
- ▶ For another, the **lack of clarity on why** a post was actioned on made her more skeptical of Facebook's systems and enforcement.

▶ *"Not myself, but the magazine got banned, it said, "You cannot sell this, it's inappropriate". And they don't tell you why. And I've written them and I never heard anything." - Female, 55-64*

“Between your clearly racist and misogynistic ‘moderation’ to your laughable appeals process, there is a reason you are called RACEBOOK. Clearly I am not included in the community you claim to protect.”

In addition, there are significant gaps in our enforcements

Quantitative analysis indicates there **are gaps in our enforcements.** →



High severity harms w/ enforcement gaps

- animal abuse
- violence incitement
- animal sales
- fraud
- bullying
- divisive content →



Low severity harms w/ enforcement gaps

- cruel/insensitive
- sexual solicitation
- firearm sales
- drug sales
- Voter misinfo
- impersonation

There are also **needs related to specific content types**, for example:

- Key user-reported pain points with **news** indicates we need to do more on: clickbait; purposefully misleading; and emotionally manipulative news →
- We over- and under-enforce in groups based on harm type →
 - For example, we underenforce V&I in complex entities, such as groups pages and events

Expanding coverage of warning screens could address key user concerns and legitimacy issues...

- Warning screen use could be expanded to cover any content a viewer may not be mentally or emotionally prepared for
- This can be done on a **personalized basis** - except in the case of misinfo treatments. Reduce thresholds for applying warning screens and use personalized models to predict how low thresholds should be for a given user based on their individual tolerances
- Increase the scope of what these personalized warning screens cover; **animal abuse, hateful/toxic language** should be candidates for expansion
- Increase use of info treatments to additional kinds of content for additional categories of misinfo content, and expand pool of raters who can trigger warning screens

Improve UI to mitigate drawbacks and solicit feedback:

- Utilize straightforward messaging that explains why and how the content was covered. Utilize humility and be open that we may have gotten it wrong
- Provide opportunity for users to give feedback on appropriateness of cover, **utilize this feedback to tune algorithm**

Personalize coverage to reduce cluttering

....though too many warning screens could lead excess friction and bad user experience for consumers

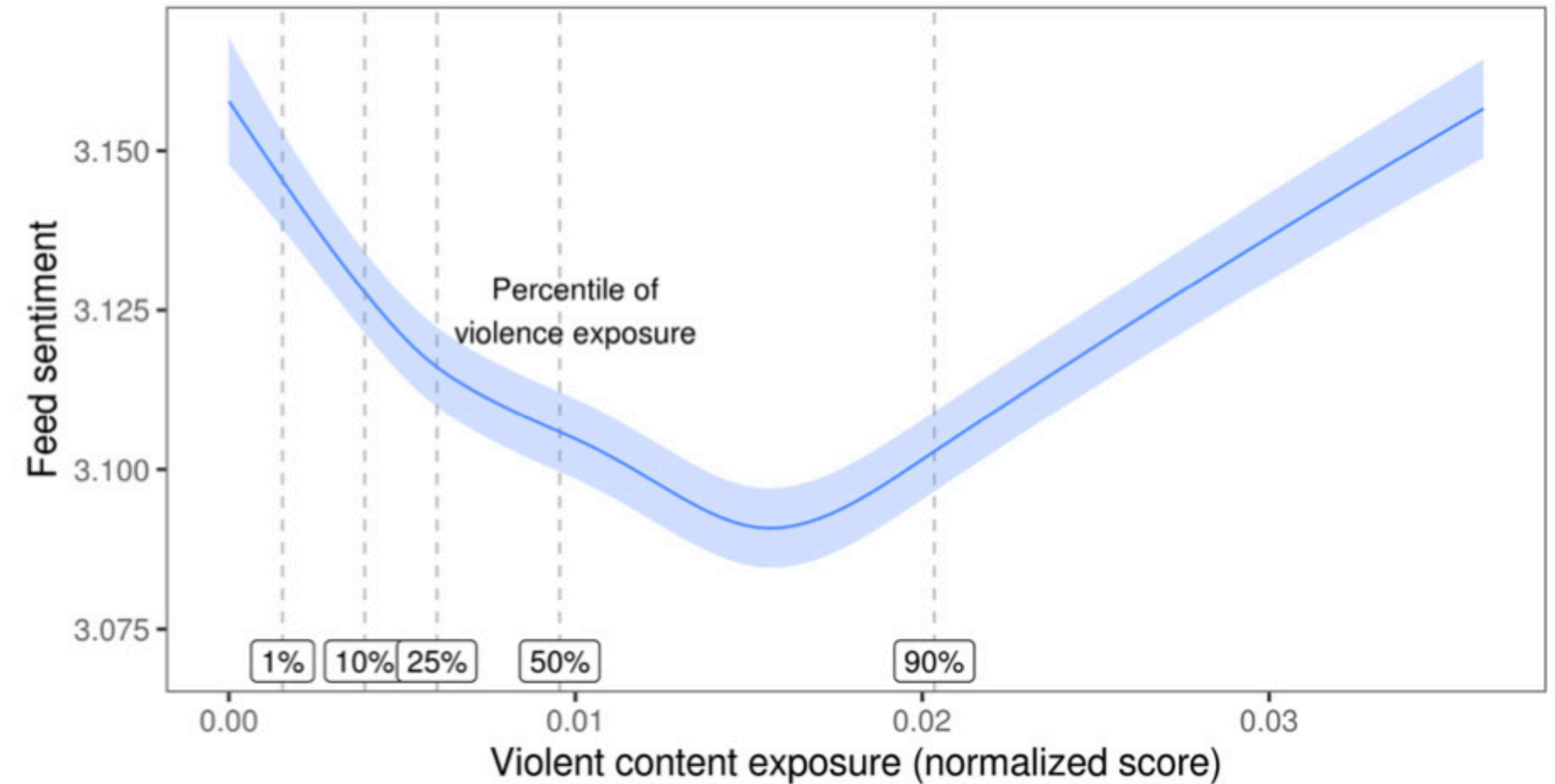
Warning screen applications could be **personalized per-user** to keep the frequency of their use low per user.

Exposure to graphic violence is most negatively associated with sentiment for users who rarely see violent content.

Personalizing warning screens would allow us to protect these most vulnerable users without cluttering feed for users with higher tolerances for violence.

Violent content exposure and feed sentiment

Controlling for age, gender, country, tenure, # friends, # vpvs, % vpvs from pages; N = 611012 respondents (all)



Why does control matter?

Currently, users feel as if they lack control over the content they see in Feed.

This sense of a lack of control, further exacerbated by the presence of unwanted content in Feed, leads to a strong want and need for user controls.

Our existing controls, broadly defined, are underused and do not properly serve users in the way we intended.

One of the greatest barriers to adoption is due to discoverability.

Providing users with greater control, either through new controls (easier ways to hide or ranking controls) or the simplification of older controls, will empower users and increase user sentiment toward Facebook.

From: Arturo [REDACTED]

Subject: Gap in our understanding of harm and bad experiences

Date: October 5, 2021 at 9:37:59 PM PDT

To: Mark Zuckerberg [REDACTED]

Cc: Sheryl Sandberg [REDACTED], Chris Cox [REDACTED], Adam Mosseri

[REDACTED], Mark Zuckerberg [REDACTED]

Dear Mark,

I saw the note you shared today after the testimony, and I wanted to bring to your attention what I believe is a critical gap in how we as a company approach harm, and how the people we serve experience it. I've raised this to Chris, Sheryl, and Adam in the last couple of weeks.

I want to start by saying that my personal experience, and what I believe, is that you and m-team care deeply about everyone we serve, and my goal in sending this is to be of service to that. It's been 2 years since I've been back part-time.

51% of Instagram users say 'yes' to having had a bad or harmful experience in the last 7 days. Out of those 1% of report and of those 2% have the content taken down (i.e. 0.02%). The numbers are probably similar on Facebook.

Two weeks ago my daughter [REDACTED], 16, and an experimenting creator on Instagram, made a post about cars, and someone commented 'Get back to the kitchen.' It was deeply upsetting to her. At the same time the comment is far from being policy violating, and our tools of blocking or deleting mean that this person will go to other profiles and continue to spread misogyny. I don't think policy/reporting or having more content review are the solutions.

There is detailed data about what people experience in TRIPS, a statistically significant survey. We ran a more detailed survey, I've attached the full age breakdown below, but here are some key numbers (these questions are in the last 7 days):

21.8% of 13-15 year olds said they were the target of bullying.

39.4% of 13-15 year olds said they experienced negative comparison.

24.4% of 13-15 year old responded said they received unwanted advances.

Why does someone think it is ok to post 'get back to the kitchen' or harass someone? I believe it is because it doesn't violate policy, and other than deleting or blocking, there is no feature that helps people know that kind of behavior is not ok. Another example, is unsolicited penis pictures.

[REDACTED] has received those from boys too since the age of 14, and her tool is to block them. I asked her why boys keep doing that? She said if the only thing that happens is they get blocked, why wouldn't they?

Why the gap between Prevalence and TRIPS? Today we don't don't know what % of content people experience as misinformation, harassment, or racism is policy violating. We have done great work in driving down prevalence, and there will always be more to do, but what if policy based solutions only cover a single digit percentage of what is harming people?

Policy is necessary when the content is unambiguously inappropriate, yet it has many limitations. It trails behavior, the interventions are heavy and risk over-enforcement and getting the border line right is extraordinarily difficult. Policy enforcement is analogous to the police, it is necessary to prevent crime, but it is not what makes a space feel safe.

What makes a workplace, or a school, or a dinner table feel safe is social norms.

If someone goes around telling women to 'get back to the kitchen', and the only thing that happens is their content is deleted or they get blocked, don't we run the risk of normalizing bad behavior? How are people to learn to be members of a safe and supportive community without visible interventions that help set the social norms for the environment? I believe social norms also protect speech.

At dinner tonight ██████ said: my car videos are getting 100,000 views, it's natural that I'm going to get a lot of hate with that. Is it? Why is it acceptable for someone to harass someone on their surface? The most powerful solution for the integrity and safety space is to affect the supply of bad experiences via the actors creating them.

I might be wrong about my assessment, and welcome feedback about any effort or data that I'm missing. I believe that it is important to get the following efforts well-funded and prioritized:

- What is the content that is causing bad experiences for our users? How intense is the experience?
- What % of that content is policy violating? (i.e. how much of TRIPS is driven by content other than what drives Prevalence?)
- What are visible product solutions that make the community better over time? e.g. actor feedback, comment covers, pinned comments, etc.

The solutions we create I believe should have the following properties:

- The person who has the negative experience should feel heard, you don't 'perceive' racism or harassment, you experience it, and you are the source of truth for that. The feedback flow should not be just about filing a report, but about understanding the experience the person is having so we can give them the right solution.
- We should empower creators, communities, and Instagram, in setting the social norms for the spaces they are a part of.
- Where appropriate we should give feedback to actors, in the belief that they are acting with good intention and might have caused unintentional harm. There can be a range of interventions that start with 'nudges' that assume positive intention. This will allow us to separate the people who would behave differently given feedback, from the ones who are intentionally causing harm. We can then approach people who are intentionally malicious with the integrity tools.

If you would like I can give more details or specifics on this. I am appealing to you because I believe that working this way will require a culture shift. I know that everyone in m-team team deeply cares about the people we serve, and the communities we are trying to nurture, and I believe that this work will be of service to that.

Arturo

Message

From: Arturo Bejar [REDACTED]
Sent: 10/14/2021 11:56:05 PM
To: [REDACTED]
Subject: Fwd: Pre-read for our conversation tomorrow

Hi [REDACTED]

Sharing with you the pre-read of my conversation with Adam tomorrow, I will keep you posted.

Arturo

Begin forwarded message:

From: Arturo Bejar [REDACTED]
Subject: Pre-read for our conversation tomorrow
Date: October 14, 2021 at 4:40:38 PM PDT
To: Adam Mosseri [REDACTED]

Hi Adam,

In order to make the best use of our time tomorrow I put together a short pre-read that I've vetted with the team in well-being.

Data points (last 7 days/more than 7 days)

Have you ever received unwanted sexual advances on Instagram?

- 13-15 year olds: 13%/27%

Have you ever seen anyone discriminating against people on Instagram because their gender, religion, race, sexual orientation, or another part of their identity?

- 13-15 year olds: 26%/31%

Has anyone done any of these things to you on Instagram? Insulted or disrespected you, contacted you in an inappropriate way, damaged your reputation, threatened you, excluded you or left you out.

- 13-15 year olds: 11%/25%

Have you ever felt worse about yourself because of other peoples' posts on Instagram?

- 13-15 year olds: 21%/23%

Questions:

- What should be the goal/number of 13-15 year olds on each of the BEEF categories?

- Are users able to express these experiences to us in the product? (e.g. for unwanted sexual advances, or negative comparison you can't)
- What would we build if >90% of the content which drives these experiences is not policy violating or borderline?

If you'd like to look at the data directly, here is the data by age:

<https://docs.google.com/spreadsheets/d/10rR5hbK4v1W-2QmUUMSLGiUFfljgm9ngf->

I also find it helpful to put the data in the context of the questions (which convey better than our labels what people are experiencing):

https://docs.google.com/spreadsheets/d/1gOGpXK7UkC_Z7_C41Z8SLab4Puom4vv4Kfx0Zs6-

Recommendations

1. For Instagram to set goals based on TRIPS/BEEF, use people's experience as the north star for the work:
 1. What would you build if the goal was to get to 1% unwanted sexual advances? Or 3% witnessed hate? Or 2% target of bullying?
 2. Change the use of the word 'perceived' to 'experienced' - people don't perceive being harassed.
2. Provide features that help us understand the issues and content that people are experiencing so that we may develop interventions/features that help them and improve the community over time.
 1. Secondary actions to block/delete where we get user experience data. This has been difficult to date because the team has been running into XFN limitations on understand efforts.
 2. Make the reporting flow, or add experiences at the beginning to make people feel heard and supported with what they are experiencing, as well as generate insights on the issues they are having.
3. Invest in features that help us learn how to develop and maintain social norms, and actor feedback.

Can we shift the conversation into one of hope and leadership?

- Everyone in the industry has the same problems right now.
- Prevalence-based measures, while necessary, don't create a safe and supportive community, you're always behind the latest harmful thing.
- We have few visible features that help create a safe and supportive environment for everyone.
- There is a great product opportunity in figuring out the features that make a community feel safe and supportive.
- It is possible and important to work these issues in partnership with other industry leaders and academics. We have much learn about each of these issues. I believe is possible to help create public conversation on these topics for good.

A point which might be good for you to know (which I did not put in the document reviewed by the team) is that many employees I've spoken who are doing this work (and are of different levels) are distraught about how the last few weeks have unfolded. These are people who love FB/IG, and are heart/mission driven to the work.

Ideas for improving report flows

- **Infuse developmental science**
 - 13/14 year olds are different from high school and college students
- **Use more kid-friendly language**
 - “Report” vs. “This post is a problem”
- **Enhance logic of the flow**
 - ‘What happened?’ to ‘how are you feeling?’ to ‘what can you do?’
- **Differentiate the experience so we could tailor support**
 - Move from just “harassing me” to real experiences of this age group
- **Empower youth to take a positive and safe action**
 - Provide simple, effective guidance (e.g., “don’t be alone with this person”)
- **Help youth to get more help from their community**
 - Encourage kids to reach out to a trusted adult

Comparing Old and New Flows

- **Were users more or less satisfied with the new report flow?**
 - One concern was that kids would be less satisfied with the new flow compared to the old flow because the new flow was longer
 - There were no significant differences

	New Flow	Old Flow
How easy?	1.89	1.92
How helpful?	2.23	2.18
How comfortable?	2.23	2.17
How satisfied?	2.19	2.22

Comparing Old and New Flows

- **Did we change actual behavior? YES!**
 - Of those who completed the report (for more extreme instances), a greater number of users in the new flow reached out to a trusted adult

	New Flow	Old Flow
Reaching out to trusted adult	43%	19%
Blocking	28%	44%

The big question:
Can social media can
incorporate design
that integrates
emotional intelligence and developmental
science to
promote pro-social behavior
– both on- and off-line?

Two (seemingly) disparate fields

Emotional Intelligence

- EI introduced to psychology in 1990; reaches public in 1995
 - EI is the *ability* to reason with and about emotions to enhance decision making and promote both personal growth and pro-social behavior.
- Hundreds of studies demonstrating that EI is associated with positive outcomes for young adolescents
- Our EI program, RULER, has demonstrated positive results in shifting school climate and children's prosocial behavior

Technology/Social media

- Internet reaches the public in 1994
- Social media evolves out of the chat room and into popular networks
- Internet keeps getting blamed for social and psychological problems that are not new
- Facebook recognizes the potential power of integrating emotional intelligence principles into reporting systems

The life of a 13-14 year old

Young Adolescent Development

Biological Changes

- Onset of puberty leads to hormonal instability
- Executive network that allows self-regulation, planning, and overall monitoring, are “under development”
- Social excitement literally overwhelms the ability to control behavior.

Cognitive Changes

- Improvements in thought complexity makes kids more vulnerable to what others think. “Imaginary audience” (thinking that everyone sees them) makes them especially self-conscious and vulnerable to embarrassment.

Self and Identity

- Separation/individuation from parents; peer group offers temporary identity so they can become “autonomous”
- Young adolescents are especially sensitive to peer relationships – power dynamics and increased risk-taking especially in presence of peers.

Overview

- The original report flows (13-14 year olds)
- Infusing emotional intelligence
- What we learned from v1.1
- Version v2.0
- What the data reveal
- What's next?



The original report flows

Is this post about you or a friend?

Yes, this post is about me or a friend:

- I don't like this post
- It's harassing me
- It's harassing a friend

No, this post is about something else:

- Spam or scam
- Hate speech
- Violence or harmful behavior
Choose a type
- Sexually explicit content
- My friend's account might be compromised or hacked

Continue **Cancel**

What You Can Do

If you are in physical danger, please contact a local authority right away.

- Block Jake Brill**
You and Jake will no longer be able to see each other or connect on Facebook
- Get help from an authority figure or trusted friend**
Forward this post to someone who can help you in person


Report to Facebook **Continue** **Cancel**

Send Message

Enter the Facebook friend you want to contact here. If the friend is not on Facebook, you can enter an email address.

To:

Message:

 **Jake Brill** ▶ **Clive Jakob**
[@272:0]
February 8 at 7:04pm near San Francisco · 🌐

Continue **Cancel**

What You Can Do

Is your friend in physical danger? If so, please report this threat to a local authority.

- Message Clive Jakob to remove**
Ask Clive to remove the video
- Unfriend Clive Jakob**
- Block Clive Jakob**
You and Clive will no longer be able to see each other or connect on Facebook

Report to Facebook **Continue** **Cancel**

Infusing emotional intelligence

- **Takeaways from initial focus groups and interviews**
 - Kids were particular about the language we used
 - E.g., *report* – meant ‘authority’ or ‘trouble’ or ‘evaluated,’ whereas ‘get help’ suggested ‘technical problem’
 - Kids helped us to differentiate bullying and non-bullying experiences
 - Kids wanted Facebook to do something about it, but were not sure what that was; wanted a ‘conversation’
 - If questions were meaningful, specific, and helpful, they would be more motivated to complete the flow
 - Kids said they wanted help crafting messages
 - Kids did not believe everything needs to be reported b/c they would just tell (call, text) someone they trusted

Infusing emotional intelligence

- **Takeaways from interviews with parents**

- Parents were mixed on whether they should be the 'trusted' adults
- Some parents enabled kids to fake their age
- If their child was threatened, they wanted to know
- Parents wanted more resources for their kids

- **Our own takeaways**

- Had to be a balance between what kids wanted and what we believed they need
 - E.g., Threatened – may not want to tell trusted adult, but they need help
- A conversational approach was ideal
- We needed to provide children, parents, and educators with more direct help

Infusing emotional intelligence

- **Infuse developmental emotion science – more adolescent-friendly language, enhanced logic, more relevant)**
 - 13/14 year olds prefer “This post is a problem” to “Report”
 - ‘What happened?’ to ‘how are you feeling?’ to ‘what can you do?’
 - Move from just “harassing me” to “saying mean things to me”
- **Integrate emotional intelligence**
 - How did the post/photo make you feel? (both emotion and intensity)
- **Empower youth to take a positive, safe action both on- and off-line**
 - Provide simple, effective guidance for less versus more threatening posts
 - Develop positive pre-populated messages to content creator/trusted adults or friends

The Present Study

Version 2.0

DEMOGRAPHICS

What we learned from v1.1

- **Most reports were about ‘self’ as opposed to others**
- **Most kids just want to be ‘untagged’ from posts/photos**
- **Photo and post report systems needed to be separated**
- **We wanted to increase messaging to content creator and trusted friends/adults and decrease blocking/unfriending**
- **We needed to improve pre-populated messages to help teens communicate with content creators and trusted friends and adults,**
- **We also wanted to help trusted friends and adults communicate with the reporter**
- **We wanted to increase completion rates**
- **Gender was a variable that needed to be explored**

Discussion

- **Gender matters**
 - Reporting behavior – girls report more than boys
 - Bullying behavior – girls are more likely than boys to be the ‘content creators’
- **Embarrassment is most frequent emotion associated with photos**
 - Kids are self-conscious about the way they look
- **Anger is most frequent emotion associated with posts**
 - Kids “say mean things” which is perceived of as an injustice
- **Emotion intensity is associated with behavior (messaging)**
 - Emotions drive decision making and action
- **Providing kids with a more emotionally intelligent report flow helps to have more positive interactions**
 - Kids are more likely to “stay in the relationship” and make constructive decisions like sending positive messages as opposed to blocking
 - In essence, we have eliminated ‘blocking’ – likely an ineffective coping strategy