

Public Comment on Gender Identity Debate Videos

The Oversight Board has selected two cases regarding gender identity and access to gendered spaces such as bathrooms and sporting competitions. I write to provide my expertise on content moderation, safety, and free expression regarding these sorts of content. My comments will discuss how further restrictions on speech will create expansive and complicated rules that do not support civil discourse. Instead, greater expression is needed in the debate around these issues for our society to better understand each other, achieve progress, and engage in the messy debates necessary for self-government.

Debating Identity

The announcement of this case specifically makes note of how in each piece of content the poster refers to the transgender individuals as males. The Board defines such references “to a person using a word, especially a pronoun or the way in which they are addressed, that does not reflect their gender identity” as misgendering. Currently, misgendering and references to others’ gender identity does not violate Meta’s Community Standards in all cases but does violate if used to discuss a minor’s gender identity. Regardless of how one feels about misgendering, a decision to label it a violation would significantly limit discussion of major questions of identity.

The question being posed is whether referring or not referring to others as certain protected characteristics is considered hate speech or bullying on Meta’s platforms. While the content in question involves gender claims, the Board’s decision must address the broader principle: does referring to someone as a certain protected characteristics or refusing to refer to someone as a certain protected characteristic comprise a violating attack? Should, for example, Catholics be able to refer to Protestants as non-Christians or should Protestants be allowed to reject Mormons as Christians? Should Ukrainians be allowed to refer to Crimeans as Ukrainian or not Russian? Should it be considered hate speech to claim that if someone votes a certain way, they are not black?ⁱ Should references to Hispanics as Latinx be considered hate speech or bullying given many Hispanics are at least somewhat offended by the term?ⁱⁱ There are countless ways people disagree about matters of identity. Shutting down such conversations is to silence fundamental social discussions of what it means to be part of a community or to participate in a group. People are passionate about matters of their identity on all sides of an issue. But forbidding discussion on these issues— or forbidding only certain perspectives— does not actually help groups resolve their differences or foster greater understanding and acceptance.

And while the definition of misgendering considered in this case is a common one, it reflects a preference around speech on issues of gender that is not shared by large parts of society. For example, recent polling in the U.S. shows that a 60-38 majority of U.S. adults say gender is determined by sex at birth.ⁱⁱⁱ For those in this 60%, they do not believe it is misgendering to refer to someone according to their biological sex. Instead, many in that group may believe it is “misgendering” to refer to individuals according to their gender identity and may be acting without any sort of malicious intent to an individual or group. While the Board may believe its view of misgendering is correct, codifying it would deter important conversations that move to a fully developed understanding of complicated issues and foster greater understanding among individuals of different viewpoints.

If the Board were to decide in this case to limit speech discussing gender identity, it would be establishing a growing precedent that certain protected characteristics or marginalized groups have

greater protections than other characteristics or groups. For example, the Board asked Meta in its *Post in Polish Targeting Trans People* case to consider attacks against transgender flags as attacks on transgender people, effectively enshrining a blasphemy or sacrilege law that only protects transgender flags and symbols from being attacked.^{iv} Such decisions create unequal rules, shunning equality in favor of a preferred point of view that is far from universally held in the countries where Meta operates. While private companies are free to engage in viewpoint discrimination, the Board often looks to legal and human rights principals to guide its decisions. Viewpoint discrimination is unconstitutional if done by governments under American conceptions of free expression, and conflicts with the universal liberal values of equality and non-discrimination enshrined in international human rights law.

A viewpoint specific approach by the Board also risks creating an increasingly complex caste system of policies based on the various protected characteristics that would further confuse users. In some cases, burning a flag is protected, but in others it is hate speech. For some groups, making identity claims is allowed, but for others it is verboten. Giving users a greater voice and treating them fairly through understandable policies are two goals that the Board should strive for whenever possible.

Expression Regarding Restricted Spaces

Another major question that the Board has raised in this case is how we should handle content regarding exclusion of transgendered individuals from bathrooms and sporting activities or spaces. This is an area that I am extremely familiar with as I was one of the policy members that has reviewed this area of the policy in the past. It is a complex issue, but current policies limit expression from certain viewpoints that reflect exclusion that isn't clearly hateful but based on widely accepted categories and offline realities.

The first consideration is that the content in question does not appear to have any explicit calls for exclusion. While we cannot see the content's exact words, the poster of the content appears to be asking about why government policies permit transgender individuals to use bathrooms that belong to their gender identity rather than biological sex. Similarly, the second post appears to have the spectators vocally disapprove of the competition including a transgender athlete, such disapproval is not explicitly calling for exclusion. In either case, if the Board were to consider such policy questions or unhappiness as violating, then vast amounts of speech would become violating and would be difficult for reviewers to handle consistently. If this content were to be deemed violating, this policy would grow beyond explicit exclusion to mere disapproval that may be interpreted or felt as being exclusive.

But such a proposed standard is very murky. Would a video with few boos count as sufficiently exclusionary? Would complete silence in protest of a transgender athlete be deemed as insufficiently praiseworthy and thus exclusionary. Would taking a knee or making an XX symbol (representing the XX chromosomes of biological females) be hate speech? Would pure policy discussions expressing concerns about the inclusion of transgender individuals in gender specific spaces by legislators on the campaign trail, with constituents, or in official legislative business be automatically deemed hateful? And if it is unacceptable to even implicitly make transgender individuals feel excluded by allowing conversations around the definition of gender for gender exclusive spaces, then such logic would extend to limiting speech around Meta's existing policy lines that allow gender restrictions in cases of health or positive support groups. In other words, while the policy currently allows a breastfeeding Facebook group or sexual assault survivors group to limit its membership to only certain genders, the logic of removing implicit exclusion would militate against allowing any gender specific spaces for being explicitly hateful and exclusionary. In essence, the rule that Board could adopt here is an expansive "prohibition of

disapproval,” which would demand that reviewer and algorithms be expected to remove any content that is not sufficiently positive or praising of transgender athletes or that raises any concerns about transgender individuals' inclusion in women's spaces or groups.

The second major consideration is whether the existing exclusion policy is properly framed to allow discussion of basic social and biological realities. While exclusion can reflect clear animus against a group of people, there are many places in societies where we simply accept that exclusion is not hateful but based on essential categories. For example, different religions may not allow members of other religions to enter their physical space or they may reserve different parts of their physical spaces for women and men. Stores may offer a discount or special hours to seniors, individuals with disabilities, or students based on their age or status. At times, for various privacy and safety reasons, almost all societies restrict access to certain gendered spaces.

As noted above, Meta's policy even acknowledges this in part, saying that exclusion is allowed for things like health and positive support groups. We understand that women may want to only be surrounded by biological women when discussing issues of rape or sexual assault. But despite sporting events and bathrooms having similar privacy and safety rationales to health and positive support groups, they are not included in Meta's existing policy. Instead, Meta's policy favors a specific ideological view of sex and gender that restricts voice and favors transgender individuals access to gendered spaces in ways that may be detrimental to others' privacy, safety, and rights. In general, while the policy against exclusion may address some hateful speech and be intended to create a more inclusive environment, it is already limiting significant types of speech and community that reflect offline realities. As such, the Board should recommend that Meta allow more expression currently classified as hateful exclusion through at scale or on-escalation policies when such distinctions are commonly and broadly justified. Additionally, Meta should be public and explicit about the distinctions its draws on these issues to be as fair and transparent to users as possible.

Challenges to Democratic Discourse

The discussions involved in this content directly relate to active political debates across the U.S. and around the world. To limit their discussion on Meta by establishing a more restrictive standard could eliminate an important forum for political debate and the growth of understanding.

If the Board decides to take down this content or recommend policies that remove more speech regarding these actively debated social issues, the Board will be placing itself directly at odds with democratic governance and the values of promoting civil discourse in the world. For example, it would directly impede discussion of constantly evolving regulation of sex and gender in schools through Title 9 of the Education Amendments Act. It would prevent fulsome debate of federal and state level policies as citizens and legislatures debate making further changes to the existing sharp split between the states—26 states currently require transgender athletes to participate in sports according to their biological sex and 24 states do not have such requirements.^v Meta's platforms will silence one side of the many lawsuits that have been and will be filed on this topic. The same dynamic is true in many other nations' politics. In sum, the ability for people and government officials of all viewpoints to weigh in on these issues being discussed and handled by legislatures, courts, executives, and civil society will be heavily restricted, with only one side able to freely speak on Meta's platforms. This impact on democratic

governance is only further heightened since the view being suppressed is the majority one as multiple polls both in the U.S. and abroad show growing and strong majority support for limiting sports, bathroom, and prisons based on biological sex.^{vi}

Given the significant implications for democratic discourse, a voice critical decision poses serious risk to society, the Board, and Meta. If most of the world cannot express its voice on a topic of significant social and political importance, then Meta's value of voice, the purpose of Meta's platforms, and the mission of the Board will have been seriously undermined. The Board will insert itself and Meta into a major cultural and political issue rather than favoring greater voice and expression. Such a decision will unfortunately empower those who believe that social media companies cannot be trusted to regulate themselves and put Meta in a difficult position. The Oversight Board may be praised by certain communities, but voice will suffer, and most will reject the legitimacy and validity of such a decision.

This is not to say that the majority is always right or that what may be acceptable today won't one day be found abhorrent in polite society. Majorities can be wrong and indeed some societies in the world still actively persecute and criminalize LGBTQ identities and activities. However, there are also many evolving debates based on logic, biology, women's rights, non-hateful but deeply held religious views, etc. that should be given the opportunity for discourse on Meta's platforms. Even if such views were a minority, they deserve to be contended with given the significant social implications of this issue and because opportunities to engage in conversation rather than being blacklisted are more likely to bring about societal change. If the view that gender identity is not correlated to biological sex is correct, then it's proponents should engage with challenges to its logic and consequences so that they can prove it to be correct and persuade others of its merits so that this view will win out in the marketplace of ideas. Whether it was American women's fight for suffrage,^{vii} the U.S. Civil Rights movement,^{viii} or the gay rights movement,^{ix} each of these movements succeeded due to free expression, peaceful protest, and direct debate and engagement with opponents. This is how liberal societies debate and reach conclusions.

In his prescient "Kindly Inquisitors," Jonathan Rauch argued that the humanitarian and protective impulse underlying these cases may seem appealing and noble- after all who wants to allow offensive, hurtful speech? But "in the pursuit of knowledge...will be hurt, and that this is a reality which no amount of wishing or regulating can ever change. It is not good to offend people, but it is necessary. A no-offense society is a no-knowledge society."^x Censorship cannot convince anyone of the rightness of a cause— if anything it suggests that the cause is incapable of winning in the marketplace of ideas. Removing speech that is critical of transgender access to women's spaces is no different in logic than the Catholic Church suppressing a heliocentric view of the universe during the Renaissance, southern states in the antebellum U.S. outlawing pro-abolition speech, or WWI authorities arresting suffragettes. Such suppression cannot achieve social change, but history shows us that it breeds resentment and further social conflict. Despite having robust and growing hate speech laws, Europe has not managed to quash the specter of hateful ideals across the continent. Former ACLU president and free speech expert Nadine Strossen reviewed the literature regarding the efficacy of hate speech laws and found they consistently pointed to no correlation between hate speech laws and a decline in hateful speech or behaviors. On the contrary, Strossen finds some evidence that hate speech laws are instead exacerbating conflict between different groups, weakening individuals' ability to grow stronger in response to perceived intolerance, and have many other harmful social effects.^{xi} While most high profile examples of suppressing critical speech in history refer to government censorship, Rauch also cites the Japanese cultural tradition against public criticism for fear of being viewed as rude, offensive, or antagonistic. Since public criticism was

strongly frowned upon, the result was stagnation rather than innovation that produced very few Nobel Prizes for Japanese scientists despite having a large, rich, and highly educated population.^{xii} The Board is not a government and Meta's policies are not laws, but both government and civil society intolerance of critical speech can have a negative impact on civil discourse and progress.

Conclusion

Sex and gender are major parts of our identity, so it is natural that debates around these identities are passionate and deeply personal for many people. While there will be assertions that misgendering or calls for limiting women's spaces based on biology are so harmful and hateful they need to be suppressed, such an approach would be ineffective in stopping hate, counterproductive in driving progress, and harm norms that encourage free expression and civil discourse. To better our society, we need more speech on these difficult topics rather than shutting down debate, even if it makes us uncomfortable. And while there will be individuals who are purposefully trying to offend and hurt others, many other aspects of Meta's policies already cover such speech. Thus, the Board should defend allowing speech that debates issues of gender identity and increase protections for this speech rather than creating further restrictions.

I make these comments in my individual capacity.

ⁱ Eric Bradner, Sarah Mucha and Arlette Saenz, "[Biden: 'If you have a problem figuring out whether you're for me or Trump, then you ain't black'](#)", CNN, May 22, 2020; Emily Jacobs, "[Tim Scott hits back against 'Uncle Tim' slur](#)", New York Post, April 30, 2021 .

ⁱⁱ "[The use of 'LatinX' among Hispanic Voters](#)", Bendixen & Amandi International.

ⁱⁱⁱ Kim Parker, Juliana Menasce Horowitz, and Anna Brown, "[Americans' Complex Views on Gender Identity and Transgender Issues](#)", Pew Research Center, June 28, 2022.

^{iv} "[Post in Polish Targeting Trans People](#)", Oversight Board, January 16, 2024; David Inserra, "[The Return of Blasphemy and Sacrilege Laws in the Most Unlikely Places](#)", Cato at Liberty, January 23, 2024.

^v "[Equality Maps: Bans on Transgender Youth Participation in Sports](#)", Movement Advancement Project, July 19, 2024.

^{vi} Jeffrey M. Jones, "[More Say Birth Gender Should Dictate Sports Participation](#)", Gallup, June 12, 2023; Chris Nesi, "[66% of US adults oppose transgender girls competing on female sports teams, new survey reveals](#)", New York Post, June 7, 2024; Kim Parker, Juliana Menasce Horowitz, and Anna Brown, "[Americans' Complex Views on Gender Identity and Transgender Issues](#)", Pew Research Center, June 28, 2022; Matthew Smith, "[Where does the British public stand on transgender rights in 2022?](#)", YouGov, July 20, 2022.

^{vii} Library of Congress, "[Tactics and Techniques of the National Woman's Party Suffrage Campaign](#)", Women of Protest: Photographs from the Records of the National Woman's Party. Accessed September 12, 2024.

^{viii} Jay Stanley, "[Civil Rights Movement Is a Reminder That Free Speech Is There to Protect the Weak](#)", ACLU, May 26, 2017.

^{ix} James Kirckick, "[The First Amendment Created Gay America](#)", The Free Press, May 31, 2022.

^x Jonathan Rauch, "Kindly Inquisitors: The New Attacks on Free Thought", Expanded edition. Chicago: University of Chicago Press, 2014. Page 125.

^{xi} Nadine Strossen, "HATE: Why We Should Resist it With Free Speech, Not Censorship", New York: Oxford University Press, 2018.

^{xii} Jonathan Rauch, "Kindly Inquisitors: The New Attacks on Free Thought", Expanded edition. Chicago: University of Chicago Press, 2014. Page 125-126.