


OVERTURNED

2021-006-IG-UA


Öcalan's isolation

The Oversight Board has overturned Facebook's original decision to remove an Instagram post encouraging people to discuss the solitary confinement of Abdullah Öcalan, a founding member of the Kurdistan Workers' Party (PKK).

Policies and topics

-  Freedom of expression, Marginalized communities, Misinformation
-  Dangerous individuals and organizations

Region and countries

-  United States & Canada
-  United States, Turkey

Platform

-  Instagram

Attachments

[2021-006-IG-UA Public Comments](#)

To read the full decision in Northern Kurdish click [here](#).

Ji bo hûn ev biryar bi Kurdiya Bakur bixwînin, [li vir](#) bitikînin.

Case summary

The Oversight Board has overturned Facebook’s original decision to remove an Instagram post encouraging people to discuss the solitary confinement of Abdullah Öcalan, a founding member of the Kurdistan Workers’ Party (PKK). After the user appealed and the Board selected the case for review, Facebook concluded that the content was removed in error and restored it. The Board is concerned that Facebook misplaced an internal policy exception for three years and that this may have led to many other posts being wrongly removed.

About the case

This case relates to Abdullah Öcalan, a founding member of the PKK. This group has used violence in seeking to achieve its aim of establishing an independent Kurdish state. Both the PKK and Öcalan are designated as dangerous entities under Facebook’s Dangerous Individuals and Organizations policy.

On January 25, 2021, an Instagram user in the United States posted a picture of Öcalan which included the words “y’all ready for this conversation” in English. In a caption, the user wrote that it was time to talk about ending Öcalan’s isolation in prison on Imrali island in Turkey. The user encouraged readers to engage in conversation about Öcalan’s imprisonment and the inhumane nature of solitary confinement.

After being assessed by a moderator, the post was removed on February 12 under Facebook’s rules on Dangerous Individuals and Organizations as a call to action to support Öcalan and the PKK. When the user appealed this decision, they were told their appeal could not be reviewed because of a temporary reduction in Facebook’s review capacity due to COVID-19. However, a second moderator did carry out a review of the content and found that it violated the same policy. The user then appealed to the Oversight Board.

After the Board selected this case and assigned it to panel, Facebook found that a piece of

internal guidance on the Dangerous Individuals and Organizations policy was “inadvertently not transferred” to a new review system in 2018. This guidance, developed in 2017 partly in

response to concern about the conditions of Öcalan’s imprisonment, allows discussion on the conditions of confinement for individuals designated as dangerous.

In line with this guidance, Facebook restored the content to Instagram on April 23. Facebook told the Board that it is currently working on an update to its policies to allow users to discuss the human rights of designated dangerous individuals. The company asked the Board to provide insight and guidance on how to improve these policies. While Facebook updated its Community Standard on Dangerous Individuals and Organizations on June 23, 2021, these changes do not directly impact the guidance the company requested from the Board.

Key findings

The Board found that Facebook’s original decision to remove the content was not in line with the company’s Community Standards. As the misplaced internal guidance specifies that users can discuss the conditions of confinement of an individual who has been designated as dangerous, the post was permitted under Facebook’s rules.

The Board is concerned that Facebook lost specific guidance on an important policy exception for three years. Facebook’s policy of defaulting towards removing content showing “support” for designated individuals, while keeping key exceptions hidden from the public, allowed this mistake to go unnoticed for an extended period. Facebook only learned that this policy was not being applied because of the user who decided to appeal the company’s decision to the Board.

While Facebook told the Board that it is conducting a review of how it failed to transfer this guidance to its new review system, it also stated “it is not technically feasible to determine how many pieces of content were removed when this policy guidance was not available to reviewers.” The Board believes that Facebook’s mistake may have led to many other posts being wrongly removed and that Facebook’s transparency reporting is not sufficient to assess whether this type of error reflects a systemic problem. Facebook’s actions in this case indicate that the company is failing to respect the right to remedy, contravening its Corporate Human Rights Policy (Section 3).

Even without the discovery of the misplaced guidance, the content should never have been removed. The user did not advocate violence in their post and did not express support for Öcalan's ideology or the PKK. Instead, they sought to highlight human rights concerns about Öcalan's prolonged solitary confinement which have also been raised by international bodies. As the post was unlikely to result in harm, its removal was not necessary or proportionate under international human rights standards.

The Oversight Board's decision

The Oversight Board overturns Facebook's original decision to remove the content. The Board notes that Facebook has already restored the content.

In a policy advisory statement, the Board recommends that Facebook:

- Immediately restore the misplaced 2017 guidance to the Internal Implementation Standards and Known Questions (the internal guidance for content moderators).
- Evaluate automated moderation processes for enforcement of the Dangerous Individuals and Organizations policy. Where necessary, Facebook should update classifiers to exclude training data from prior enforcement errors that resulted from failures to apply the 2017 guidance.
- Publish the results of the ongoing review process to determine if any other policies were lost, including descriptions of all lost policies, the period they were lost for, and steps taken to restore them.
- Ensure the Dangerous Individuals and Organizations "policy rationale" reflects that respect for human rights and freedom of expression can advance the value of "Safety." The policy rationale should specify in greater detail the "real-world harms" the policy seeks to prevent and disrupt when "Voice" is suppressed.
- Add to the policy a clear explanation of what "support" excludes. Users should be free to discuss alleged abuses of the human rights of members of designated organizations.
- Explain in the Community Standards how users can make the intent behind their posts clear to Facebook.
- Ensure meaningful stakeholder engagement on the proposed changes to its Dangerous Individuals and Organizations policy through Facebook's Product Policy Forum, including through a public call for inputs.
- Ensure internal guidance and training is provided to content moderators on any

proposed policy changes.

- Ensure that users are notified when their content is removed. The notification should note whether the removal is due to a government request or due to a violation of the Community Standards, or due to a government claiming a national law has been violated (and the jurisdictional reach of any removal).
- Clarify to Instagram users that Facebook’s Community Standards apply to Instagram in the same way they apply to Facebook.
- Include information in its transparency reporting on the number of requests received for content removals from governments based on Community Standards violations (as opposed to violations of national law), and the outcomes of those requests.
- Include more comprehensive information in its transparency reporting on error rates for enforcing rules on “praise” and “support” of dangerous individuals and organizations, broken down by region and language.

Full case decision

1. Decision summary

The Oversight Board has overturned Facebook’s original decision to remove an Instagram post encouraging people to discuss the solitary confinement of Abdullah Öcalan, a person designated by Facebook as a dangerous individual. After the user appealed and the Board selected the case for review, Facebook concluded that the content was removed in error and restored the post to Instagram.

Facebook explained that in 2018 it “inadvertently failed to transfer” a piece of internal policy guidance that allowed users to discuss conditions of confinement of designated dangerous individuals to a new review system. The Board believes that if Facebook were more transparent about its policies the harm from this mistake could have been mitigated or avoided altogether. Even without the misplaced internal policy guidance, the Board found that the content never should have been removed. It was simply a call to debate the necessity of Öcalan’s detention in solitary confinement and its removal did not serve the aim of the Dangerous Individuals and Organizations policy “to prevent and disrupt real-world harm.” Instead, the removal resulted in a restriction on freedom of expression about a human rights concern.

2. Case description

The case concerns content related to Abdullah Öcalan, a founding member of the Kurdistan Workers' Party (PKK). The PKK was founded in the 1970s with the aim of establishing an independent Kurdish state in South-Eastern Turkey, Syria, and Iraq. The group uses violence in seeking to achieve its aim. The PKK has been designated as a terrorist organization by the United States, the EU, the UK, and Turkey, among others. Öcalan has been imprisoned on Imrali Island, Turkey, since his arrest and sentencing in 1999 for carrying out violent acts aimed at the secession of a part of Turkey's territory ([Case of Ocalan v Turkey, European Court of Human Rights](#)).

On January 25, 2021, an Instagram user in the United States posted a picture of Öcalan, which included the words "y'all ready for this conversation" in English. Below the picture, the user wrote that it was time to talk about ending Öcalan's isolation in prison on Imrali Island. The user encouraged readers to engage in conversation about Öcalan's imprisonment and the inhumane nature of solitary confinement, including through hunger strikes, protests, legal action, op-eds, reading groups, and memes. The content did not call for Öcalan's release, nor did it mention the PKK or endorse violence.

The post was automatically flagged by Facebook and, after being assessed by a moderator, was removed on February 12 for breaching the policy on Dangerous Individuals and Organizations. The user appealed the decision to Facebook and was informed that the decision was final and could not be reviewed because of a temporary reduction in Facebook's review capacity due to COVID-19. However, a second moderator still carried out a review of the content, also finding a breach of the Dangerous Individuals and Organizations policy. The user received a notification explaining that the initial decision was upheld by a second review. The user then appealed to the Oversight Board.

The Board selected the case for review and assigned it to a panel. As Facebook prepared its decision rationale for the Board, it found a piece of internal guidance on the Dangerous Individuals and Organizations policy that allows discussion or debate about the conditions of confinement for individuals designated as dangerous. This guidance was developed in 2017 partly in response to international concern about the conditions of Öcalan's imprisonment. Facebook explained that in 2018 the guidance was "inadvertently not transferred" to a new review system. It was also not shared within Facebook's policy team which sets the rules for

what is allowed on the platform. While the guidance remained technically accessible to content moderators in a training annex, the company acknowledges that it was difficult to find during standard review procedures and that the reviewer in this case did not have access to it. This guidance is a strictly internal document designed to assist Facebook's moderators and was not reflected in Facebook's public-facing Community Standards or Instagram's Community Guidelines.

Facebook only learned that this policy was not being applied because of the user who decided to appeal Facebook's decision to remove their content to the Board. If not for this user's actions, it is possible this error would never have come to light. As of June 29, Facebook has yet to reinstate the misplaced internal policy into its guidance for content moderators. The company explained to the Board that it "will work to ensure that the guidance it provides to its content reviewers on this subject is clear and more readily accessible to help avoid future enforcement errors."

Facebook restored the content to Instagram on April 23 and notified the Board that it "is currently working on an update to its policies to make clear that users can debate or discuss the conditions of confinement of designated terrorist individuals or other violations of their human rights, while still prohibiting content that praises or supports those individuals' violent actions." The company welcomed "the Oversight Board's insight and guidance into how to strike an appropriate balance between fostering expression on subjects of human rights concern while simultaneously ensuring that its platform is not used to spread content praising or supporting terrorists or violent actors."

3. Authority and scope

The Board has the power to review Facebook's decision following an appeal from the user whose post was removed (Charter Article 2, Section 1; Bylaws Article 2, Section 2.1). The Board may uphold or reverse that decision (Charter Article 3, Section 5). In line with case decision 2020-004-IG-UA, Facebook reversing a decision a user appealed against does not exclude the case from review.

The Board's decisions are binding and may include policy advisory statements with recommendations. These recommendations are non-binding, but Facebook must respond to them (Charter Article 3, Section 4).

The Board is an independent grievance mechanism to address disputes in a transparent and principled manner.

4. Relevant standards

The Oversight Board considered the following standards in its decision:

I. Facebook's content policies:

Instagram's Community Guidelines state that Instagram is not a place to support or praise terrorism, organized crime, or hate groups. This section of the Guidelines includes a link to Facebook's Community Standard on Dangerous Individuals and Organizations (a change log reflecting the June 23 update to the Community Standards is here). In response to a question from the Board, Facebook has confirmed that the Community Standards apply to Instagram in the same way they apply to Facebook.

The Dangerous Individuals and Organizations Community Standard states 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 stated, at the time it was enforced, 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 limits "Voice" in the service of four values, and one is relevant here:

"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. 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. In March 2021, [Facebook announced its Corporate Human Rights Policy](#), where it recommitted to respecting human rights in accordance with the UNGPs. 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](#)); Article 6, Declaration on Human Rights Defenders, [A/RES/53/144](#) (1998); Human Rights Committee, [General Comment No. 34](#) (2011); Special Rapporteur on freedom of opinion and expression, [A/HRC/38/35](#) (2018), [A/74/486](#) (2019); Special Rapporteur on human rights and counter-terrorism, [A/HRC/40/52](#) (2019)
- [Right to remedy](#): Article 2, ICCPR; Human Rights Committee, [General Comment No. 31](#) (2004)
- [Prohibition on torture, cruel, inhuman, or degrading treatment or punishment](#): Rule 43, UN Standard Minimum rules for the Treatment of Prisoners (the Nelson Mandela Rules) [A/Res/70/175](#) (2016).

5. User statement

In their appeal to the Board, the user explained that they posted the content to spur discussion about Öcalan's philosophy and to end his isolation. The user said that they believed banning communicating about Öcalan and his philosophy prevents discussions that could lead to a peaceful settlement for Kurdish people in the Middle East. They also stated that they did not wish to promote violence but believed there should not be a ban on posting pictures of Öcalan on Instagram. The user claimed that the association of Öcalan's face with violent organizations is not based on fact, but rather is slander and an ongoing effort to silence an important conversation. They compared Öcalan's imprisonment to that of former South African President Nelson Mandela, noting that the international community has a role in illuminating Öcalan's imprisonment just as it did with Mandela.

6. Explanation of Facebook's decision

Facebook initially concluded that the content was a call to action to support Öcalan and the PKK, which violated the Dangerous Individuals and Organizations policy. Öcalan co-founded the PKK, which Facebook notes has been designated as a Foreign Terrorist Organization by the United States. Based on this designation of the organization, Facebook added Öcalan to its list of designated dangerous individuals. Under its Dangerous Individuals and Organizations policy, Facebook removes all content that it deems to support such individuals.

After the Board selected this case for review, Facebook evaluated the content against its policies again and found that it developed internal guidance in this area in 2017. In explaining the situation, Facebook stated that it inadvertently failed to transfer this guidance when it switched to a new review system in 2018 and did not share it throughout its policy team.

This guidance allows content where the poster is calling for the freedom of a terrorist when the context of the content is shared in a way that advocates for peace or debate of the terrorist's incarceration. Applying that guidance, Facebook found that the content in this case fell squarely within it and restored the content.

7. Third-party submissions

The Board received 12 public comments related to this case. Six came from the United States and Canada, four from Europe, and two from the Middle East and North Africa.

The submissions covered themes including the lack of transparency around the Dangerous Individuals and Organizations policy as well as its inconsistency with international human rights law, and that calls for discussion of solitary confinement do not constitute praise or support.

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

8. Oversight Board analysis

8.1 Compliance with Facebook's content policies

The Board found that Facebook's decision to remove the content was not in line with the company's Community Standards. The Community Standard on Dangerous Individuals and Organizations did not define what constituted "support" for a designated dangerous

individual or organization until it was updated on June 23, 2021.

In January 2021, the Board recommended that Facebook publicly define praise, support, and representation, as well as provide more illustrative examples of how the policy is applied ([case 2020-005-FB-UA](#)). In February, Facebook committed to “add language to our Dangerous Individuals and Organizations Community Standard within a few weeks explaining that we may remove content if the intent is not made clear [as well as to] add definitions of “praise,” “support” and “representation” within a few months.” On June 23, 2021, Facebook updated this standard to include definitions. The Board also recommended that Facebook clarify the relationship between Instagram’s Community Guidelines and the Facebook Community Standards ([case 2020-004-IG-UA](#)). As of June 29, Facebook has yet to inform the Board of its actions to implement this commitment.

In the present case, following a request from the Board, Facebook shared internal guidance for content moderators about the meaning of “support” of designated individuals and organizations. That defines a “call to action in support” as a call to direct an audience to do something to further a designated dangerous organization or its cause. This language was not reflected in the public-facing Community Standards at the time this content was posted and is not included in the update published on June 23, 2021.

Further to this, the misplaced and non-public guidance created in 2017 in response to Öcalan’s solitary confinement makes clear that discussions of the conditions of a designated dangerous individual’s confinement are permitted, and do not constitute support. In the absence of any other context, Facebook views statements calling for the freedom of a terrorist as support and such content is removed. Again, this language is not reflected in the public-facing Community Standards.

The Board is concerned that specific guidance for moderators on an important policy exception was lost for three years. This guidance makes clear that the content in this case was not violating. Had the Board not selected this case for review, the guidance would have remained unknown to content moderators, and a significant amount of expression in the public interest would have been removed.

This case demonstrates why public rules are important for users: they not only inform them of what is expected, but also empower them to point out Facebook’s mistakes. The Board

appreciates Facebook’s apprehension of fully disclosing its internal content moderation rules due to concerns over some users taking advantage of this to spread harmful content. However, Facebook’s policy of defaulting towards removing content showing “support” for designated individuals, while keeping key exceptions hidden from the public, allowed this mistake to go unnoticed by the company for approximately three years without any accountability. The June 2021 update to the Dangerous Individuals and Organizations Community Standard provides more information on what Facebook considers to be “support” but does not explain to users what exceptions could be applied to these rules.

Even without the discovery of the misplaced guidance, the content should not have been removed for “support.” This kind of call to action should not be construed as supporting the dangerous ends of the PKK. The user only encouraged people to discuss Öcalan’s solitary confinement, including through hunger strikes, protests, legal action, op-eds, reading groups, and memes. Accordingly, the removal of content in this case did not serve the policy’s stated aim of preventing and disrupting real-world harm.

8.2 Compliance with Facebook’s values

The Board found that Facebook’s decision to remove this content did not comply with Facebook’s values of “Voice” and “Safety.”

The user carefully crafted their post calling for a discussion about ending Öcalan’s isolation in prison on Imrali Island. They encouraged readers to discuss the inhumanity of solitary confinement and why it would be necessary to keep Öcalan confined in such a manner. The user advocated peaceful actions to provoke this discussion, and did not advocate for or incite violence in their post. They also did not express support for Öcalan’s ideology or the PKK.

The Board found that expression which challenges human rights violations is central to the value of “Voice.” This is especially important with reference to the rights of detained people, who may be unable to effectively advocate in support of their own rights, particularly in countries with alleged mistreatment of prisoners and where human rights advocacy may be suppressed.

The value of “Safety” was notionally engaged because the content concerned a designated dangerous individual. However, removing the content did not address any clear “Safety” concern. The content did not include language that incited or advocated for the use of

violence. It did not have the potential to “intimidate, exclude, or silence other users.” Instead, Facebook’s decision illegitimately suppressed the voice of a person raising a human rights concern.

8.3 Compliance with Facebook’s human rights responsibilities

Removing this content was inconsistent with the company’s commitment to respect human rights, as set out in its [Corporate Human Rights Policy](#). In relation to terrorist content, Facebook is a signatory of the [Christchurch Call](#), which aims to “eliminate” the dissemination of “terrorist and violent extremist content” online. While the Board is cognizant of [human rights concerns raised by civil society](#) concerning the Christchurch Call, it nevertheless requires social media companies to enforce their community standards “in a manner consistent with human rights and fundamental freedoms.”

1. Freedom of expression (Article 19 ICCPR)

Article 19 states that everyone has the right to freedom of expression, which includes freedom to seek, receive and impart information. The right to freedom of expression includes discussion of human rights (General Comment 34, para. 11) and is a “necessary condition for the realization of the principles of transparency and accountability” (General Comment 34, para. 3). Furthermore, the UN Declaration on Human Rights Defenders provides that everyone has the right to “study, discuss, form and hold opinions on the observance, both in law and in practice, of all human rights and fundamental freedoms and, through these and other appropriate means, to draw public attention to those matters” (A/RES/53/144, Article 6(c)).

The user sought to highlight concerns about an individual’s solitary and prolonged confinement. The Board notes that international bodies have raised human rights concerns about such practices. The Nelson Mandela Rules set out that states should prohibit indefinite as well as prolonged solitary confinement as a form of torture or cruel, inhuman, or degrading treatment of punishment ([A/Res/70/175](#), Rule 43). Discussing the conditions of any individual’s detention and alleged violations of their human rights in custody falls squarely within the types of expression protected by Article 19 of the ICCPR, as emphasized in the UN Declaration on Human Rights Defenders seeks to protect.

While the right to freedom of expression is fundamental, it is not absolute. It may be restricted, but restrictions should meet the requirements of legality, legitimate aim, and

necessity and proportionality (Article 19, para. 3, ICCPR). The Human Rights Committee has stated that restrictions on expression should not “put in jeopardy the right itself,” and has emphasized that “the relation between right and restriction and between norm and exception must not be reversed” (General Comment No. 34, para. 21). The UN Special Rapporteur on freedom of expression has emphasized that social media companies should seek to align their content moderation policies on Dangerous Individuals and Organizations with these principles (A/74/486, para. 58(b)).

a. Legality (clarity and accessibility of the rules)

Restrictions on expression should be formulated with sufficient precision so that individuals understand what is prohibited and act accordingly (General Comment 34, para. 25). Such rules should also be made accessible to the public. Precise rules are important for those enforcing them: to constrain discretion and prevent arbitrary decision-making, and also to safeguard against bias.

The Board recommended in [case 2020-005-FB-UA](#) that the Community Standard on Dangerous Individuals and Organizations be amended to define “representation,” “praise,” and “support,” and reiterated these concerns in [case 2021-003-FB-UA](#). The Board notes that Facebook has now publicly defined those terms. The UN Special Rapporteur on freedom of expression has described social media platforms prohibitions on both “praise” and “support” as “excessively vague” (A/HRC/38/35, para. 26; see also: General Comment No. 34, para. 46). In a public comment submitted to the Board (PC-10055), the UN Special Rapporteur on human rights and counter-terrorism noted that although Facebook has made some progress to clarify its rules in this area, “the Guidelines and Standard are [still] insufficiently consistent with international law and may function in practice to undermine certain fundamental rights, including but not limited to freedom of expression, association, participation in public affairs and non-discrimination.” Several public comments made similar observations.

The Board noted Facebook provides extensive internal and confidential guidance to reviewers to interpret the company’s public-facing content policies, to ensure consistent and non-arbitrary moderation. However, it is unacceptable that key rules on what is excluded from Facebook’s definition of support are not reflected in the public-facing Community Standards.

b. Legitimate aim

Restrictions on freedom of expression should pursue a legitimate aim. The ICCPR lists legitimate aims in Article 19, para. 3, which includes the protection of the rights of others. The Board notes that because Facebook reversed the decision the user appealed against, following the Board's selection of that appeal, the company did not seek to justify the removal as pursuing a legitimate aim but instead framed the removal as an error.

The Community Standards explain that the Dangerous Individuals and Organizations policy seeks to prevent and disrupt real world harm. Facebook has previously informed the Board the policy limits expression to protect "the rights of others," which the Board has accepted ([case 2020-005-FB-UA](#)).

c. Necessity and proportionality

Restrictions on freedom of expression should be necessary and proportionate to achieve a legitimate aim. This requires there to be a direct connection between the expression and a clearly identified threat (General Comment 34, para. 35), and restrictions "must be the least intrusive instrument amongst those which might achieve their protective function; they must be proportionate to the interest to be protected" (General Comment 34, para. 34).

As Facebook implicitly acknowledged by reversing its decision following the Board's selection of this case, the removal of this content was not necessary or proportionate. In the Board's view, the breadth of the term "support" in the Community Standards combined with the misplacement of internal guidance on what this excludes, meant an unnecessary and disproportionate removal occurred. The content in this case called for a discussion about ending Öcalan's isolation in prolonged solitary confinement. It spoke about Öcalan as a person and did not indicate any support for violent acts committed by him or by the PKK. There was no demonstrable intent of inciting violence or likelihood that leaving this statement or others like it on the platform would result in harm.

The Board is particularly concerned about Facebook removing content on matters in the public interest in countries where national legal and institutional protections for human rights in particular freedom of expression are weak (cases [2021-005-FB-UA](#) and [2021-004-FB-UA](#)). The Board shares the concern articulated by the UN Special Rapporteur on human rights and counter-terrorism in her submissions, that the "sub-optimal protection of human rights on the platform [...] may be enormously consequential in terms of the global protection

of certain rights, the narrowing of civic space, and the negative consolidation of trends on governance, accountability and rule of law in many national settings.” The Board notes that

the UN Special Rapporteur on freedom of expression has expressed specific concerns in this regard on Turkey ([A/HRC/41/35/ADD.2](#)).

II. Right to remedy (Article 2 ICCPR)

The right to remedy is a key component of international human rights law ([General Comment No. 31](#)) and is the third pillar of the UN Guiding Principles on Business and Human Rights. The UN Special Rapporteur on freedom of expression has stated that the process of remediation “should include a transparent and accessible process for appealing platform decisions, with companies providing a reasoned response that should also be publicly accessible” (A/74/486, para 53).

In this case, the user was informed an appeal was not available due to COVID-19. However, an appeal was then carried out. The Board once again stresses the need for Facebook to restore the appeals process in line with recommendations in cases [2020-004-IG-UA](#) and [2021-003-FB-UA](#).

While the user in this case had their content restored, the Board is concerned at what may be a significant number of removals that should not have happened because Facebook lost internal guidance which allowed for discussion on conditions of confinement for designated individuals. Facebook informed the Board that it is undertaking a review of how it failed to transfer this guidance to its new review system, as well as whether any other policies were lost. However, in response to a Board question, the company said that “it is not technically feasible to determine how many pieces of content were removed when this policy guidance was not available to reviewers.”

The Board is concerned that Facebook’s transparency reporting is not sufficient to meaningfully assess if the type of error identified in this case reflects a systemic problem. In questions submitted to Facebook, the Board requested more information on its error rates for enforcing its rules on “praise” and “support” of dangerous individuals and organizations. Facebook explained that it did not collect error rates at the level of the individual rules within the Dangerous Individuals and Organizations policy, or in relation to the enforcement of specific exceptions contained only in its internal guidance. Facebook pointed the Board to

publicly available information on the quantity of content restored after being incorrectly removed for violating its policy on Dangerous Individuals and Organizations. The Board notes that this does not provide the same kind of detail that would be reflected in internal audits and quality control to assess the accuracy of enforcement. While Facebook acknowledged it internally breaks down error rates for enforcement decisions by moderators and by automation, it refused to provide this information to the Board on the basis that “the information is not reasonably required for decision-making in accordance with the intent of the Charter.”

Furthermore, the Board asked Facebook whether content is appealable to the Board if it has been removed for violating the Community Standards following a government flagging the content. Facebook confirmed such cases are appealable to the Board. This is distinct from those removed on the basis of a government requesting removal to comply with local law, which are excluded from review by Article 2, Section 1.2 of the Bylaws. While the Board does not have reason to believe that this content was the subject of a government referral, it is concerned that neither users whose content is removed on the basis of the Community Standards, nor the Board, are informed where there was government involvement in content removal. This may be particularly relevant for enforcement decisions that are later identified as errors, as well as where users suspect government involvement but there was none. Facebook’s transparency reporting is also limited in this regard. While it includes statistics on government legal requests for the removal of content based on local law, it does not include data on content that is removed for violating the Community Standards after a government flagging the content.

This collection of concerns indicates that Facebook is failing to respect the right to remedy, in contravention of its Corporate Human Rights Policy (Section 3).

9. Oversight Board decision

The Oversight Board overturns Facebook's original decision to take down the content, requiring the post to be restored. The Board notes that Facebook has accepted that its original decision was incorrect and has already restored the content.

10. Policy advisory statement

As noted above, Facebook changed its Community Standard on Dangerous Individuals and

Organizations after asking the Board to provide guidance on how this Community Standard should function. These recommendations take into account Facebook's updates.

The misplaced internal guidance

Pending further changes to the public-facing Dangerous Individuals and Organizations policy, the Board recommends Facebook should take the following interim measures to reduce the erroneous enforcement of the existing policy:

1. Immediately restore the misplaced 2017 guidance to the Internal Implementation Standards and Known Questions (the internal guidance for content moderators), informing all content moderators that it exists and arranging immediate training on it.
2. Evaluate automated moderation processes for enforcement of the Dangerous Individuals and Organizations policy and where necessary update classifiers to exclude training data from prior enforcement errors that resulted from failures to apply the 2017 guidance. New training data should be added that reflects the restoration of this guidance.
3. Publish the results of the ongoing review process to determine if any other policies were lost, including descriptions of all lost policies, the period the policies were lost for, and steps taken to restore them.

Updates to the Dangerous Individuals and Organizations policy

Facebook notified the Board that it is currently working on an update to its policies to make clear that its rules on "praise" and "support" do not prohibit discussions on the conditions of confinement of designated individuals or other violations of their human rights.

As an initial contribution to this policy development process, the Board recommends that Facebook should:

4. Reflect in the Dangerous Individuals and Organizations "policy rationale" that respect for human rights and freedom of expression, in particular open discussion about human rights violations and abuses that relate to terrorism and efforts to counter terrorism, can advance the value of "Safety," and that it is important for the platform to provide a space for these discussions. While "Safety" and "Voice" may sometimes be in tension, the policy rationale

should specify in greater detail the “real-world harms” the policy seeks to prevent and disrupt when “Voice” is suppressed.

5. Add to the Dangerous Individuals and Organizations policy a clear explanation of what “support” excludes. Users should be free to discuss alleged violations and abuses of the human rights of members of designated organizations. This should not be limited to detained individuals. It should include discussion of rights protected by the UN human rights conventions as cited in Facebook’s Corporate Human Rights Policy. This should allow, for example, discussions on allegations of torture or cruel, inhuman, or degrading treatment or punishment, violations of the right to a fair trial, as well as extrajudicial, summary, or arbitrary executions, enforced disappearance, extraordinary rendition and revocation of citizenship rendering a person stateless. Calls for accountability for human rights violations and abuses should also be protected. Content that incites acts of violence or recruits people to join or otherwise provide material support to Facebook-designated organizations should be excluded from protection even if the same content also discusses human rights concerns. The user’s intent, the broader context in which they post, and how other users understand their post, is key to determining the likelihood of real-world harm that may result from such posts.

6. Explain in the Community Standards how users can make the intent behind their posts clear to Facebook. This would be assisted by implementing the Board’s existing recommendation to publicly disclose the company’s list of designated individuals and organizations (see: case 2020-005-FB-UA). Facebook should also provide illustrative examples to demonstrate the line between permitted and prohibited content, including in relation to the application of the rule clarifying what “support” excludes.

7. Ensure meaningful stakeholder engagement on the proposed policy change through Facebook’s Product Policy Forum, including through a public call for inputs. Facebook should conduct this engagement in multiple languages across regions, ensuring the effective participation of individuals most impacted by the harms this policy seeks to prevent. This engagement should also include human rights, civil society, and academic organizations with expert knowledge on those harms, as well as the harms that may result from over-enforcement of the existing policy.

8. Ensure internal guidance and training is provided to content moderators on any new policy. Content moderators should be provided adequate resources to be able to understand the

new policy, and adequate time to make decisions when enforcing the policy.

Due process

To enhance due process for users whose content is removed, Facebook should:

9. Ensure that users are notified when their content is removed. The notification should note whether the removal is due to a government request or due to a violation of the Community Standards or due to a government claiming a national law is violated (and the jurisdictional reach of any removal).

10. Clarify to Instagram users that Facebook’s Community Standards apply to Instagram in the same way they apply to Facebook, in line with the recommendation in [case 2020-004-IG-UA](#).

Transparency reporting

To increase public understanding of how effectively the revised policy is being implemented, Facebook should:

11. Include information on the number of requests Facebook receives for content removals from governments that are based on Community Standards violations (as opposed to violations of national law), and the outcome of those requests.

12. Include more comprehensive information on error rates for enforcing rules on “praise” and “support” of dangerous individuals and organizations, broken down by region and language.

***Procedural note:**

The Oversight Board’s decisions are prepared by panels of five Members and approved by a majority of the 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.

