

Appearance comparison: Defining sensitive ("High-NAC") content using FIT and X-Ray

██████████, ██████████, ██████████, ██████████ and ██████████

- We identified 37 FIT subtopics (e.g., eye liners, skin care) and 15 X-Ray concepts (e.g., women's swimwear, chest focus) that are sensitive: greater exposure to content with these features is associated with more negative appearance comparison (High-NAC).
 - Content that was sensitive for women was also sensitive for men. Men saw a lower proportion of sensitive content, but those who see it more have higher levels of appearance comparison, just as women do.
 - On average, 14% of the content that people saw on IG was from one or more of these sensitive categories (18% for women / teen girls).
 - Using survey responses of comparison, we found that when people saw more than 11-13% sensitive content, the majority reported experiencing appearance comparison at least half the time. By that measure, approximately 70% of teen girls may see "too much" sensitive content.
 - While people who see more sensitive content are likely to experience more negative comparison, we don't know if sensitive content causes comparison, even though qualitative research implicates them.
- "Content that was sensitive for women was also sensitive for men. Men saw a lower proportion of sensitive content, but those who see it more have higher levels of appearance comparison, just as women do"
 - "On average, 14% of the content that people saw on IG was from one or more of these sensitive categories (18% for women/teen girls)"
 - "Using survey responses of comparison, we found that when people saw more than 11-13% sensitive content, the majority reported experiencing appearance comparison at least half the time"

- We can use this sensitive content definition to:
 - Measure how much sensitive content specific top accounts (or classes of top accounts, such as fashion models) produce in order to inform decisions about account recommendations
 - Understand prevalence of sensitive content by surfaces (e.g., Explore)
 - Measure changes in views of sensitive content in experiments as a proxy metric (along with survey measures of comparison, which may take longer to move)
 - Build a sensitive content classifier (once we have content-level comparison labels, which is in progress)

Introduction

When people compare themselves to others they see on Instagram, they often feel worse about themselves. On the Mental Well-Being team, we're identifying ways to help people, particularly teens, reflect on and take control over these experiences, such as nudges to unfollow accounts that may elicit comparison.

As part of this work, we first need to identify which kinds of content induce social comparison, referred to as "sensitive content" here. To do this, we paired [survey reports](#) of appearance comparison with characteristics of the content people saw in the two weeks prior to the survey. Specifically, we've examined the Facebook Interest Taxonomy (FIT) subtopics (e.g., Makeup & Cosmetics, Jeans & Denim) and X-Ray concepts (image features, such as "chest focus") that were most strongly associated with appearance comparison after controlling for viewer characteristics.

Sensitive FIT subtopics

As reported in a [previous note](#), FIT subtopics related to beauty and globally western pop music were associated with higher levels of appearance comparison. Because our [qualitative research](#) implicated body-focused content in comparison but not music, we focused specifically on the FIT subtopics within the "Fashion & Style" parent topic – which includes concepts like beauty, makeup, and clothing – that were statistically significantly associated with appearance comparison in regressions controlling for viewer characteristics (country, gender, whether they're a teen, following count, followers, L28, and time spend on IG). We do not consider the music subtopics as sensitive *per se*; the association with comparison is likely due to cases of some pop stars emphasizing their bodies, which the X-Ray features in the next section capture directly.

List of sensitive FIT subtopics. There were 37 FIT subtopics within Fashion & Style that were statistically significantly associated with appearance comparison (more details in the [FIT topic/subtopic notebook](#)). Figure 1 includes a few examples and you can browse more in the [FIT subtopic dashboard](#):

- Makeup & Cosmetics

- Lipstick
- Makeup Artists
- Beauty
- Eye Liners
- Jeans & Denim
- Eye Makeup & Cosmetics
- Skin Care
- Eye Shadows
- Fashion Modelling
- Outerwear
- False Eyelashes & Adhesives
- Hair Coloring
- Shirts & Tops
- Fashion Photography
- Costume Makeup & Makeup Art
- Foundation (Beauty)
- Makeup Brands
- Face Care
- Hair Styling Tools & Products
- Hair Removal
- Boots
- Face Makeup & Cosmetics
- Clothing & Accessories
- Lip Makeup & Cosmetics
- Sleepwear & Loungewear
- Celebrity Style
- Hair
- Purses
- Piercings
- Hair Care
- Nail Art
- Hair Extensions
- Modeling & Photography
- Body Modification
- Tattoos & Body Art
- Lingerie



Figure 1. Examples of content with sensitive FIT subtopics within the "Fashion & Style" top-level topic. Subtopics include makeup & cosmetics, lipstick, celebrity style, skin care, and outerwear.

Within Fashion & Style, a few FIT subtopics were linked with lower levels of comparison, and others had no statistically significant effect detected. Specifically, topics related to weddings ("Bridal Fashion," "Wedding Hairstyles," and "Wedding Fashion & Beauty") were linked to lower levels of comparison. This finding is surprising since romantic relationships are a common vector of comparison, but it might also reflect a selection effect in which viewers of wedding-related content were themselves about to get married and felt more assured in their relationships. Notably, while "Formal Wear" was associated with lower comparison among women, it was associated with higher comparison among men and teen boys. [Internal research indicates](#) that teen boys talk more about the fit of their clothes rather than the shape of their bodies. Fashion subtopics with no statistical association with comparison were typically pieces of clothing: earrings, watches, jewelry, hats, scarves. (Full list of [subtopics](#).)

Outside of Fashion & Style, many FIT subtopics were associated with lower comparison: most had to do with being outside. See the full list below.

Sensitive X-Ray concepts

X-Ray (Cortex) concepts are based on the images themselves and come in three flavors: [XRAY_2018A](#) ("gold-standard" concepts like *face*, *blond*, *smiling*), [XRAY_2018A_EXPLORATORY](#) (other concepts such as *closeup*, *selfie*, *woman*), and [XRAY_OC](#) (concepts relating to objectionable content such as *chest focus*, *facial sexual expression*). Here, we focus on OC concepts as these concepts have already been flagged as likely policy violating or borderline.

Because our analyses are correlational, not causal, we are being conservative in defining sensitive content. Many of these OC concepts also align with qualitative findings on social comparison in that they emphasize body parts rather than the whole person, their context, and their journey. As we collect content-level survey data ("how does this post make you feel") and session-level data ("how did this session make you feel") rather than person-level survey data ("how do you usually feel using IG"), we may expand to all OC categories and weight features using a classifier instead.

List of sensitive X-Ray concepts. There are 15 OC X-Ray concepts that are statistically significantly associated with appearance comparison (more details in the [X-Ray notebook](#)). Figure 2 shows some examples and you can browse more in the [X-Ray dashboard](#) (use caution; images in some of these categories are disturbing). Descriptions of many of these concept classifiers are in the [content integrity wiki](#).

- chest focus
- penis
- women's swimwear
- adult nudity & sexual activity (ANSA) borderline
- ANSA violating
- ANSA violating ID photo
- suggestive external image
- facial sexual expression
- breast covered with hand
- nudity at a protest
- vagina
- ssi
- suicide
- delete-worthy gv video
- violence non-medical vs. medical



Figure 2. Examples of content with sensitive X-Ray concepts (women's swimwear, chest focus, ANSA violating, external suggestive image).

Other potentially useful X-Ray concepts. If we also consider concepts that have *not* been flagged as likely policy-violating or borderline, multiple other X-Ray concepts were also significantly associated with greater negative appearance comparison ([full list](#)). These include:

- Closeup
- Selfie
- Woman
- Hair
- Face
- Mirror
- [Eugenia](#) (largely closeups of women)
- Blond
- Sunglass
- People

The above ten concepts are all more strongly associated with negative appearance comparison than any of the OC concepts. Still, these topics are relatively broad and affect a large amount of content. Perhaps to a greater extent than for the OC concepts, understanding the implications of filtering or targeting based on these broader concepts (e.g., on creators) remains important future work.

Content that was sensitive for women was also sensitive for men

How much of this content do people see?

On average, 14.1% of the content that people see comes from one of these sensitive FIT subtopics [1]. Women and teen girls see the most, with roughly 1 in 5 pieces of content they see coming from a sensitive FIT subtopic (18.7 and 18.2%, respectively), while teen boys see the least, at 9.9% (Figure 4). For the sensitive X-Ray concepts (Figure 5), differences in prevalence between subgroups (e.g., women vs. men) are smaller. (The prevalence rates for sensitive X-Ray concepts are likely overestimates as we used a relatively low threshold for identifying content as containing these concepts.)

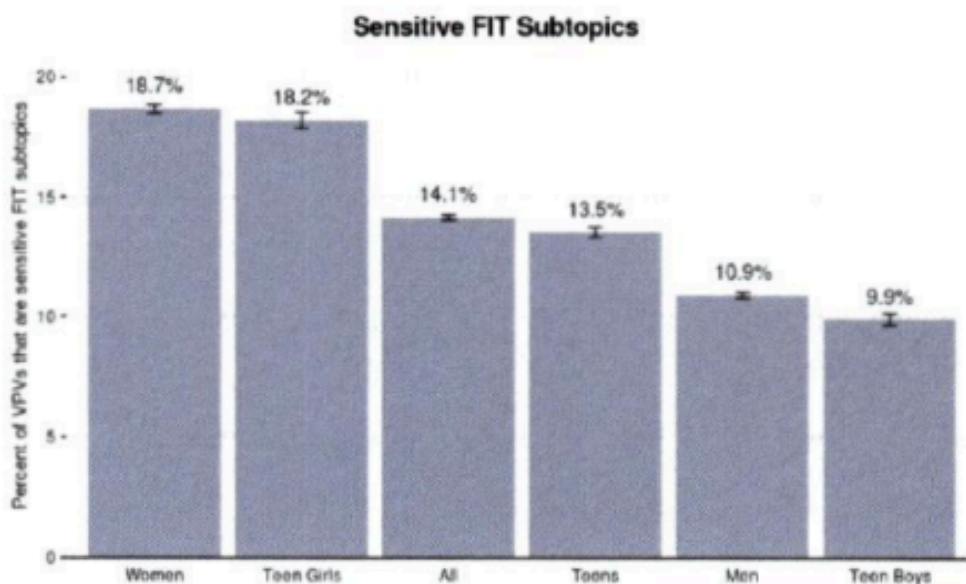


Figure 4. Average fraction of content seen that is from a sensitive FIT subtopic.

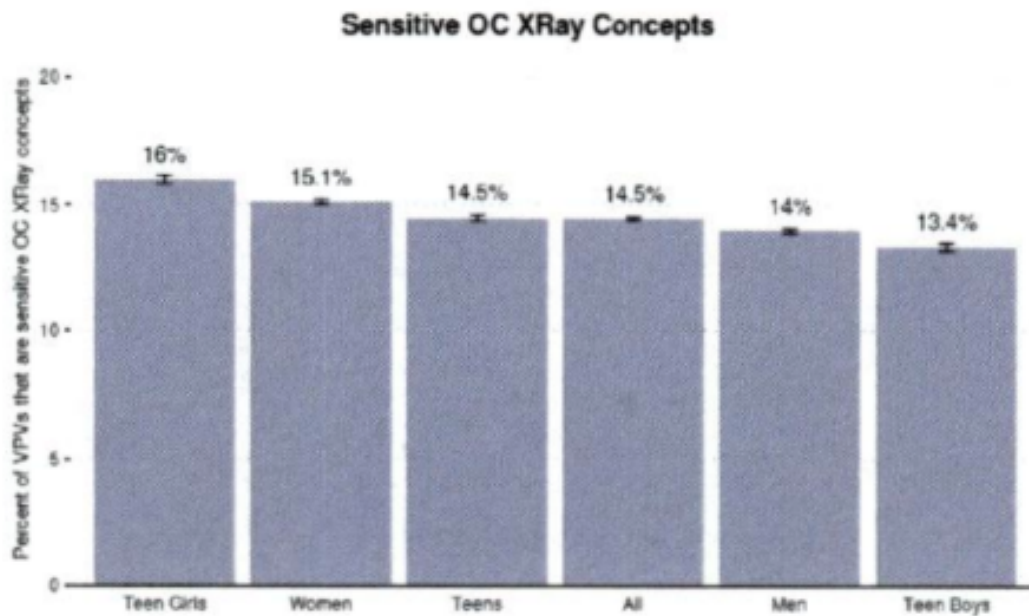


Figure 5. Average fraction of content seen that is from a sensitive OC X-Ray concept.

How much sensitive content is “too much”?

Some of the qualitative research suggests that individual pieces of body-emphasizing content may not trigger social comparison among viewers, but seeing [multiple pieces in a session](#) can.

And while individual tolerances vary, we wanted to understand whether there are turning points in this sensitive content beyond which the majority of viewers report experiencing comparison.

Thus, we turned to our appearance comparison survey responses. As described in the [initial note in this series](#), respondents answered three questions about appearance comparison (how often they see content on IG that makes them feel worse about their bodies or appearances, how often they compare themselves to others on IG, and how much pressure to be perfect they feel). These responses are combined into a single person-level “appearance comparison” measure, on a scale from 1 to 5. Here, we binarize the outcome such that people with a mean score of 3 or higher (those who experience comparison at least “half the time” on average) are considered to experience issues with comparison, and those with mean scores below 3 are not.

Then we examine how experiences of comparison vary based on the fraction of content people see that was sensitive (see [sensitive-content thresholds notebook](#)). Figure 6 shows this relationship for teen girls. Each point is one decile of teen girls based on how much sensitive content they saw (x-axis). The y-axis is the fraction of them that reported issues with comparison. One way to think about “too much” sensitive content is the point at which the majority of teen girls experience comparison, which happens around 11-13% sensitive content. Since seven deciles see at least this much sensitive content, **approximately 70% of teen girls may see “too much” sensitive content.**

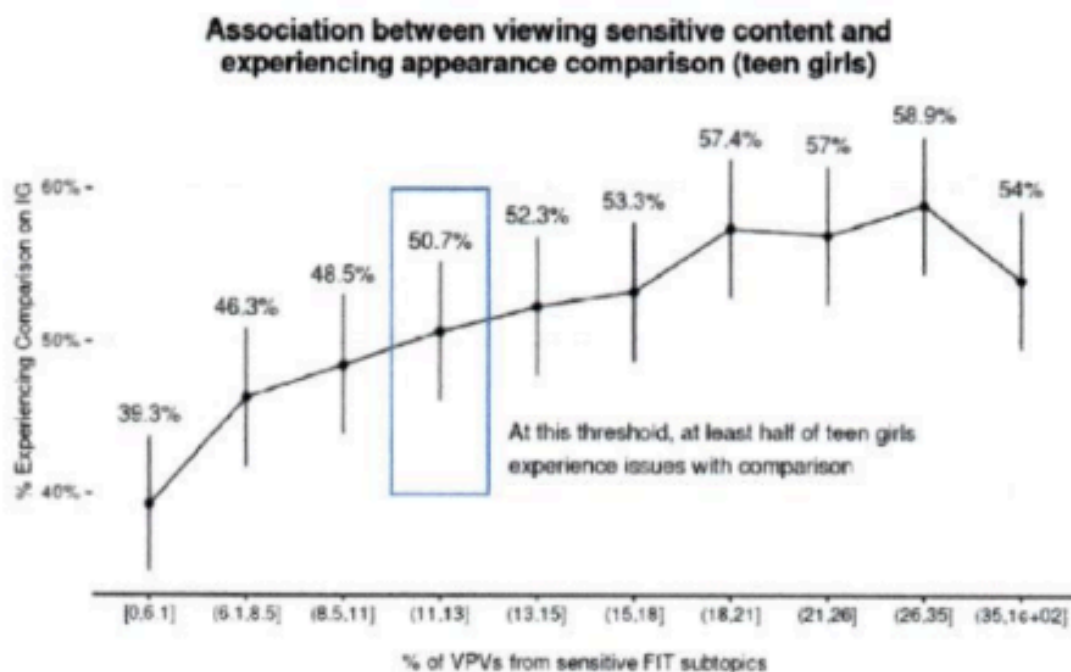


Figure 6. Among teen girls who see 11-13% sensitive FIT subtopics or more, at least half of them report experiencing problems with appearance comparison (experiencing it at least “half the time”).

However, we should be cautious in applying these thresholds since individual tolerances vary, and we don't know that these specific topics caused comparison, just that seeing them is associated with higher levels of comparison and that qualitative research implicates them. As a result of doing these analyses, we determined that for our upcoming “Update Following” experiment, we don't need to target teens based on the fraction of sensitive content viewed. Because negative comparison affects such a large proportion of teens, our “precision” (nudging those who experience comparison) is only slightly lower when targeting all teens (51.7%), rather than targeting those who view more than 11-13% sensitive content (54.8%).

What types of content can we recommend instead?

Here, we examine the X-Ray concepts and FIT subtopics that were statistically significantly associated with reductions in negative appearance comparison. People who saw more of this content felt less pressure to be perfect. Most of the content relates to being outdoors, playing sports, eating, art, or music. This content is less likely to emphasize physical appearances. Some concepts, such as food, may still be sensitive for individuals with eating disorders.

Do these concepts just not have people in them? The X-Ray concept of “people” is associated with greater appearance comparison generally, but just because a photo has a person in it doesn’t mean it induces comparison. For example, [sports photos with and without people were associated with reductions in comparison](#), as were photos with and without people about games, religion, vehicles, politics, crime, TV, and history. On the other hand, fashion content with people is linked to negative comparison, but fashion content without people is not linked to comparison. Some topics are always associated with comparison, regardless of whether a person is in them: relationships, performing arts, and music.

X-Ray concepts and FIT subtopics associated with significantly lower levels of appearance comparison, taking into account viewer demographics:

	<ul style="list-style-type: none"> • Ball • Stadium inside • Quarterfinal • Court • Basketball 	
		FIT topics (high-level):
		<ul style="list-style-type: none"> • Visual arts, architecture, & crafts • Games, puzzles, & play • Sports • TV & movies • Vehicles & transportation • Travel & leisure activities • Animals & pets • Food & drink • Books & literature • Religion & spirituality • History & philosophy • Politics • Business, finance, & economics • Crime & tragedy (includes humanitarian aid)
Outdoors:	Food:	
<ul style="list-style-type: none"> • Grass • Outdoor • Nature • Cloud • Sky • Catamount • Tree • Wildlife • Flower 	<ul style="list-style-type: none"> • Cake • Meat • Beef steak • Espresso maker • Confiture • Crustacean 	
	Art & Music:	
	<ul style="list-style-type: none"> • Comics • Animation • Anime • Instrument 	
Sports:		
<ul style="list-style-type: none"> • Soccer 		

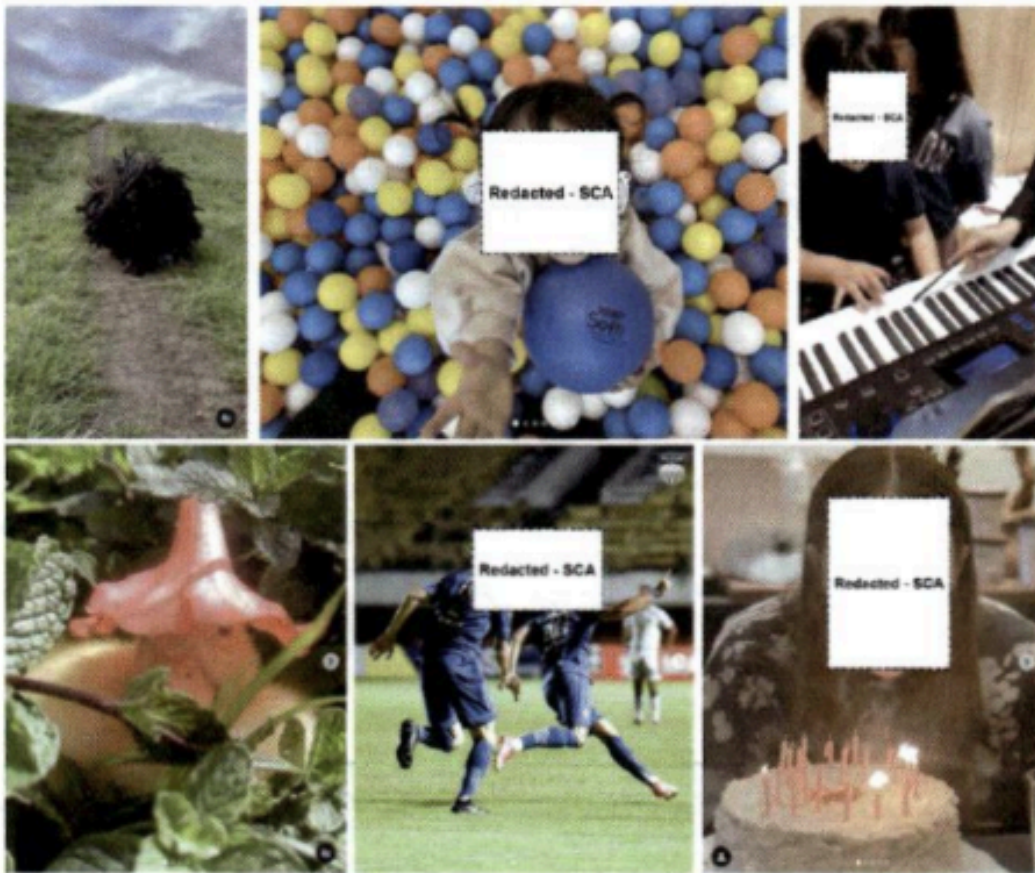


Figure 7. Content with X-Ray features that were associated with reduced appearance comparison, including *outdoor*, *ball*, *instrument*, *flower*, *soccer*, *cake*.

What can we use this sensitive content identification for?

Combining these findings with qualitative research in order to build conviction that these sensitive topics cause appearance comparison (not just that they're correlated with it), we can apply these findings in the following ways:

- **Unfollowing accounts:** Measuring how much sensitive content specific top accounts (or classes of top accounts, such as fashion models) produce. This allows us to think about whether to encourage teens to consider unfollowing interest accounts that make

them feel worse about their bodies. We're planning to test this with "[Update Following](#)," and are currently evaluating offline whether simpler sort orders for accounts to unfollow such as "Most seen" would rank accounts similarly to sorting by sensitive content production.

- **Following accounts:** Similarly, we can take the proportion of sensitive content top accounts (or classes of top accounts) produce into account when recommending accounts to follow, either to new or existing users.
- **Opportunity sizing:** Understanding the prevalence of sensitive content by surface (e.g., Explore) will allow us to build features for surfaces where these features are likely to have greater impact.
- **Building a classifier:** Once we have content-level labels about comparison (e.g., "this post makes me feel worse about my appearance or body") rather than viewer-level labels (e.g., "I usually feel a lot of pressure to be perfect on IG"), we can expand the feature set to build a sensitive content classifier, which would allow us to build a sensitive control that users could opt in to.
- **Metrics:** While our goal is to reduce negative experiences of comparison on Instagram, which would be most appropriately measured through surveys, sensitive content views could be used as a supplementary metric in experiments, especially in cases where feelings of comparison measured in surveys may lag several months from experiment start.

Next steps and limitations

Fairness: As we think about the impact this sensitive content has on viewers, we also need to consider the impact that any changes we make would have on creators. We're beginning to pull data about the kinds of creators that disproportionately produce sensitive content in order to evaluate fairness.

Causal conviction: These FIT subtopics and X-Ray features are correlated with appearance comparison, but we don't have evidence of causation. We have multiple qualitative and quantitative studies that implicate these features, but we're thinking about causation in two ways.

- **Content-level survey:** First, we're launching a content-level survey, asking people how specific posts make them feel. This provides more granular content-level conviction for our features. One challenge is that individual pieces of content may not trigger comparison, but rather a large collection of content may.
- **Session-level survey:** Thus, we'll be comparing rates of comparison in the content-level survey and a separate post-session survey, in which people report how their whole session made them feel.

Experiments: By running experiments, measuring both sensitive content impressions as well as negative comparison via surveys, then examining the relationship between the two, we'll gain more evidence for causality.

Footnotes

1. This is among content that has any FIT Subtopic classification at 80% precision, which was about 26% of the VPVs among our survey data. At the time of data collection, content had to receive at least 100 views to be classified by FIT, and not all of those had p80 subtopic classifications. (Now content only needs 10 views for classification.) Since we don't know the nature of the non-classified VPVs, take the estimates of sensitive content viewed with a grain of salt.
2. We are examining a large number of models and thus some are likely to be statistically significant at random. We partially account for this by using a more conservative threshold ($p < 0.01$), but since this work is exploratory and we have lower power among smaller subgroups (e.g., women who saw any specific subtopic within a two-week period), we're not doing additional corrections for multiple comparison here.

 21

7 Comments

 Like Comment Share

[REDACTED]
This is awesome work -- I'd love to sync up sometime to see if I can help connect this to the Interests work on IG/FB!

Like Reply 45w

 2

[REDACTED]
cc: [REDACTED]

Like Reply 45w



[REDACTED]
Cc [REDACTED]

Like Reply 45w



[REDACTED]
cc: [REDACTED]

Like Reply 45w



[REDACTED]
cc: [REDACTED]

Like Reply 45w



[REDACTED]
cc: [REDACTED]

 2

Like Reply 45w



[REDACTED]
Have you considered interventions that target the creators of sensitive content with education or nudges? They may be amenable to suggestions that make their content less likely to trigger comparison.

Like Reply 42w

 3

Write a comment...



Press Enter to post.