

## Case decision 2021-003-FB-UA

### Case summary

The Oversight Board has overturned Facebook's decision to remove a post under its Dangerous Individuals and Organizations Community Standard. After the Board identified this case for review, Facebook restored the content. The Board expressed concerns that Facebook did not review the user's appeal against its original decision. The Board also urged the company to take action to avoid mistakes which silence the voices of religious minorities.

### **About the case**

In November 2020, a user shared a video post from Punjabi-language online media company Global Punjab TV. This featured a 17-minute interview with Professor Manjit Singh who is described as "a social activist and supporter of the Punjabi culture." The post also included a caption mentioning Hindu nationalist organization Rashtriya Swayamsevak Sangh (RSS) and India's ruling party Bharatiya Janata Party (BJP): "RSS is the new threat. Ram Naam Satya Hai. The BJP moved towards extremism."

In text accompanying the post, the user claimed the RSS was threatening to kill Sikhs, a minority religious group in India, and to repeat the "deadly saga" of 1984 when Hindu mobs massacred and burned Sikh men, women and children. The user alleged that Prime Minister Modi himself is formulating the threat of "Genocide of the Sikhs" on advice of the RSS President, Mohan Bhagwat. The user also claimed that Sikh regiments in the army have warned Prime Minister Modi of their willingness to die to protect the Sikh farmers and their land in Punjab.

After being reported by one user, a human reviewer determined that the post violated Facebook's Dangerous Individuals and Organizations Community Standard and removed it. This triggered an automatic restriction on the user's account. Facebook told the user that they could not review their appeal of the removal because of a temporary reduction in review capacity due to COVID-19.

## **Key findings**

After the Board identified this case for review, but prior to it being assigned to a panel, Facebook realized that the content was removed in error and restored it. Facebook noted that none of the groups or individuals mentioned in the content are designated as "dangerous" under its rules. The company also could not identify the specific words in the post which led to it being removed in error.

The Board found that Facebook's original decision to remove the post was not consistent with the company's Community Standards or its human rights responsibilities.

The Board noted that the post highlighted the concerns of minority and opposition voices in India that are allegedly being discriminated against by the government. It is particularly important that Facebook takes steps to avoid mistakes which silence such voices. While recognizing the unique circumstances of COVID-19, the Board argued that Facebook did not give adequate time or attention to reviewing this content. It stressed that users should be able to appeal cases to Facebook before they come to the Board and urged the company to prioritize restoring this capacity.

Considering the above, the Board found the account restrictions that excluded the user from Facebook particularly disproportionate. It also expressed concerns that Facebook's rules on such restrictions are spread across many locations and not all found in the Community Standards, as one would expect.

Finally, the Board noted that Facebook's transparency reporting makes it difficult to assess whether enforcement of the Dangerous Individuals and Organizations policy has a particular impact on minority language speakers or religious minorities in India.

### **The Oversight Board's decision**

The Board overturns Facebook's original decision to remove the content. In a policy advisory statement, the Board recommends that Facebook:

- Translate its Community Standards and Internal Implementation Standards into Punjabi. Facebook should also aim to make its Community Standards accessible in all languages widely spoken by its users.
- Restore both human review of content moderation decisions and access to a human appeals process to pre-pandemic levels as soon as possible, while protecting the health of Facebook's staff and contractors.
- Increase public information on error rates by making this viewable by country and language for each Community Standard in its transparency reporting.

\*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 the content. The Board notes that, after it selected the case but before it was assigned to a panel, Facebook determined that the content was removed in error and restored it. The Board found that the content in question did not praise, support or represent any dangerous individual or organization. The post highlighted the alleged mistreatment of minorities in India by government and pro-government actors and had public interest value. The Board was concerned about mistakes in the review of the content and the lack of an effective appeals process available to the user. Facebook's mistakes undermined the

user's freedom of expression as well as the rights of members of minorities in India to access information.

## **2. Case description**

The content touched on allegations of discrimination against minorities and silencing of the opposition in India by "Rashtriya Swayamsevak Sangh" (RSS) and the Bharatiya Janata Party (BJP). RSS is a Hindu nationalist organization that has allegedly been involved in violence against religious minorities in India. "BJP" is India's ruling party to which the current Indian Prime Minister Narendra Modi belongs, and has close ties with RSS.

In November 2020, a user shared a video post from Punjabi-language online media Global Punjab TV and an accompanying text. The post featured a 17-minute interview with Professor Manjit Singh, described as "a social activist and supporter of the Punjabi culture." In its post, Global Punjab TV included the caption "RSS is the new threat. Ram Naam Satya Hai. The BJP moved towards extremism." The media company also included an additional description "New Threat. Ram Naam Satya Hai! The BJP has moved towards extremism. Scholars directly challenge Modi!" The content was posted during India's mass farmer protests and briefly touched on the reasons behind the protests and praised them.

The user added accompanying text when sharing Global Punjab TV's post in which they stated that the CIA designated the RSS a "fanatic Hindu terrorist organization" and that Indian Prime Minister Narendra Modi was once its president. The user wrote that the RSS was threatening to kill Sikhs, a minority religious group in India, and to repeat the "deadly saga" of 1984 when Hindu mobs attacked Sikhs. They stated that "The RSS used the Death Phrase 'Ram naam sat hai'." The Board understands the phrase "Ram Naam Satya Hai" to be a funeral chant that has allegedly been used as a threat by some Hindu nationalists. The user alleged that Prime Minister Modi himself is

formulating the threat of “Genocide of the Sikhs” on advice of the RSS President, Mohan Bhagwat. The accompanying text ends with a claim that Sikhs in India should be on high alert and that Sikh regiments in the army have warned Prime Minister Modi of their willingness to die to protect the Sikh farmers and their land in Punjab.

The post was up for 14 days and viewed fewer than 500 times before it was reported by another user for “terrorism.” A human reviewer determined that the post violated the Community Standard on [Dangerous Individuals and Organizations](#) and took down the content, which also triggered an automatic restriction on the use of the account for a fixed period of time. In its notification to the user, Facebook noted that its decision was final and could not be reviewed due to a temporary reduction in its review capacity due to COVID-19. For this reason, the user appealed to the Oversight Board.

After the Case Selection Committee identified this case for review, but prior to it being assigned to a panel, Facebook determined the content was removed in error and restored it. The Board nevertheless proceeded in assigning the case to panel.

### **3. Authority and scope**

The Board has authority to review Facebook's decision under Article 2 (Authority to Review) of the Board's Charter and may uphold or reverse that decision under Article 3, Section 5 (Procedures for review: Resolution of the Charter). Facebook has not presented reasons for the content to be excluded in accordance with Article 2, Section 1.2.1 (Content not Available for Board Review) of the Board's Bylaws, nor has Facebook indicated that it considers the case to be ineligible under Article 2, Section 1.2.2 (Legal obligations) of the Bylaws. Under Article 3, Section 4 (Procedures for Review: Decisions) of the Board's Charter, the final decision may include a policy advisory statement, which will be taken into consideration by Facebook to guide its future policy development.

Facebook restored the user's content after determining their error, which likely would not have happened if the Board had not identified the case. In line with [case decision 2020-004-IG-UA](#), Facebook's choice to restore content does not exclude the case from review. Concerns over why the error occurred, the harm stemming from it, and the need to ensure it is not repeated remain pertinent. The Board offers users a chance to be heard and receive a full explanation of what happened.

#### **4. Relevant standards**

The Oversight Board considered the following standards in its decision:

##### **I. Facebook's Community Standards:**

Facebook's [Dangerous Individuals and Organizations](#) policy explains that "in an effort to prevent and disrupt real-world harm, we do not allow any organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on Facebook." The Standard further states that Facebook removes "content that expresses support or praise for groups, leaders, or individuals involved in these activities."

##### **II. Facebook's values:**

Facebook's values are outlined in the introduction to the Community Standards.

The value of "Voice" 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 may limit “Voice” in service of the values of four values, including “Safety” and “Dignity”:

*“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.*

*“Dignity” : We believe that all people are equal in dignity and rights. We expect that people will respect the dignity of others and not harass or degrade others.*

### III. Human rights standards:

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. Facebook’s commitment to respect human rights standards in line with the UNGPs was elaborated in [a new corporate policy](#) launched in March 2021. The Board's analysis in this case was informed by the following human rights standards:

Freedom of expression: Article 19, International Covenant on Civil and Political Rights (ICCPR); Human Rights Committee, [General Comment No. 34](#) (2011); Special Rapporteur on freedom of opinion and expression, reports: [A/HRC/38/35](#) (2018) and [A/74/486](#) (2019)

The right to non-discrimination: Article 2, para. 1 and Article 26, ICCPR; Human Rights Committee, [General Comment No. 23](#) (1994); General Assembly, Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities, as interpreted in by the Independent Expert on Minority Issues in [A/HRC/22/49](#) para. 57-58 (2012); Special Rapporteur on Minority Issues, [A/HRC/46/57](#) (2021)

The right to an effective remedy: Article 2, para. 3, ICCPR; Human Rights Committee, General Comment 31 (2004); Human Rights Committee, General Comment No. 29 (2001);

The right to security of person: Article 9, para. 1, ICCPR, as interpreted in General Comment No. 35, para. 9.

## **5. User statement**

The user indicated to the Board that the post was not threatening or criminal but simply repeated the video's substance and reflected its tone. The user complained about account restrictions imposed on them. They suggested that Facebook should simply delete problematic videos and avoid restricting users' accounts, unless they engage in threatening or criminal behavior. The user also claimed that thousands of people engage with their content and called on the account to be restored immediately.

## **6. Explanation of Facebook's decision**

According to Facebook, following a single report against the post, the person who reviewed the content wrongly found a violation of the of the Dangerous Individuals and Organizations Community Standard. Facebook informed the Board that the user's post included no reference to individuals or organizations designated as dangerous. It followed that the post contained no violating praise.

Facebook explained that the error was due to the length of the video (17 minutes), the number of speakers (two), the complexity of the content, and its claims about various political groups. The company added that content reviewers look at thousands of pieces of content every day and mistakes happen during that process. Due to the volume of content, Facebook stated that content reviewers are not always able to watch videos in



full. Facebook was unable to specify the part of the content the reviewer found to violate the company's rules.

While the user appealed the decision to Facebook, they were informed that Facebook could not review the post again due to staff shortages caused by COVID-19.

## **7. Third-party submissions**

The Oversight Board received six public comments related to this case. Two comments were submitted from Europe and four from the United States and Canada. The submissions covered the following themes: the scope of political expression, Facebook's legal right to moderate content, and the political context in India.

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

## **8. Oversight Board analysis**

### **8.1 Compliance with Community Standards**

The Board concluded that Facebook's original decision to remove the content was inconsistent with its Dangerous Individuals and Organizations Community Standard.

The content referred to the BJP as well as the RSS and several of its leaders. Facebook explained that none of these groups or individuals are designated as "dangerous" under its Community Standards, and it was unable to identify the specific words in the content that led to its removal. The Board noted that, even if these organizations had been designated as dangerous, the content clearly criticized them. The content did praise one group – Indian farmers who were protesting. It therefore appears that inadequate time or attention was given to reviewing this content.

The Board finds that the Dangerous Individuals and Organizations Community Standard is clear that violating content will be removed. The [Introduction to the Community Standards](#), as well as [Facebook's Help Centre](#) and [Newsroom](#), explain that severe or persistent violations may result in a loss of access to some features. In this case, Facebook explained to the Board that it imposed an automatic restriction on the user's account for a fixed period of time for repeat violations. The Board found this would have been consistent with the company's Community Standards had there been a violation.

Facebook explained to the Board that account restrictions are automatic. These are imposed once a violation of the Community Standards has been determined and depend on the individual's history of violations. This means that a person reviewing the content is not aware of whether removal will lead to an account restriction and is not involved in selecting that restriction. The Board notes that the consequences of enforcement mistakes can be severe and expresses concern that account level restrictions were wrongly applied in this case.

## **8.2 Compliance with Facebook's values**

The Board found that Facebook's decision to remove the content was inconsistent with its values of "Voice," "Dignity" and "Safety." The content linked to a media report and related to important political issues, including commentary on the alleged violation of minority rights and the silencing of opposition by senior BJP politicians and the RSS. Therefore, the incorrect removal of the post undermined the values of "Voice" and "Dignity."

Facebook has indicated it prioritizes the value of "Safety" when enforcing the Community Standard on Dangerous Individuals and Organizations. However, in this case, the content did not refer to, or praise, any designated dangerous individual or organization. Instead, the Board found that the content criticized governmental actors and political groups.

### **8.3 Compliance with Facebook’s human rights responsibilities**

Facebook’s application of the Community Standard on Dangerous Individuals and Organizations was inconsistent with the company’s human rights responsibilities and its publicly stated commitments to the UNGPs. Principles 11 and 13 call on businesses to avoid causing or contributing to adverse human rights impacts that may arise from their own activities or their relationships with other parties, including state actors, and to mitigate them.

#### *I. Freedom of Expression and Information (Article 19, ICCPR)*

Article 19 of the ICCPR guarantees the right to freedom of expression, and places particular value on uninhibited public debate, especially concerning political figures and the discussion on human rights (General Comment 34, paras 11 and 34).

Article 19 also guarantees the right to seek and receive information, including from the media (General Comment 34, para. 13). This is guaranteed without discrimination, and human rights law places particular emphasis on the importance of independent and diverse media, especially for ethnic and linguistic minorities (General Comment 34, para. 14).

#### *a. Legality*

The Board has previously raised concerns with the accessibility of the Community Standard on Dangerous Individuals and Organizations, including around Facebook’s interpretation of “praise,” and the process for designating dangerous individuals and organizations ( case decision 2020-005-FB-UA). Precise rules are important to constrain discretion and prevent arbitrary decision-making (General Comment No. 34, para. 25), and also to safeguard against bias. They also help Facebook users understand the rules being enforced against them. The UN Special Rapporteur on

freedom of expression has raised concern at social media companies adopting vague rules that broadly prohibit “praise” and “support” leaders of dangerous organizations (report A/HRC/38/35, para. 26).

The consequences of violating a rule, e.g. suspension of account functionalities or account disabling, must also be clear. The Board is concerned that information on account restrictions is spread across many locations, and not all set out in the Community Standards as one would expect. It is important to give users adequate notice and information when they violate rules so they can adjust their behavior accordingly. The Board notes its previous recommendations that Facebook should not expect users to synthesize rules from across multiple sources, and for rules to be consolidated in the Community Standards ( [case decision 2020-006- FB-FBR](#), Section 9.2).

The Board is concerned that the Community Standards are not translated into Punjabi, a language widely spoken globally with 30 million speakers in India. Facebook’s Internal Implementation Standards are also not available in Punjabi for moderators working in this language. This will likely compound the problem of users not understanding the rules, and increase the likelihood of moderators making enforcement errors. The possible specific impacts on a minority population raise human rights concerns (A/HRC/22/49, para. 57).

#### *b. Legitimate aim*

Article 19, para. 3 of the ICCPR states that legitimate aims include respect for the rights or reputations of others, as well as the protection of national security, public order, or public health or morals. Facebook has indicated that the aim of the Dangerous Individuals and Organizations Community Standard is to protect the rights of others. The Board is satisfied that the policy pursues a legitimate aim, in particular to protect

the right to life, security of person, and equality and non-discrimination (General Comment 34, para. 28; Oversight Board decision 2020-005-FB-UA).

*c. Necessity and proportionality*

Restrictions must be necessary and proportionate to achieve a legitimate aim. There must be a direct connection between the necessity and proportionality of the specific action taken and the threat stemming from the expression (General Comment 34, para. 35). Facebook has acknowledged that its decision to remove the content was a mistake, and does not argue that this action was necessary or proportionate.

Mistakes which restrict expression on political issues are a serious concern. It is particularly worrying if such mistakes are widespread, and especially if this impacts minority language speakers or religious minorities who may already be politically marginalized. The UN Special Rapporteur on minority issues has expressed concern at hate speech targeting minority groups on Facebook in India (A/HRC/46/57, para. 40). In such regional contexts, errors can silence minority voices that seek to counter hateful and discriminatory narratives, as in this case.

The political context in India when this post was made, with mass anti-government farmer protests and increasing governmental pressure on social media platforms to remove related content, underscores the importance of getting decisions right. In this case, the content related to the protests and the silencing of opposition voices. It also included a link to an interview from a minority language media outlet on the topics. Dominant platforms should avoid undermining the expression of minorities who are protesting their government and uphold media pluralism and diversity (General Comment 34, para. 40). The account restrictions which wrongfully excluded the user from the platform during this critical period were particularly disproportionate.

Facebook explained that they could not carry out an appeal on the user's content due to reduced capacity during the COVID-19 pandemic. While the Board appreciates these unique circumstances, it again stresses the importance of Facebook providing transparency and accessible processes for appealing their decisions (UNGPs, Principle 11; A/74/486, para. 53). As the Board stated in case decision [2020-004-IG-UA](#), cases should be appealed to Facebook before they come to the Board. To ensure users' access to remedy, Facebook should prioritize the return of this capacity as soon as possible.

The Board acknowledges that mistakes are inevitable when moderating content at scale. Nevertheless, Facebook's responsibility to prevent, mitigate and address adverse human rights impacts requires learning from these mistakes (UNGPs, Principles 11 and 13).

It is not possible to tell from one case whether this enforcement was symptomatic of intentional or unintentional bias on behalf of the reviewer. Facebook also declined to provide specific answers to the Board's questions regarding possible communications from Indian authorities to restrict content around the farmer's protests, content critical of the government over its treatment of farmers, or content concerning the protests. Facebook determined that the requested information was not reasonably required for decision-making in accordance with the intent of the Charter and/or cannot or should not be provided because of legal, privacy, safety, or data protection restrictions or concerns. Facebook cited the Oversight Board's Bylaws, Article 2, Section 2.2.2, to justify its refusal.

Facebook answered the Board's question on how Facebook's moderation in India is independent of government influence. The company explained that its staff receive training specific to their region, market, or role as part of the Global Ethics and Compliance initiative, which fosters a culture of honesty, transparency, integrity,

accountability and ethical values. Further, Facebook's staff are bound by a Code of Conduct and an Anti-Corruption Policy.

The Board emphasizes the importance of processes for reviewing content moderation decisions, including auditing, to check for and correct any bias in manual and automated decision-making, especially in relation to places experiencing periods of crisis and unrest. These assessments should take into account the potential for coordinated campaigns by governments and non-state actors to maliciously report dissent.

Transparency is essential to ensure public scrutiny of Facebook's actions in this area. The lack of detail in Facebook's transparency reporting makes it difficult for the Board or other actors to assess, for example, if enforcement of the Dangerous Individuals and Organizations policy has particular impacts on users, and particularly minority language speakers, in India. To inform the debate, Facebook should make more data public, and provide analysis of what it means.

## **9. Oversight Board Decision**

The Oversight Board overturns Facebook's decision to take down the content and requires the post to be restored. The Board notes that Facebook has already taken action to this effect.

## **10. Policy advisory statement**

The following recommendations are numbered, and the Board requests that Facebook provides an individual response to each as drafted.

*Accessibility*

1. Facebook should translate its Community Standards and Internal Implementation Standards into Punjabi. Facebook should aim to make its Community Standards accessible in all languages widely spoken by its users. This would allow a full understanding of the rules that users must abide by when using Facebook's products. It would also make it simpler for users to engage with Facebook over content that may violate their rights.

#### *Right to remedy*

2. In line with the Board's recommendation in case [2020-004-IG-UA](#), the company should restore human review and access to a human appeals process to pre-pandemic levels as soon as possible while fully protecting the health of Facebook's staff and contractors.

#### *Transparency reporting*

3. Facebook should improve its transparency reporting to increase public information on error rates by making this information viewable by country and language for each Community Standard. The Board underscores that more detailed transparency reports will help the public spot areas where errors are more common, including potential specific impacts on minority groups, and alert Facebook to correct them.

#### **\*Procedural note:**

The Oversight Board's decisions are prepared by panels of five Members and approved 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: Apr 29, 2021
- Platform: Facebook
- Relevant Community Standard: Dangerous Individuals and Organizations
- Related Topics: Politics
- Region: Central and South Asia
- Country(s) Affected: India

Policies Cited: Dangerous Individuals and Organizations