***The Case of Call for women's protest in Cuba***

**Case Analysis**

***Meta-Data*:**

* **Case Number**:
* **Date of decision**:
* **Featured case**: N/A
* **Region**:
* **Country**:
* **Type of expression**:
* **Judicial Body**:
* **Type of law**:
* **Main Themes**:
* **Outcome**:
* **Status**:
* **Tags**:

***Analysis:***

* **Summary and Outcome**:

On October 3, 2023, the Oversight Board overturned Meta’s decision to remove a video of a woman criticizing men by comparing them to animals and encouraging women to join her posted by a Cuban news platform on Instagram. The Board found the women’s statements to be a qualified behavioral statement allowed under Meta’s Hate Speech policy. The Board further highlighted the importance of external context analysis in Meta’s decisions, especially with content from countries where freedom of expression and peaceful assembly are suppressed or met with violence.

* **Facts**:

In July 2022, a Cuban news platform’s Instagram account posted a video of a woman calling other women to protest alongside her. In the video the woman called Cuban men “rats” for not standing up to repression by the government and “mares”. The text overlying the video connected political change and women’s protests. The caption was in Spanish and it included quotes from the video and hashtags referring to the “dictatorship” and “regime” in Cuba, and called for international attention to the humanitarian situation there by using #SOSCuba. The video gathered 90,000 plays and fewer than 1,000 shares.

The post was sharing around the anniversary of nationwide protests that occurred in July 2021 in Cuba in response to lack of economic, social, and cultural rights. The protests were the largest in Cuba’s recent history. Between July 2021 and July 2022, state repression intensified, and the post was shared in a very tense social context. A few days before the post, the Cuban police killed a young Cuban man and the woman in the video referred to the incident by saying, “we cannot keep allowing the killing of our sons.”

Furthermore, while the terms “rats” and “mares” might be offensive, they were not dehumanizing as they are used by Cubans to imply cowardice. The term “mares” in particular is often used as a homophobic slur, however, its combination with the reference of urinal pots allude to its usage to show discontent towards men as experts have submitted to the Board.

Furthermore, experts highlighted the importance of social media to raise awareness about the struggles of Cubans as in the wake of the 2021 protests, the Cuban government has intensified restraints on all forms of dissent and public criticism as well as criminalization of online speech through introducing new imprecise crimes with severe sanctions in the Cuban new penal code.

Moreover, only a small fraction of Cubans has access to independent news websites and blogs due to near-complete government control over the internet, censorship, obstruction of communications and a very high cost of accessing the internet. Meta’s February 2023 report on Adversarial Threat found links between the Cuban government and a network of 363 Facebook accounts, 270 pages, 229 groups, and 72 Instagram accounts violating Meta’s policy against coordinated inauthentic behavior.

Seven days after the video was posted, it was identified by a hostile speech classifier as potentially violating and sent it for human review. The day after the content was removed for violating Meta’s Hate Speech policy, the content was later escalated for secondary review due to its [cross-check status](https://www.oversightboard.com/decision/PAO-NR730OFI/). The first moderator found the post violating on July 21, 2022, and the second moderator found the post violating on February 24, 2023, and the content was removed seven months after it was initially removed, and a standard strike was applied to the user’s account without any feature limit. The delay in the review was because a backlog in Meta’s review queues under the cross-check systems.

On the same day of the removal, the user appealed the decision. Another moderator reviewed the content on February 26, 2023, and upheld the decision, and the content was escalated for additional review. The user then appealed the case to the Board.

* **Decision Overview**:

The main issue before the Oversight Board was whether the removal of the post was consistent with Meta’s content policies, values, and human rights obligations.

In their appeal to the Board, the user urged social media companies to better understand the situation in Cuba. The user explained that the video was referencing the July 2021 protests, and that the woman was calling Cuban men to solve the crisis.

On the other hand, Meta submitted that the post was removed under Tier 1 of the Hate Speech policy which prohibits content comparing on group to "animals that are culturally perceived as intellectually or physically inferior”. Meta admitted that the company did not understand the cultural meaning of "mares loaded with chamber pots or toilets" but found the phrase violating of the policy because it compared men to animals carrying human feces. Meta explained that comparing men to rats and toilet laden horses dehumanized men based on their sex and might exclude them from the discussion.

The Board asked 17 written questions. The questions covered topics such as Meta's content moderation strategy in Cuba, the impact of asymmetrical power dynamics on the Hate Speech policy and its enforcement after both automated and human review, and opportunities for context assessment, particularly in relation to the Early Response Secondary Review (ERSR) section of Meta's cross-check system. ERSR is a kind of cross-checking that keeps content available while offering further human evaluation for specific postings that were first found to be against Meta's regulations. Meta responded to all 17 questions.

In one of its answers, Meta explained that it considered applying a “spirit of the policy” allowance in this case, which allows content to remain on the platform which strict enforcement of a policy produce results inconsistent with the policy’s rationale or objectives. Meta refrained from this because they deemed the content violating of both the letter and the spirit of the policy. Meta further explained that it treats all groups defined by protected characteristics equally and removes any attacks against any group even if the attacked group is more culturally and socially powerful in what it refers to as the "protected characteristic-agnostic" approach. In response to another question, Meta explained that its classifies only analyze the context of the post itself and disregards any wider social or political context.

*Compliance with Meta’s content policies and values*

1. Content Rules and Values

The Board found the content to be a qualified behavioral statement and not hate speech, thus non-violating of the Hate Speech policy and removing it was inconsistent with the policy. The Board recognized that the comparison of men to “rats” or “mares loaded with urinal pots could be interpreted as violating of the policy if read out of context. However, the post, when read in whole and in context, was not aiming to dehumanize men or incite to violence against them. The Board noted that these statements were qualified as they called attention to the behavior of Cuban men in the context of the situation in Cuba and that it was just a commentary about how men acted. This is evident by the usage of the #SOSCuba hashtag.

Furthermore, the Board referred to the public comments and experts who have submitted that “rats” and “mares” are used in Cuba to imply cowardice. The Board also highlighted the context where the content was posted as the content was posted during a wave of state repression. The Board noted the context, the usage of hashtag, statements such as “we cannot keep allowing the killing of our sons” and the woman’s call to the street as indicators of the content being an opinion about the behavior of Cuban men during the context of the on-going struggle of Cubans.

Moreover, a minority of the Board questioned the “protected characteristic-agnostic" approach, especially in cases where its enforcement led to restricting historically marginalized groups. Those members suggested that a proportionate Hate Speech policy should acknowledge societal and cultural power dynamics to prevent stifling under-represented groups.

Finally, the Board agreed that the content fell within the value of “Voice” and that its removal was inconsistent with Meta’s values. This approach adopted in this case echoes the approach in the “Violence Against Women” decision when the Board decided that content should be read as whole and recognized it as a qualified behavioral statement.

1. Enforcement action

The Board noted the 7-month delay in this case. The delay meant the content remained on the platform until cross-check secondary review was done which aligns with the Board’s analysis of the Hate Speech policy, however the outcome did not align with Meta’s decision that the content was harmful.

The Board expressed concern over how contextual information was considered in decisions on content benefiting from additional human review. The Board recalled its stance on how challenge it was to assess hate speech and relevant context at scale, especially in cases of dehumanizing speech, whether implicit or explicit, which has previously led to atrocities. Furthermore, the Board reiterated that content moderation with the aim to address cumulative harms of hate speech might be consistent with Meta’s human rights responsibilities even if some content didn’t directly incite to violence or discrimination.

The Board highlighted that Meta established policy exceptions in order to avoid stifling public debate on relevant issues such as political speech on historical events in the case at hand. This was why it is necessary for Meta’s reviewers to be able to distinguish between qualified and unqualified behavioral statements, and it is necessary for Meta’s automated systems and reviewers to be able to factor contextual information in their decisions and not rely solely on the post’s context like what happened in this case. The Board additionally noted that reviews at escalation level were expected to deliver better results as they had better access to context, however, even after escalated review, Meta failed to issue the right decision.

The markets team examines the material first for ERSR. This group has more linguistic and contextual expertise on a certain region. The market in Cuba is not an independent market and it is categorized in general Spanish language queues, meaning that reviewers review Spanish language material in general and are not focused exclusively on that country. After then, the Early Response team might examine content to determine whether it is violating. This team may additionally make use of Meta's "newsworthiness" and "spirit of the policy" allowances since they have "deeper policy expertise and the ability to factor in additional context". Nevertheless, lacking language or regional experience, the Early Response team evaluates the material based on translations and contextual data from the applicable Regional Market team.

Meta’s decision in the case raised the Board’s concern regarding contextual information was not sufficiently considered during the cross-check escalations process. In response the Board reiterated that Meta must improve its consideration of context and language expertise in its workflow to prevent future similar content from being removed.

*Compliance with Meta’s human rights responsibilities*

Freedom of expression, including political speech, is protected under 19 of the International Covenant on Civil and Political Rights (ICCPR) and the right to freedom of peaceful assembly is protected under article 21 of the ICCPR and extends to related online activities. The Board noted that Meta’s fulfillment of its protection of these rights is crucial due to the extreme restrictions on them in Cuba.

Furthermore, the Board underlined the heightened protection the content in the case has due to its nature as a woman’s call for protest to defend rights at a significant political moment. The Board also highlighted the continuation of public anger and criticism of Cuban government for their repression of Cubans.

The Board implemented the three-part test established in article 19 of the ICCPR to analyze whether Meta fulfilled its human rights obligations.

1. Legality (clarity and accessibility of the rules)

The principle of legality requires rules to be specific, clear, and accessible. Additionally, unfettered discretion for restricting freedom of expression shouldn’t be given to those who execute the rules. This means Meta users should have access and understand the rules and content reviewers must have clear guidance on enforcement.

The Hate Speech policy prohibits attacks on groups based on their protected characteristics, dehumanizing speech is considered an attack under the policy. Dehumanizing speech includes comparisons, generalizations, or unqualified behavioral statements about or to animals culturally perceived as inferior. However, the same policy permits qualified behavioral statements. Meta defines unqualified behavior statements as explicit attribution of a behavior to all, or a majority of a group defined by a protected characteristic. Meta admitted to the Board in the “Violence against women” decision that it can be hard for reviewers to distinguish between qualified and unqualified behavioral statements if they don’t consider the context. The Board noted that the guidance to reviewers limits their ability to analyze content even if there are clear cues in the context itself that it is a qualified behavioral statement. Meta stated that it instructs its reviewers to default to removing statements if the user has not made it clear whether the statement is qualified or not due to the challenging nature of determining intent.

The Board noted that in the case at hand it was clear the content was a statement commenting on what the woman recognized as a pattern during specific historical and conflict events as she was referring to Cuban men in the context of the 2021 protests and the 2022 repressions. Furthermore, the inclusion of #SOSCuba made it clearer.

The Board reiterated that content reviewers should have the opportunities and resources to do contextual analysis while enforcing policies. The Board found the language of the policy and internal review vague as the company’s confusing and contradictory guidance made it challenging for reviewers to reach the right conclusion.

1. Legitimate aim

For a restriction on freedom to be valid, it must pursue one of the legitimate aims outlined in the ICCPR. The Board has previously found the Hate Speech policy to in pursuant of protecting people from the harm of hate speech.

1. Necessity and proportionality

For a restriction to be necessary and proportionate, it must correspond a pressing social need and the least restrictive measure to achieve the aim pursued. While the Board disagreed with Meta’s classification of the content as hate speech, it recognized the challenges of hate speech moderation, especially hate speech comparing others to animals. The Board acknowledged that some of Meta’s restrictions wouldn’t be valid with the human rights obligations of states, in those cases social media companies are expected to give reasoned explanations of their restrictive policies.

The Board noted that in Meta’s decision not to apply any policy exceptions was due to Meta’s failure to consider important context which made the decision neither necessary nor proportionate.

The Board implemented the Rabat Plan in its analysis with the focus on the elements of social and political context, the author, and the content. The Board considered the strong wave of repression following the 2021 protests and the death of a young Cuban man at the hands of the policy as relevant context that catalyzed calls for protesting the government. As for the content, the woman was expressing her opinion on the behavior of Cuban men during the events in Cuba and calling on women to protest to defend the lives of “our sons” and the post explicitly featured hashtag #SOSCuba in an explicit reference to the protests. The Board found the post not inciting to violence against men as it did not attribute a behavior to them or dehumanize them based on their gender, instead it used harsh language to urge Cuban men to join the protests. The Board highlighted the significant negative impact the removal had on both the woman in the video, the user who shared it and on political debate.

The Board found the removal decision disproportionate on the women in the video and the user who have come many obstacles existing in Cuba, including access to the internet and risks of criticizing the government. Furthermore, the strike applied to the user’s account could have imposed further unnecessary burden if it has resulted in the account’s suspension. Additionally, the Board noted that the post did not incite to violence and was in the public interest, therefore, the removal also affected public debate in Cuba, where it is already significantly limited.

The Board reaffirmed that context analysis is crucial in decision making for Meta, especially in the possibility of political speech. In the “Colombia Protests” decision, the Board analyzed the political context and public interest of using a homophobic slur during a protest. Furthermore, the Board acknowledged in the “Iran Protest Slogan” decision that Meta’s over-removal of political expression created risks to human rights in Iran. Finally, in the “Pro-Navalny Protests in Russia” decision the Board underlined the importance of external content analysis in enforcement of Meta’s policies.

The Board highlighted the physical and legal risks that come with criticizing the government. These risks alongside the difficulty of internet access in Cuba raised the stakes of content moderation from content in the country. Furthermore, the Board underlined the persecution of women in response to the July 2021 protests in Cuba as women are subjected to political and sexual violence.

The Board urged Meta to exercise more care with content from regions were political speech and peaceful assembly ate suppressed or met with violence. Social media platforms in Cuba offer a limited but significant means for government criticism and social activism. Meta stated that it implemented risk mitigation measures during the July 2021 protests and November 2021 event in Cuba, however, it did not provide any risk mitigation measures at the time the content at hand was posted.

Ultimately, the Oversight Board overturned Meta’s decision to remove the post.

*Recommendations*

The Board decided not to issue any recommendations and instead reiterated recommendations no. 1, no.3, and no. 8 from its cross-check policy advisory opinion. Recommendation no. 1 urged Meta to safeguard expression by implementing a list-based over-enforcement prevention program. This system should be separate from the one that protects expression, which Meta considers to be a commercial priority. It should also ensure that content submitted by human rights advocates and other parties receives further review by Meta.

Recommendation no. 3 recommended that Meta integrated linguistic and contextual knowledge for better assessment, particularly at the decision-making stages. Finally, recommendation no. 8 recommended Meta should use specialized staff, with the benefit of local input, to create over-enforcement prevention lists.

Additionally, the Board recalled recommendation no. 2 from the “Violence Against Women” decision for Meta to update its guidance on what constitutes qualified behavioral statement as the current guidance makes it almost impossible for reviewers to reach the correct decision.

***Direction:***

* **Outcome**: Contracts Expression/Mixed Outcome/Expands Expression
	+ This should be based on international standards
	+ However, if you have knowledge of national standards, and can provide insights into how the decision impact precedent nationally, please do so.
* **Explanation for why and how it contracts or expands expression or has a mixed outcome.** You can also provide additional context about the case here.

***Perspective***:

* **Related International and/or regional laws**:

**Example:**

[ECHR, art. 10](https://www.echr.coe.int/Documents/Convention_ENG.pdf);

[ECHR, art. 11](https://www.echr.coe.int/Documents/Convention_ENG.pdf);

* **National law or jurisprudence**:
* **Example:** [Sp. Constitution art. 14](http://www.congreso.es/portal/page/portal/Congreso/Congreso/Hist_Normas/Norm/const_espa_texto_ingles_0.pdf);

**Other national law or jurisprudence**:

* **List here any references to national case law outside the Court’s jurisdiction**. For instance, if a UK Court relies on Canadian or Australian case law, it would be listed here.

***Significance***:

* **Binding or persuasive precedent within jurisdiction**; **Decision establishes influential or persuasive precedent outside jurisdiction**; **Explanation**:

Standard I: The decision establishes a binding or persuasive precedent within its jurisdiction.

Information: i.e. Judgments of the European Court of Human Rights are binding upon parties to the decision.

Standard II: Decision (including concurring or dissenting opinions) establishes influential or persuasive precedent outside its jurisdiction.

Information: i.e. The decisions of the European Court of Human Rights have precedential value on the interpretation of the right to freedom of expression for other States Parties to the European Convention for the Protection of Human Rights and Fundamental Freedoms.

* **Related Cases**: Self-generated
* **Date updated**: N/A

***Docs***:

* **Official Case Documents**:

**Examples:**

Judgment (in English) [Attached]

Press Release issued by …. (in English) [Attached]

* **Reports, Analysis, and News Articles**:
	+ Please list any titles and URLS of interesting articles or analyses about the case that might be of interest to readers