

Case Decision 2020-007-FB-FBR

Case Summary

The Oversight Board has overturned Facebook's decision to remove a post under its Violence and Incitement Community Standard. While the company considered that the post contained a veiled threat, a majority of the Board believed it should be restored. This decision should only be implemented pending user notification and consent.

About the case

In late October 2020, a Facebook user posted in a public group described as a forum for Indian Muslims. The post contained a meme featuring an image from the Turkish television show "Diriliş: Ertuğrul" depicting one of the show's characters in leather armor holding a sheathed sword. The meme had a text overlay in Hindi. Facebook's translation of the text into English reads: "if the tongue of the kafir starts against the Prophet, then the sword should be taken out of the sheath." The post also included hashtags referring to President Emmanuel Macron of France as the devil and calling for the boycott of French products.

In its referral, Facebook noted that this content highlighted the tension between what it considered religious speech and a possible threat of violence, even if not made explicit.

Key findings

Facebook removed the post under its Violence and Incitement Community Standard, which states that users should not post coded statements where "the threat is veiled or implicit." Facebook identified "the sword should be taken out of the sheath" as a veiled threat against "kafirs," a term which the company interpreted as having a retaliatory tone against non-Muslims.

Considering the circumstances of the case, the majority of the Board did not believe that this post was likely to cause harm. They questioned Facebook's rationale, which indicated that threats of violence against Muslims increased Facebook's sensitivity to such threats, but also increased sensitivity when moderating content from this group.

While a minority viewed the post as threatening some form of violent response to blasphemy, the majority considered the references to President Macron and the boycott of French products as calls to action that are not necessarily violent. Although the television show character holds a sword, the majority interpreted the post as criticizing Macron's response to religiously motivated violence, rather than threatening violence itself.

The Board notes that its decision to restore this post does not imply endorsement of its content.

Under international human rights standards, people have the right to seek, receive and impart ideas and opinions of all kinds, including those that may be controversial or deeply offensive. As such, a majority considered that just as people have the right to criticize religions or religious figures, religious people also have the right to express offense at such expression.

Restrictions on expression must be easily understood and accessible. In this case, the Board noted that Facebook's process and criteria for determining veiled threats is not explained to users in the Community Standards.

In conclusion, a majority found that, for this specific post, Facebook did not accurately assess all contextual information and that international human rights standards on expression justify the Board's decision to restore the content.

The Oversight Board's decision

The Board overturns Facebook's decision to take down the content, requiring the post to be restored.

As a policy advisory statement, the Board recommends that:

- This decision should only be implemented pending user notification and consent.
- Facebook provide users with additional information regarding the scope and enforcement of restrictions on veiled threats. This would help users understand what content is allowed in this area. Facebook should make their enforcement criteria public. These should consider the intent and identity of the user, as well as their audience and the wider context.

*Case summaries provide an overview of the case and do not have precedential value.

Full Case Decision

1. Decision Summary

The Oversight Board has overturned Facebook’s decision to remove content it considered a veiled threat under its Violence and Incitement Community Standard. A majority of the Board found that restoring the content would comply with Facebook’s Community Standards, its values, and international human rights standards.

2. Case Description

In late October 2020, a Facebook user posted in a public group that describes itself as a forum for providing information for Indian Muslims. The post contained a meme featuring an image from the Turkish television show “Diriliş: Ertuğrul” depicting a character from the show in leather armor holding a sheathed sword. The meme had a text overlay in Hindi. Facebook’s translation of the text into English reads: “if the tongue of the kafir starts against the Prophet, then the sword should be taken out of the sheath.” The accompanying text in the post, also in English, stated that the Prophet is the user’s identity, dignity, honor and life, and contained the acronym “PBUH” (peace be upon him). This was followed by hashtags referring to President Emmanuel Macron of France as the devil and calling for the boycott of French products. The post was viewed about 30,000 times, received less than 1,000 comments and was shared fewer than 1,000 times.

In early November 2020, Facebook removed the post for violating its policy on Violence and Incitement. Facebook interpreted “kafir” as a pejorative term referring to nonbelievers in this context. Analyzing the photo and text, Facebook concluded that the post was a veiled threat of violence against “kafirs” and removed it.

Two Facebook users had previously reported the post; one for Hate Speech and the other for Violence and Incitement, and Facebook did not remove the content. Facebook then received information from a third-party partner that this content had the potential to contribute to violence. Facebook confirmed that this third-party partner is a member of its trusted partner network and is not linked to any state. Facebook described this network as a way for the company to obtain additional local context. According to Facebook, the network consists of non-governmental organizations, humanitarian organizations, non-profit organizations, and other international organizations. After the

post was flagged by the third-party partner, Facebook sought additional contextual information from its local public policy team, which agreed with the third-party partner that the post was potentially threatening. Facebook referred the case to the Oversight Board on November 19, 2020. In its referral, Facebook stated that it considered its decision to be challenging because the content highlighted tensions between what it considered religious speech and a possible threat of violence, even if not made explicit.

3. Authority and Scope

The Oversight Board has the authority to review Facebook’s decision under the Board’s Charter Article 2.1 and may uphold or overturn that decision under Article 3.5. This post is within the Oversight Board’s scope of review: it does not fit within any excluded category of content set forth in Article 2, Section 1.2.1 of the Board’s Bylaws and it does not conflict with Facebook’s legal obligations under Article 2, Section 1.2.2 of the Bylaws.

4. Relevant Standards

The Oversight Board considered the following standards in its decision:

I. Facebook’s Community Standards

The Community Standard on Violence and Incitement states that Facebook “aim[s] to prevent potential offline harm that may be related to content on Facebook” and that Facebook restricts expression “when [it] believe[s] there is a genuine risk of physical harm or direct threats to public safety.” Specifically, the standard indicates users should not post coded statements “where the method of violence or harm is not clearly articulated, but the threat is veiled or implicit.” Facebook also notes that it requires additional context to enforce this section of the standard.

II. Facebook’s Values

The Facebook values relevant to this case are outlined in the introduction to the Community Standards. The first is “Voice,” which is described as “paramount”:

The goal of our Community Standards has always been to create a place for expression and give people a voice. [...] We want people to be able to talk openly about the issues that matter to them, even if some may disagree or find them objectionable.

Facebook limits "Voice" in service of four other values: "Authenticity," "Safety," "Privacy" and "Dignity." The Board considers that the value of "Safety" is relevant to this decision:

Safety: We are committed to making Facebook a safe place. Expression that threatens people has the potential to intimidate, exclude or silence others and isn't allowed on Facebook.

III. Relevant Human Rights Standards Considered by the Board

The UN Guiding Principles on Business and Human Rights (UNGPs), endorsed by the UN Human Rights Council in 2011, establish a voluntary framework for the human rights responsibilities of private businesses. Drawing upon the UNGPs, the following international human rights standards were considered in this case:

- The right to freedom of expression: International Covenant on Civil and Political Rights (ICCPR), Article 19; General Comment No. 34, Human Rights Committee (2011) (GC34); Rabat Plan of Action.
- The right to life and security of person: ICCPR Articles 6 and 9, para. 1.

5. User Statement

Facebook notified the user it had referred the case to the Oversight Board, and gave the user the opportunity to share further context about the post with the Board. The user was given a 15-day window to submit their statement from the time of the referral. The Board received no statement from the user.

6. Explanation of Facebook's Decision

Facebook first evaluated the post for a possible Hate Speech violation and did not remove the content. Facebook did not indicate that the term "kafir" appears on a list of banned slurs or that the post otherwise violated the Hate Speech policy.

Facebook then removed this content based on its Violence and Incitement Community Standard. Under that standard, Facebook prohibits content that creates a "genuine risk of physical harm or direct threats to public safety," including coded statements "where the method of violence or harm is not clearly articulated, but the threat is veiled or implicit." Facebook explained that, in its view, veiled threats "can be as dangerous to

users as more explicit threats of violence.” According to Facebook, veiled threats are removed when certain non-public criteria are met.

Based on these criteria, Facebook determined “the sword should be taken out of the sheath” was a veiled threat against “kafirs” generally. In this case, Facebook interpreted the term “kafir” as pejorative with a retaliatory tone against non-Muslims; the reference to the sword as a threatening call to action; and also found it to be an “implied reference to historical violence.”

Facebook stated that it was crucial to consider the context in which the content was posted. According to Facebook, the content was posted at a time of religious tensions in India related to the Charlie Hebdo trials in France and elections in the Indian state of Bihar. Facebook noted rising violence against Muslims, such as the attack in Christchurch, New Zealand, against a mosque. It also noted the possibility of retaliatory violence by Muslims as leading to increased sensitivity in addressing potential threats both against and by Muslims.

Facebook further stated that its Violence and Incitement policy aligns with international human rights standards. According to Facebook, its policy is “narrowly framed to uphold the rights of others and to preserve the ‘necessity and proportionality’ elements required for permissible restriction of freedom of expression.”

7. Third party submissions

The Board received six public comments related to this case. The regional breakdown of the comments was: one from Asia Pacific and Oceania, one from Latin America and Caribbean and four from the United States and Canada. The submissions covered various themes, including: the importance of knowing the identity and influence of the user, including where it was posted and in what group; the importance of recognizing who the target is; whether the post targeted public figures or private individuals; whether the user intended to encourage the harmful stereotype of Indian Muslims as violent; whether the content met the standard of veiled threat under Facebook’s Community Standards; whether the Violence and Incitement policy was applicable in this case; whether the post could be deemed as violent speech under Facebook’s Hate Speech policy; as well as feedback for improving the Board’s public comment process.

To read public comments submitted for this case, please click [here](#).

8. Oversight Board Analysis

8.1 Compliance with Community Standards

A majority of the Board found that restoring this content would comply with Facebook's Community Standards.

Facebook indicated that the content was a veiled threat, prohibited by the Violence and Incitement Community Standard. The standard states users should not post coded statements "where the method of violence or harm is not clearly articulated, but the threat is veiled or implicit." Facebook stated in its rationale to the Board that it focuses on "imminent physical harm" in interpreting this provision of the standard.

Board Members unanimously considered it important to address veiled threats of violence, and expressed concern around users employing veiled threats to evade detection of Community Standards violation. Members also acknowledged the challenges Facebook faces in removing such threats at scale, given that they require contextual analysis.

Board Members differed in their views on how clearly the target was defined, the tone of the post, and the risk of physical harm or violence posed by this content globally and in India. A majority of the Board considered that the use of the hashtag to call for a boycott of French products was a call to non-violent protest and part of discourse on current political events. The use of a meme from a popular television show within this context, while referring to violence, was not considered by a majority as a call to physical harm.

In relation to Facebook's explanation, the Board noted that Facebook justified its decision by referring to ongoing tensions in India. However, the examples cited were not related to this context. For example, the protests that occurred in India in reaction to President Macron's statement following the killings in France in response to cartoon depictions of the Prophet Muhammad were not reported to be violent. Facebook also cited the November 7, 2020 elections in the Indian state of Bihar, yet the Board's research indicates that these elections were not marked by violence against persons based on their religion. The Board unanimously found that analysis of context is essential to understand veiled threats, yet a majority did not find Facebook's contextual rationale in relation to possible violence in India in this particular case compelling.

A minority found that Facebook’s internal process, which relied upon a third party partner assessment, was commendable, and would defer to Facebook’s determination that the post presented an unacceptable risk of promoting violence. This view acknowledged that Facebook consulted regional and linguistic experts, and shared the assessment that the term “kafir” was pejorative. The minority did not consider the Board had strong basis to overturn.

That said, a majority found that the Board’s independent analysis supported restoring the post under the Violence and Incitement Community Standard.

8.2 Compliance with Facebook Values

A majority of the Board found that restoring the content would comply with the company’s values. Although Facebook’s value of “Safety” is important, particularly given heightened religious tensions in India, this content did not pose a risk to “Safety” that justified displacing “Voice.” The Board also recognized the challenges Facebook faces in balancing these values when dealing with veiled threats. A minority considered these circumstances justified displacing “Voice” to err on the side of “Safety.”

8.3 Compliance with Human Rights Standards

A majority of the Board found that restoring this content would be consistent with international human rights standards.

According to Article 19 of the ICCPR individuals have the right to seek, receive and impart ideas and opinions of all kinds, including those that may be controversial or deeply offensive (General Comment No. 34, para. 11). The right to freedom of expression includes the dissemination of ideas that may be considered blasphemous, as well as opposition to such speech. In this regard, freedom of expression includes freedom to criticize religions, religious doctrines, and religious figures (General Comment No. 34, para. 48). Political expression is particularly important and receives heightened protection under international human rights law (General Comment No. 34, at para. 34 and 38) and includes calls for boycotts and criticism of public figures.

At the same time, the Board recognizes that the right to freedom of expression is not absolute and can exceptionally be subject to limitations under international human rights law. In this case, after discussing the factors in the Rabat Plan of Action, the Board did not consider the post to be advocacy of religious hatred reaching the threshold of

incitement to discrimination, hostility or violence, which states are required to prohibit under ICCPR Article 20, para. 2. ICCPR Article 19, para. 3 requires restrictions on expression to be easily understood and accessible (legality requirement), to have the purpose of advancing one of several listed objectives (legitimate aim requirement), and to be necessary and narrowly tailored to the specific objective (necessity and proportionality requirement). The Board discussed Facebook's removal decision against these criteria.

I. Legality

On legality, the Board noted that Facebook's process and criteria for determining veiled threats is not explained to users in the Community Standards, making it unclear what "additional context" is required to enforce the policy.

II. Legitimate aim

The Board further considered that the restriction on expression in this case would serve a legitimate aim: the protection of the rights of others (the rights to life and integrity of those targeted by the post).

III. Necessity and proportionality

A majority of the Board considered that the removal of the post was not necessary, emphasizing the importance of assessing the post in its particular context.

They considered that just as people have the right to criticize religion and religious figures, adherents of religions also have the right to express their offense at such expression. The Board recognized the serious nature of discrimination and violence against Muslims in India. The majority also considered the references to President Macron and the boycott of French products as non-violent calls to action. In this respect, although the post referenced a sword, the majority interpreted the post to criticize Macron's response to religiously motivated violence, rather than credibly threaten violence.

The Board considered a number of factors in determining that harm was improbable. The broad nature of the target ("kafirs") and the lack of clarity around potential physical harm or violence, which did not appear to be imminent, contributed to the majority's conclusion. The user not appearing to be a state actor or a public figure or otherwise

having particular influence over the conduct of others, was also significant. In addition there was no veiled reference to a particular time or location of any threatened or incited action. The Board's research indicated that protests in India following Macron's statements were not reportedly violent. In this respect, some Board Members noted the Facebook group was targeted towards individuals in India and partly in Hindi, which suggests the scope of impact may have been more limited to an area that did not see violent reactions. Additionally, some Board Members considered that the examples cited by Facebook largely related to violence against the Muslim minority in India, which Board Members considered to be a pressing concern, and not retaliatory violence by Muslims. Therefore, the majority concluded that as well as not being imminent, these factors meant physical harm was unlikely to result from this post.

A minority interpreted the post as threatening or legitimizing some form of violent response to blasphemy. Although the "sword" is a reference to nonspecific violence, the minority considered that the Charlie Hebdo killings and recent beheadings in France related to blasphemy mean this threat cannot be dismissed as unrealistic. The hashtags referencing events in France support this interpretation. In this case, the minority expressed that Facebook should not wait for violence to be imminent before removing content that threatens or intimidates those exercising their right to freedom of expression, and would have upheld Facebook's decision.

The majority, however, found that Facebook did not accurately assess all contextual information. The Board emphasized that restoring the content does not imply agreement with this content, and noted the complexities in assessing veiled or coded threats. Nonetheless, for this specific piece of content, international human rights standards on expression justify the Board's decision to restore the content.

9. Oversight Board Decision

9.1 Content Decision

The Oversight Board overturns Facebook's decision to take down the content, requiring the post to be restored.

9.2 Policy Advisory Statement

This decision should only be implemented pending user notification and consent.

To ensure users have clarity regarding permissible content, the Board recommends that Facebook provide users with additional information regarding the scope and enforcement of this Community Standard. Enforcement criteria should be public and align with Facebook's Internal Implementation Standards. Specifically, Facebook's criteria should address intent, the identity of the user and audience, and context.

***Procedural note:**

The Oversight Board's decisions are prepared by panels of five Members and must be agreed by a majority of the Board. Board decisions do not necessarily represent the personal views of all Members.

For this case decision, independent research was commissioned on behalf of the Board. An independent research institute headquartered at the University of Gothenburg and drawing on a team of over 50 social scientists on six continents, as well as more than 3,200 country experts from around the world, provided expertise on socio-political and cultural context. The company Lionbridge Technologies, LLC whose specialists are fluent in more than 350 languages and work from 5,000 cities across the world, provided linguistic expertise.

- Date Published: Feb 12, 2021
- Platform: Facebook
- Relevant Community Standard: Violence and Incitement
- Related Topics: Religion, Violence
- Region: Central and South Asia
- Country(s) Affected: France, India