



Twitter UK  
20 Air Street  
London  
W1B 5AN

twitter.com

**29 September 2021**

Dear Chair,

Thank you for inviting me to provide oral evidence earlier this month. I am writing to follow up on a number of issues raised during the discussion.

### **Rules and Enforcement**

During and following the hearing, the Committee asked us to review three hypothetical Tweets. Our policy is not to review hypothetical Tweets, as it is generally not an accurate representation of how rules are enforced. This is because some Tweets that seem abusive when viewed in isolation may not be when viewed in the context of a larger conversation (for example, documenting an incident that occurred offline, or sharing a trope satirically in order to discredit it). In addition, when we review this type of content, it may not be clear whether it is intended to harass an individual, or if it is part of a consensual conversation (for example, friends or family who may engage in satirical discussion that could appear abusive to bystanders). Indeed, at times we may need to hear directly from the person being targeted, to ensure that we have the information needed prior to taking any enforcement action.

Given the Committee's concerns, however, we wanted to clarify that our policy on [Hateful Conduct](#) prohibits abuse that targets individuals or groups of people belonging to protected categories. This includes content intended:

- To incite fear or spread fearful stereotypes about a protected category, including asserting that members of a protected category are more likely to take part in dangerous or illegal activities, e.g., "all [religious group] are terrorists."
- To incite others to harass members of a protected category on or off platform, e.g., "I'm sick of these [religious group] thinking they are better than us, if any of you see someone wearing a [religious symbol of the religious group], grab it off them and post pics!"
- To incite others to discriminate in the form of denial of support to the economic enterprise of an individual or group because of their perceived membership in a protected category, e.g., "If you go to a [religious group] store, you are supporting those [slur], let's stop giving our money to these [religious slur]." This may not include content intended as

political in nature, such as political commentary or content relating to boycotts or protests.

We are, however, always alert to potential gaps in our policies and enforcement. A global review of this policy is currently underway and will be informed by recent events, including the hearing in which we participated earlier this month. I can provide a further written update on that work early in 2022.

### **Identifiable accounts**

Following Euro 2020, we undertook our own analysis of the Tweets removed and accounts suspended. We wanted to better understand the users we had permanently suspended over the course of the tournament. While we have always welcomed the opportunity to hear ideas from partners on what will help, including from within the football community, our data suggests that ID verification would have been unlikely to prevent the abuse from happening - as the overwhelming majority of the accounts we suspended were not anonymous. Of the permanently suspended accounts from the Tournament, 99% of account owners were identifiable.

During the hearing, the Committee asked why the 1% would be not identifiable. To confirm, our analysis suggests this represents historic users who set up their Twitter accounts prior to changes we introduced in 2018 requiring verification of an account upon sign-up with an email address or phone number.

We do not have an equivalent percentage for all users who would be identifiable, as it would involve analysing hundreds of millions of accounts. We shared the data for the Euros, however, as we were able to assess this information for a limited number of users. This does not prohibit external research being undertaken using public signals looking at relationships of identity and abuse online.

### **Signify research**

During the hearing, the Chair shared research published by Signify, and commissioned by the PFA. I can confirm that we have reviewed all 1772 Tweets shared, and removed 445.

However, there appears to be some disconnect between our data, and that of Signify's. For example, in reviewing the data they shared, we saw that a number of Tweets were already deleted. Of the 1772 Tweets shared, 30% had already been removed. In addition, 69% of what they reported was from last year - which predates a number of changes we have made.

We determine whether or not to take down a Tweet based on whether or not it breaks our rules - a number of Tweets did not. As described above, we acknowledge there is ongoing work to be done when it comes to our global policies. However, while there were clearly some egregious violations, which we did remove, it is important to note that there was some discrepancy

between Signify's definition of abuse and our own. For instance, while homophobic abuse is against our rules, describing someone as gay is not something we consider abusive of itself.

With the above in mind, we have invited Signify to meet.

### **Our staff**

Twitter's purpose is to serve the public conversation, and we want to ensure our company reflects our service. This means building a team at Twitter as diverse as the people who use it. This includes the professionals we hire to lead our work on safety on Twitter - many of whom come from an activist background, have directly experienced these issues, and/or have led work on online safety prior to joining the company.

We have established objectives in order to attract, retain, and develop the talent necessary to deliver on our commitment. Critically, we've also been focusing on transparency around workforce representation. One example is through our regular [Inclusion and Diversity \(I&D\) reports](#), which outline our progress in making Twitter a diverse and global workforce.

In 2020, we announced a set of new diversity goals for 2025: at least half of our global staff will be women and, in the US, at least a quarter will be underrepresented minorities, specifically Black, Latinx, Native American, Alaskan or Hawaiian/Pacific Islander, or Multiracial.

Starting in 2021, our executive team's annual compensation will be determined in part by their participation in our inclusion and diversity efforts, including their contributions toward meeting these workforce representation targets.

As of June 2021:

- Women make up just under 44% of our global workforce, up 15% since 2017. Women also hold 38% of all leadership roles, up 16% since 2017.
- The representation of Black members of staff in the US has more than doubled since 2017 to 8% of our team and 7% of our leadership (director level and above).
- The representation of Asian members of staff in the US is 29%, an increase of just under 12% since 2017, and 18% of our leadership (director level and above).
- Latinx colleagues make up 7% of our US staff, an increase of 81% since 2017, and 4% of our leadership (director level and above).

Our quarterly progress reports are available [here](#).

Please do let us know if you have any further questions.

Yours sincerely,

Katy Minshall  
Head of UK Public Policy