Qualitative research into the impact of online hate

Final report

Ofcom

TRAVERSE

FEBRUARY 2023

CONTENT WARNING

This report explores the impact of online hate, including hate speech and hateful abuse, relating to people with different protected characteristics.

While offensive words and phrases have not been included, some people may find the content of this report distressing.

TRAVERSE

Ofcom's preface

Ofcom has a statutory duty to promote and research media literacy. A key way we seek to fulfil this duty is through our *Making Sense of Media* programme, which aims to help improve the online skills, knowledge and understanding of children and adults in the UK. Ofcom was also given powers in autumn 2020 to regulate UK-established video-sharing platforms (VSPs). And Ofcom is to gain new responsibilities as the regulator for online safety in the UK, under the Online Safety Bill, which is currently in Parliament.

This report is one in a series of research studies into online safety that will inform our preparations for implementing the new online safety laws. As part of these preparations, we are building a robust evidence base, bringing together internal and external data, collected using different methods, from a variety of different sources. This programme of research further develops our understanding of online harms and how we can help to promote a safer user experience.

Ofcom commissioned independent research organisation Traverse to carry out a qualitative study to understand the impact of exposure to online hate and hateful abuse on people with protected characteristics, focusing on content found on user-to-user services.

The findings should not be considered a reflection of any final policy position that Ofcom may adopt when we take up our role as the online safety regulator.

Executive summary

Traverse is an independent research organisation – we were commissioned by Ofcom to carry out qualitative research with a diverse sample of **39** people who had experienced online hate and hateful abuse. Eleven of these participants also participated in an online task and online discussion. The fieldwork took place in April and May 2022.

The following working definitions were co-developed with Ofcom to use in the research. These are not the full and legal definitions of these terms and the definitions do not directly correspond to the definitions under VSP regulation, how illegal or harmful content will be defined under the Online Safety Bill, or definitions in platforms' terms of service. Simplified definitions were used within the research to support a common understanding of key terms amongst participants.

Protected characteristic: A protected characteristic is a characteristic that someone may have or be perceived to have which means they might be discriminated against. Examples of protected characteristics include gender, disability, race, religion or belief, sexual orientation or whether someone is transgender.

Online hate: hateful content directed at **a group of people** on the basis of a particular protected characteristic.

Hateful abuse: hateful content directed at **an individual** on the basis of a protected characteristic they have or are perceived to have.

Executive summary

The key findings of the research are below:

1. Participants reported that being exposed to **online hate** is a common feature of their online experience**.** Frequency of exposure often **increased around particular events** (e.g. Euro 2020).
2. The frequency and types of **hateful abuse** experienced were strongly determined by context, including **how often** participants used different platforms and how they used them (e.g. those who frequently 'pushed back' against other users perceived as hateful, or those who shared a lot about themselves could experience it more often).
3. Impacts tended to be **more pronounced** where content targeted characteristics. The emotional and psychological impacts of hateful content included:
   - **Surprise and shock** – especially where they saw or received hate where it was unexpected.
   - **Anger and disappointment** – because it suggested that these views were more prevalent than they had realised; because they felt incensed by the hateful behaviour; alongside anger and frustration where hateful behaviour had evaded moderation.
   - **Embarrassment and shame** – especially where the experience played out in open/public spaces where friends, family and strangers could see the exchanges or become involved.
   - **Anxiety and fear** – in the abuse context, participants sometimes felt threatened and feared for their safety or felt uncertain about who was targeting them, how long it would go on for and if it might 'flare-up' again.
   - **Hopelessness and exhaustion** – because online hate was so pervasive and taking action often didn't lead to desired outcomes, some participants became desensitised and no longer reported hateful content.

Executive summary

4. In terms of **behaviours**, anxiety and fear could lead to participants limiting what they shared/expressed or where they went online. **Offline**, they could also become more guarded and less trusting of others; participants described feeling less at ease when in public/ interacting with people they did not know due to the fear people could be harbouring similar views to those they had experienced online.
5. The types of **online** reactions and coping strategies in response to hateful content included:
   - **Blocking and reporting**
   - **Challenging and engaging**
   - **Seeking support**
   - **Self-censoring and retreating**
6. The research found that some felt **compelled** or **duty bound** to challenge hate (as part of an 'activist' identity); while others sought to **minimise exposure to harmful content.**
7. The factors that could make the hateful content **more severe** included:
   - **Scale and exposure:** where participants were targeted by multiple users or seeing hate from many users tended to increase severity.
   - **Intensity and threat level:** where language and behaviour was particularly malicious, prolonged or threatening.
   - **The inability of those targeted to take/ see action:** where participants did not feel able to get the content removed and felt users doing it faced no consequences.
   - **The characteristics of those being targeted:** where multiple characteristics were targeted (intersectionality), and where the characteristics targeted were things about themselves they could not hide.
   - **The format and nature of the content:** where it was felt that the content had the potential to be widely shared and could normalise the views.
   - **Who was being hateful:** where the hateful behaviour came from people with status, influence or from people who participants knew and previously trusted.

Executive summary

8. Despite the pervasiveness of online hate and abuse, participants often wanted to protect free speech and it was felt almost unanimously that mandatory user verification via uploading a form of ID was not a good idea. However, whilst freedom of speech was valued it was common to say that there should not be freedom from consequences. Harming and threatening others was often seen as the 'red line' in terms of free speech and could have a chilling effect on others.
9. Hateful content experienced was mostly seen as **not compliant** with platform's policies. Participants called for platforms to have more **active and consistent moderation** and to consider the capacity, skills and make up of their staff.
10. Participants felt that **platforms had the primary responsibility** to moderate/remove hateful content in line with their policies and the law.
11. It was felt that a regulator should ensure that platforms are following rules and are taking **robust action** to enforce their own policies or removing any illegal content**.**
12. Participants also thought a regulator should be **promoting best practice** by sharing examples of how best to tackle online hate and abuse.
13. There were also calls for a greater emphasis on education and awareness raising to shift negative behaviours amongst offending users, alongside guidance and making improvements to platform functionality to help people to minimise exposure to online hate (e.g. creating more private circles or the filtering out of non-verified user content).

Contents

Background and approach

TRAVERSE

The brief

TRAVERSE

To capture insight into how exposure to online hate **impacts** people with different protected characteristics.

To explore a range of **impacts** including physical/behavioural, psychological and social impacts, both online and offline.

To understand how **severe** the impact of online hate can be, compared with other online and offline harms.

To carry out **qualitative research** with at least 30 people from a range of backgrounds.

To be delivered **safely, legally and ethically** with appropriate safeguarding in place for participants and staff.

Key terms

The following working definitions were co-developed with Ofcom to use in the research. These are not the full and legal definitions of these terms and do not directly correspond to the definitions under VSP regulation or how illegal or harmful content will be defined under the Online Safety Bill, or in platforms' terms of service. Simplified definitions were used to support a common understanding of key terms amongst participants.

Protected characteristic: A protected characteristic is a characteristic that someone may have or be perceived to have which means they might be discriminated against. Examples of protected characteristics include gender, disability, race, religion or belief, sexual orientation or whether someone is transgender.

Online hate: hateful content directed at **a group of people** on the basis of a particular protected characteristic.

Hateful abuse: hateful content directed at **an individual** on the basis of a protected characteristic they have or are perceived to have.

Approach

Depth interviews
75mins

1. Semi-structured guide covering:
2. The type of experience they had - what happened, immediate and long-term response
3. Framing the severity of online hate
4. What participants think platforms, regulators and others could or should be doing to address online hate and hateful abuse

Online task
15 mins

1. Independent activity and a pre-requisite to taking part in the online workshop.
2. Captured experiences of online hate
3. Asked participants to rank sets of factors in terms of severity of impact

Online workshop
95 mins

1. Emphasis on setting participant ground rules and highlighting welfare
2. Further exploration of online hate experiences and impacts
3. Discussion of the ranking exercise
4. What further can be done to address online hate

Co-design: data collection tools reviewed by panel with lived experience and trained counsellor

Wellbeing offer: access to professional counsellor sessions and signposting resource for participants; and welfare focused debriefs for research team

All relevant fieldwork materials are available on Ofcom's website

Sample achieved

In total 43 participants were recruited, and **39** interviews were completed. Our screening process ensured that all participants had experienced online hate and or abuse (with the emphasis on the former) and all had been negatively impacted by this (indicating a degree of severity). As shown in the final sample of targeted protected characteristics, intersectionality was a common feature amongst project participants. **Intersectionality** often plays a key part in the online hate and abuse that people experience. Rarely is one protected characteristic singled out and targeted in isolation.

Original quota

Targeted protected characteristic	No of participants
Race-based	Minimum 6
Sexuality-based	Minimum 3
Gender-based	Minimum 3
Religious hate	Minimum 6
Transgender based	Minimum 3
Disability based	Minimum 3
Impacted by hate aimed at a protected characteristic they don't identify with	Total 4

Final sample

Targeted protected characteristic	No of participants
Race-based	20
Sexuality-based	13
Gender-based	13
Religious	14
Transgender based	10
Disability based	17
Impacted by hate aimed at a protected characteristic they don't identify with	4

Recruitment and consent process

TRAVERSE

Worked with specialist market recruitment agency

Mixed method approach which combined drawing on recruiters' database; advertising the opportunity through networks; and promotion via social media.

Completion of screener questionnaire

Captured information about participant demographics, including protected characteristics, and nature of online hate experiences.

Information and consent

Provision of participation information sheet, illegal behaviour policy and consent form, to be signed in advance of interview.

Further verbal consent

Key information explained to participants and further verbal consent sought at start of the interview, right to pause/end interview or withdraw from study explained.

Experiences of online hate and abuse

TRAVERSE

Experiences of online hate

How often had they been exposed to it?

- Participants felt that there was '**lots**' of online hate which was seen across a range of platforms – often on a daily basis and which **flared-up** around key events (e.g. Euro 2020).
- Frequency of experiences was also linked to the extent of their activity online, including the **number of platforms** participants used and how often they posted and engaged with other users.
- Some participants suggested becoming '**blind' or 'desensitised'** to it after seeing it so often and some made efforts to **avoid 'hot spots'.**
- Some platforms were felt to have **more online hate than others**. This was felt to be linked to the effectiveness of different moderation/policies and or the cultures and behaviours that had set in across different platforms.

What types of experiences were had?

- **Reactive** hateful content, including comments, memes, emojis underneath news stories and original content.
- **Original** hateful content – views, observations in the form of tweets, videos etc. directed at groups with particular protected characteristics.
- Views interpreted as hateful that were expressed during **discussion and debate.**
- The use of **slurs/ insults/ strong language** that had a hateful dimension but which might be used in different settings/contexts, e.g. during live sporting events or in gaming settings.

Experiences of hateful abuse

What types of experiences were had?

- **Sharing/posting something** (e.g. a video or photo or a political view) and then being sought out/ targeted for hateful abuse.
- Interactions (e.g. topical discussions or friendly/romantic interactions) which then **became abusive**.
- Pushing back against something seen as hateful or **defending** someone else and then being targeted with hateful abuse.
- Participants noted that hateful abuse sometimes came from users **who shared certain characteristics** with the person they were targeting (e.g. racial abuse from someone of the same race, or homophobic abuse from someone of the same race).
- It was also noted that those perpetrating hateful abuse would sometimes create **new accounts** or would use **multiple platforms** or **different channels** to sustain the abuse and/or increase its intensity.
- Hateful abuse sometimes overlapped with other forms of behaviour such as **bullying and harassment** and exists amongst a **broader set of dangers** (revenge porn, fraud, identity theft) so it is not always the worst thing participants have experienced.

How often had they been exposed to it?

- The frequency of hateful abuse experienced was **mixed** and was strongly linked with **how** and **how often** participants used online platforms. The likelihood of receiving abuse could increase where users:
  - **commented/interacted** lots with other users
  - used multiple **platforms**
  - shared **information** about themselves/their identity
  - spent time/shared in more **open and public** online spaces
- Having had harmful experiences, several participants talked about taking steps to minimise the risk of being targeted again (see section Reactions and coping behaviours). In most instances this has led to a reduction in their exposure.

Impacts of online hate and abuse

Impacts of online hate and online abuse

- Due to the frequency of exposure, the impact of a single incident of **online hate** could be difficult to discern from the cumulative impacts (e.g. feeling desensitised or hopeless).
- Conversely, participants found it easier to recall and discuss incidents of **hateful abuse** because these were less common and stood out more clearly as events which had had more significant impacts (compared with online hate).
- In making sense of different impacts and reactions, a key distinction between participants was between:

Those who felt more duty bound or compelled to **challenge** and may not see themselves as a victim	VS.	Those who sought to **avoid** confrontation and **minimise** exposure to potentially hateful content

"If I could go back, I would just remove the person [from my social media feed] and not have had the interaction – would have stopped them sending harmful words my way. To protect my wellbeing I would have just avoided the confrontation."

"The video creators being abused need defending so I felt it was my duty to report negative comments."

Impacts of online hate and online abuse

The types of emotional and psychological impacts identified by participants are set out below.

Impacts of online hate and online abuse

Surprise and shock

Participants reported feeling surprised or shocked. In the **online hate** context, this could be as a result of seeing a "wall" of hateful comments or emojis underneath a news story or underneath something seemingly innocent.

In the **online abuse** context, they could feel shocked or surprised because they had not expected to receive abuse from others (e.g. receiving abuse having shared a family photo, or sharing a beauty video). In a few instances of **hateful abuse**, as the event unfolded, participants talked about having physiological responses to the situation (e.g. reporting that their heart was pounding).

"Honestly it felt like, like a violation really, innocently watching something, you make a comment, and someone can attack you and say anything they want to you, I felt unsafe. I felt let down."

Impacts of online hate and online abuse

Anger and disappointment

It was common for participants who had experienced both **online hate** and **hateful abuse** to feel anger and disappointment. This was often because their experiences suggested that hateful views were more prevalent than they had realised or because they felt incensed by the hateful behaviour directed at them or at others.

Participants also often expressed anger and frustration where they felt this behaviour had **evaded moderation** (e.g. due to perceptions of "crude" key word based moderation that had missed more subtle hateful content or because it was not deemed as hateful by platforms). For a few participants feelings of anger and disappointment hardened their resolve to challenge online hate, while others chose to avoid exposure.

"The algorithm focuses on specific words, whereas the content is reprehensible, they [the account being hateful] know how to play the game. They are absolutely vile. But nothing happens, and I keep reporting them."

"Initially I actually came off the app, and told myself I'm leaving it and probably shut my account. When I'd calmed down I thought to myself that this is what the haters want – for trans people to become invisible. So I went back on to spite them."

Impacts of online hate and online abuse

Embarrassment and shame

Online abuse sometimes played out in online public spaces where friends, family and strangers could see the exchanges. In some instances this could lead to participants feeling humiliated or shamed by the experience. A few participants described feeling like they were at the centre of a public spectacle, which increased as other users contributed to the 'pile-on' in different ways or attempted to come to their defence.

Experiencing **online abuse** in several cases led to participants questioning elements of their identity or to feeling more negative about who they were. For example, people early in their transitioning journey or coming to terms with their sexual orientation, or where somebody received racist insults that led to them feeling unhappy with their physical appearance.

"In the end I went to 6th form the next day, concerned about how people would react. I broke down to my Latin teacher. I couldn't talk to my parents due to embarrassment".

Impacts of online hate and online abuse

Anxiety and fear

It was quite common for participants to report feelings of anxiety and fear, especially when they had experienced **hateful abuse**. This included both while it was happening and more longer term.

During or closer to the event, participants described feeling anxious or afraid because they:

- had been threatened and feared for their safety;
- felt uncertain about how long the abuse might go on for; because it could intensify;
- feared that abuse could flare up again; or
- because they were uncertain about who was attacking them (where users were unknown or anonymous) and about what personal information those attacking them possessed.

Experiencing **online hate and abuse** often led to fear and anxiety associated with the realisation that hateful views were more prevalent in society than they had realised.

Offline, some participants said their experiences left them feeling less safe in public spaces or worried that they might be mocked, threatened or insulted.

Having experienced hateful abuse online, a few also talked about feeling less trusting towards people they interacted with when offline (e.g. work colleagues), out of a suspicion that they could be harbouring hateful or intolerant views.

"I remember the time it happened. My heart was pounding, I am a grown man. But I felt threatened and scared by how they were behaving. Especially when talking about wanting to hurt people."

"In the longer term, I questioned my ability to judge people's characters and I lost faith in people more generally due to questioning how frequently individuals hold racist views."

Impacts of online hate and online abuse

Hopelessness and exhaustion

Online hate, and to a lesser extent **online abuse**, was seen as pervasive and an integral part of most participants experiences of going online. For some participants, this led to feeling hopeless or exhausted; with participants resigning themselves to the fact that being online and having protected characteristics inevitably meant having to deal with hate and abuse. This also applies to participants who had regularly seen hateful content aimed at protected characteristics which they don't have themselves.

Participants who reported becoming desensitised in the long term often stopped reporting or challenging **online hate.** This was because they saw it as a poor return on their investment of time and emotional energy, since action was not always taken or because the sheer volume of hateful content led to feelings of futility.

"Online hate is something you're endlessly exposed to, whether you challenge it or you don't, you feel hurt and exhausted; it drains your energy."

"In the longer term I have learnt not to engage – I can still write comments, and am true to myself, but if someone instigates a reaction, I do a thumbs up, after 2 or 3 messages, it's just not worth it."

Reactions and coping behaviours

The types of reactions and coping behaviours reported by participants can be clustered into the following categories:

- **Blocking and reporting** – this could drive further anger and frustration when no action was taken
- **Challenging and engaging** – some felt it was their duty to challenge
- **Seeking support** – typically from close friends
- **Self censoring and retreating**, e.g. not sharing/contributing, removing personal information, avoiding spaces

Reactions and coping behaviours

Blocking and reporting

The majority of participants said that at some point they had resorted to blocking and/or reporting accounts behaving in a hateful way, both to prevent themselves from being further impacted by **online hate or abuse** and in the hope that the offending user/s would be temporarily or permanently banned from the platform in question.

However, as noted previously, many participants who took this approach described feeling further anger and disappointment when no action was taken by the platforms, or they did not get their desired outcome.

Participants who had seen hateful content aimed at protected characteristics which they don't have themselves also blocked and reported users behaving in a hateful way. This is both in an effort to protect the other users being attacked, as well as to prevent themselves from being exposed to the hateful content.

"Reading through the tabloid articles and negative comments underneath articles makes me so upset and I comment back but there's 1000s of them. I do report them but there's too many."

Reactions and coping behaviours

Challenging and engaging

It was quite common for participants to report instances where they engaged or confronted those being hateful online. As noted previously, several participants reported that they saw it as their duty to push back against **online hate**. This was especially the case when they had been very negatively impacted by online hate aimed at a set of protected characteristics that they identify as having, or targeting a particular community they belong to. In this scenario, the anger or sense of injustice could drive towards being more confrontational.

For some, the emotional toll of reacting this way led to longer-term negative impacts, including feelings of "burnout". In one example, a transgender woman had to take a few days off social media after challenging hundreds of hateful comments targeting LGBTQ+ users in a single day. She did eventually return back to the platform because she didn't want those posting hateful content to 'win', despite the negative impact on her mental health.

Several participants said that following a negative experience of **hateful abuse**, they no longer tried to engage with people who posted hateful views and instead chose to make comments that they felt neutralised/closed down the situation.

Whilst not common, some participants who received **online abuse** retaliated against the perpetrators either by reporting illegal behaviour to the police or by contacting the employer of the person who had behaved in an abusive way.

"I emailed their workplace and called them to make a complaint explaining the death threats. I then shared screenshots and then he got fired. I saw on their social media that they'd been fired – felt like justice."

Reactions and coping behaviours

Seeking support

Participants tended to seek support from close friends, people within their community and to a lesser extent family members, especially in cases where they had experienced **hateful abuse**.

Very few participants reported seeking support from helplines, or other services.

One participant said that they still sometimes posted content that could provoke hateful comments and reactions. When they were planning on posting, they said they encouraged online friends to post supportive follow-up comments, which they said had the effect of reducing negative and hateful comments and reactions.

"I met with a friend for a meal and I wasn't myself. Then my friend posted a statement online condemning my abuse… It was helpful at the time but I still took a week to get over it, but the statement from my friend did calm the situation down."

Reactions and coping behaviours

Self-censoring and retreating

In order to avoid becoming the target of **online hate** and **online abuse**, participants sometimes reported that they self-censored or retreated from online spaces (e.g. social media sites and discussion forums) in different ways. This included:

- participants who **no longer shared their views** on certain topics, **meaning that their views are potentially not being represented online**;
- **removing personal information** on their profiles in order to make themselves less vulnerable to being a target but **becoming more anonymous as a consequence**;
- **avoiding certain online spaces/platforms** when a news story likely to trigger hate breaks or leaving them all together; therefore **missing out on information and not being able to contribute to public debate and discourse**

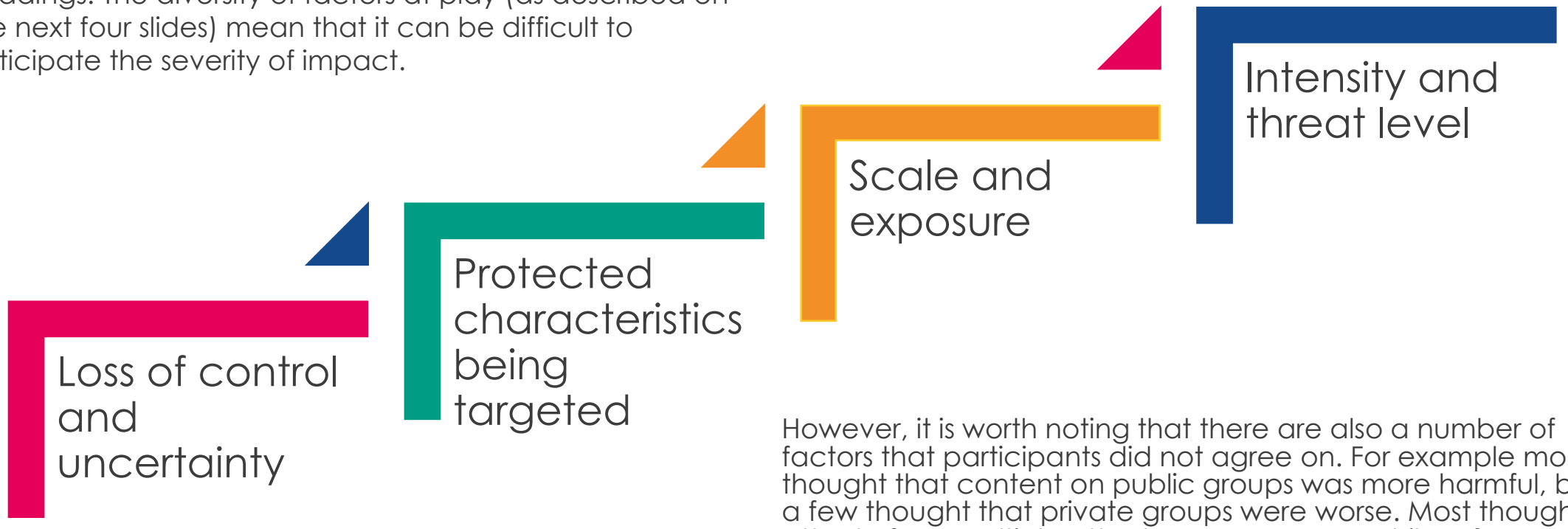
"Offline I probably don't take as many photos, or document my life online in the same way; I've become more guarded as a person."

Factors affecting severity of impact

TRAVERSE

Factors affecting severity of impact

The research identified a wide range of factors that can increase the severity of impact of online hate and hateful abuse which have been clustered under four broad headings. The diversity of factors at play (as described on the next four slides) mean that it can be difficult to anticipate the severity of impact.



However, it is worth noting that there are also a number of factors that participants did not agree on. For example most thought that content on public groups was more harmful, but a few thought that private groups were worse. Most thought attacks from multiple attackers were worse, while a few thought an attack from a single individual was worse. There was no issue on which the participants were unanimous, demonstrating how complex and personal this topic is.

Loss of control and uncertainty

Across both online hate and hateful abuse, seeing hateful content or receiving abuse that was not expected could heighten the sense of surprise or shock – this was reflected in both the online task results and in the interviews.

"It feels more shocking or surprising [where the hate is unexpected]"

"[Where it's unexpected] it contributes to that feeling that you are vulnerable wherever you are online"

Likewise, drawing on the interviews, across both hate and hateful abuse, feelings of anger and frustration could be increased where those being hateful were perceived not to be moderated or could evade moderation (e.g. by creating a new account) and therefore acted with a sense of impunity.

In the hateful abuse context, participants talked about feeling anxious or uncertain because they could not rule out that the hateful abuse might start up again, either by the same users or by new ones – since the content was still online. Some also mentioned not being able to successfully block users since they could keep creating new accounts, which again drove feelings of uncertainty.

Factors that can increase severity:

ONLINE HATE	HATEFUL ABUSE
Online hate is not anticipated, e.g. under a shared family photo	Hateful abuse is not anticipated e.g. abuse after sharing a beauty video
Those being hateful act with a sense of impunity e.g. identity not hidden	Those being hateful act with a sense of impunity e.g. identity not hidden
It is difficult or impossible to remove hateful content	Inability to hide or delete the abusive/humiliating comments
Knowledge that the content won't be removed because it does not break the rules	Uncertainty about whether the abuse may continue/flare up again
	Inability to block/prevent the abuse (e.g. attackers create new accounts, evade moderation)

Protected characteristics being targeted

Where hateful content or abuse targeted multiple characteristics rather than a single one, this could heighten the potential to feel harmed. There was also a sense that online hate targeting multiple characteristics impacts on a greater number of people and was therefore more damaging.

"The hate feels more extensive"

"More people will be targeted/ effected"

A small number of participants also suggested that where online hate or hateful abuse targeted characteristics that cannot be hidden from others (e.g. targeting someone on the basis of their race) this could potentially feel more damaging or hurtful.

"When its about something you can't change or hide it feels worse."

Factors that can increase severity:

ONLINE HATE
Directed at characteristics you can't change/hide
When the hate targets multiple characteristics (i.e. intersectional hate)

HATEFUL ABUSE
Directed at characteristics you can't change/hide
When the abuse targets multiple characteristics

Scale and exposure

A majority of participants agreed that being exposed to hate or receiving hateful abuse from multiple users rather than a single user increased the severity of impact. This was because the experience could feel more intense or overwhelming and because multiple users acting in a way that was hateful indicated that the views were prevalent in society and the content had greater potential to normalise the views and behaviours.

"It's more disturbing to know that many people hold hateful views"

It was also noted that where multiple users posted hateful content, it was more difficult to report and remove, thus heightening feelings of frustration or anger. Where hateful views were shared by users who were well known, who had status or influence or where it was delivered in a humorous way, participants again noted that the impact was likely to be more severe because it again had greater potential to normalise the view and influence others.

"Humour could be more easy to spread"

When it came to hateful abuse, the severity of impact could be raised where growing numbers of users 'piled-on' or where large numbers 'liked' hateful comments. Additionally, where the hateful abuse played out in a online public forum this could increase feelings of embarrassment or shame.

Factors that can increase severity:

ONLINE HATE
Comes from multiple users
Is 'liked' or endorsed by others
Comes from figures with respect/status/influence
Comes from what appears to be 'real' people rather than bots/anonymous accounts
Is posted by someone within your community / known to you / somebody influential
Potential of hateful content to be influential/ widely shared e.g. video/humour, posted publicly

HATEFUL ABUSE
Attacks being 'liked'/ endorsed by others
Pile-ons – with growing number of users joining in
Becomes a public spectacle
It comes from those close to you / within your community

Intensity and threat level

Where hateful abuse was particularly threatening or used violent language, was prolonged, used multiple channels or used particularly spiteful or personalised attacks this could heighten the severity of impact.

"He started being provocative and at the time I wasn't confident so ignored it. Eventually he found out things about me, and because I ignored it the aggression became more – he found out my name and assumed my heritage and became racist"

In both the online hate and hateful abuse contexts, participant's experiences also demonstrated that the severity of impact could be increased depending on the participants' level of vulnerability or emotional state. Here, participants could feel more harm where they:

- had low self-esteem or a negative body image
- low resilience or poor mental health
- were experiencing difficult life events at the time when they were targeted or exposed to online hate (e.g. those at an earlier stage in coming to terms with their trans identity or sexual orientation)

Factors that can increase severity:

ONLINE HATE
The intensity of the hateful content, including threats of violence
Content is solely motivated by desire to offend/ attack
The emotional state/ level of vulnerability of those being targeted

HATEFUL ABUSE
When those abusing have knowledge of/ reference their target (e.g. name, appearance, characteristics, biographical details)
Repeated and prolonged attacks (harassment/stalking dynamic)
Combines mockery/ humiliation
Doxing[1] attempts and/or physical threats
Attacks via. multiple channels, including direct messages
The emotional state/ level of vulnerability of those being attacked

1 Doxing is the act of publicly revealing previously private personal information about an individual, usually via the internet.

Participants' suggestions for addressing online hate

TRAVERSE

Online hate and freedom of speech[1]

- Participants were asked about whose responsibility they think it is to address hateful content online, and whether they think there any risks with limiting what people say online. Most participants felt that freedom of speech was **important to consider** in the context of addressing online hate.
- However, few felt that there should be **no limits** on what people can say – with the freedom to harm and threaten others often seen as the 'red line'.
- A common position was to remain committed to protecting freedom of speech but avoid a situation where there was "**freedom from all consequences**" – where hateful and abusive views can be expressed with impunity.
- A few felt that a consequence of being 'hateful' should include being "**exposed**" for behaving in this way.
- It was felt that in the current situation, many people with protected characteristics are having to **self censor**, in order to avoid being targeted and abused.
- Several acknowledged that moderation/decision making had "grey areas" and sometimes involved making **tricky judgement calls**.
- A few participants had the view that speech alone (including speech which could be interpreted as hateful) should not be moderated, so long as it **doesn't involve threats** of violence or harm.

"I don't agree with an overzealous censoring culture, it should only happen where content can cause violence and material harm – I think there should be 'freedom of speech but not freedom of consequence'."

"As a result of my experience I'm now reluctant to post anything on social media that relates to my faith. Although I will make a point of liking other peoples' posts about Judaism; but I wont post a comment myself."

1 We use the phrase freedom of speech in this report (rather than freedom of expression), as this was the phrase used by research participants.

Online hate and freedom of speech

Several **risks** associated with restricting speech and online behaviour were identified by participants:

- Overzealous moderators could stifle debate and public discourse
- Political activists may reasonably want to protect their anonymity
- Some users may want to protect their anonymity due to sensitivities related to their immigration status
- Stricter approaches could drive views 'underground' where they could intensify
- Policy may become dictated by a minority of users who are most sensitive

"There are definitely risks in terms of limiting what people can say; people can feel marginalised if they can't share their views. That can fester and bubble and grow – if they can't be hateful here they'll find somewhere else if they feel it enough."

"There's risks of limiting what people say – an example is [a video sharing platform] removing the counter on the number of 'dislikes' for a video –this can stop people protesting content they find harmful or offensive."

"I don't agree with proposals to require proof of ID before creating an account as this will exclude marginalised communities who may not have a passport or other ready means of proving ID and anyway, trolls can always create fake IDs."

Platforms – what participants think is working well and less well

What's working well?

- A few participants felt that the following was working well:
  - Key word **bans** are positive (although it was noted that this is only part of the solution to addressing the problem).
  - Where hateful content is **demonetised** - so that content creators can not receive ad revenues.
  - Where platforms have **filtering functions** (e.g. you only see content from verified users or from a trusted 'circle').
  - Having the ability to **block** users.
  - **User-led moderation** in online forums which can mean that hateful content is often swiftly removed.
  - Platforms who are fairly **responsive** in taking action, especially for content more likely to be harmful.

What's working less well?

- Some participants reported that content is often not being reviewed/removed **quickly enough or at all,** and that it is still too easy to **evade moderation** (with perceptions that platforms are often relying on simplistic algorithmic approaches, e.g. by writing in Punjabi using English characters, avoiding the use of key words, use of "dog whistles"[1]).
- A few participants noted that:
  - People can create **new accounts** once they've been banned/deactivated.
  - The ability to post **anonymously** encourages hateful behaviour.
  - Existing rules and guidelines are adequate but are **not being properly followed**/enforced by platforms.
  - The ability to delete and edit posts and retrieve messages can **embolden** those posting hateful content.

1 Internet dog whistles can be defined as a coded message communicated through words, images or phrases understood by a particular group of people.

Addressing online hate – role of platforms

Participants tended to say that platforms had the primary responsibility for addressing online hateful content:

Many called for more **active moderation**, including more proactive searching/removing of hateful content, more nuanced and human-led moderation, and taking faster and more consistent action

Several called for tackling the challenge of users with **multiple accounts and repeat offenders**

A few called for greater **external oversight** of platforms

Many called for consideration of the **capacity, skills and diversity** of those moderating and the quality and comprehensiveness of their guidance

Several placed an emphasis on creating settings that give users greater **control** to filter out hate, alongside community standards and guidelines and specialist support

Addressing online hate – participants' views on the role of users, regulators and government

Users

- Some felt users have a duty to report online hate while others emphasised that the onus should not be on users - due to harm caused by prolonged exposure.
- Participants felt that users tempted to post potentially hateful content, need to become more aware of the harm that result from behaving in this way.
- When it came to online abuse, it was felt that users need to act responsibly to minimise harms (e.g. utilise privacy settings, block users, utilise the 'delete comments' function, avoid 'leaning in' to debates).

Regulators

- Regulators should ensure that platforms are following rules and are taking robust action to enforce their own policies or removing any illegal content.
- Review and highlight best practice by platforms in tackling online hate.
- Consider introducing new fines/sanctions for platforms.
- Invest in training and resources to support moderation.
- Track the prevalence of online hate.
- Ensure that the public know what regulators' roles are in this policy area.

Government

- Set up an independent panel to review cases, to benefit users who don't believe the response is adequate.
- Enlist influencers and work with educators, targeting those who are ignorant/ insensitive and teaching young people about digital welfare/ethics.
- Review hate speech legislation – is it tough enough/effective?
- Consider the challenge of users who run multiple accounts.
- Form a 'cyber' police force.

Participant case studies

Case study participants (slides 44 - 48) have been given fictional names and details about their experiences have been edited to protect their anonymity.

TRAVERSE

Online hate case study

Michelle - a black woman in her fifties with an autistic daughter

- She describes various experiences of online hate targeting people on the basis of their race and ethnicity and people with disabilities.
- One event she spoke about specifically was a news article on a social media platform about the Euro 2020 final that included various derogatory and racially charged hateful comments.

"Reading through posts and articles and negative comments underneath articles makes me so upset and I comment back but there's 1000s of them. I do report them but there's too many."

- She explained that she isn't particularly impacted by hateful content as she is 'used to it'. She also said that, whilst she doesn't fear for herself, she fears for the life her daughter will live having to potentially deal with discrimination in person.
- She pushes back on hateful comments when she can but still feels as though people should have the right to say whatever they want online.
- She believes that users should educate themselves about the impacts of their words and comments and that the media should consider whether their articles will generate hateful comments.

"Seeing racist or disablist comments so frequently makes me scared and concerned about my kids and how people react to them being mixed race or having a disability. I used to cry when I saw hateful comments about [a certain celebrity] and her child because I can relate."

Online abuse case study

Connie, a trans, non-binary person married to a Muslim Iranian woman

- The participant and their wife have experienced abuse online on multiple occasions, primarily after posting a picture of their wedding. Their social media account was hacked and the hacker started posting transphobic content.
- This was often homophobic abuse but they also received hateful comments that suggested that Muslim people cannot be homosexual. These comments led to threats of violence.
- These incidents also escalated into people misconstruing their words from posts and suggesting that they were responsible for child abuse.
- In response, they display a 'rules of engagement' message on their account profile and asked friends to make supportive comments whenever they post to deter hateful abuse.
- Connie now thinks very carefully about when and how they post content that may trigger abuse.
- The participant noted that they won't post immediately after someone more influential than them (e.g. official government accounts or celebrities) are in the news discussing some LGBTQ+ issues. This helps to reduce unwanted replies and interactions.
- They do not agree with the suggestion that platform users should have to upload ID when signing up for a platform, however they do not think platforms are currently doing enough to tackle online abuse.
- It was suggested that more investment is needed to effectively deal with hateful abuse online.

"I've pinned some rules of engagement - if you are abusive I wont engage with you I will just block, delete and report you. It has become too time consuming to engage with people."

Online hate case study

Kedija, woman in her 20s who was born in East Africa

- She described regularly seeing online hate from 'right wing trolls' focused on immigration, with much of the content containing racist and sexist language. She has pushed back at these comments in the past but this led to hate directed at her personally.
- This led to her feeling angry and upset about the content she was seeing, making her question how people view her in a real-life context as she feels indirectly attacked by the online hate.
- She has shared her experiences with friends and family but has resorted to only posting in online spaces that she deems to be 'safer'.
- It has left her less vocal in general, including in the real world – she used to always express her opinions on topics but now only does with people she feels safe around. This reminds her of the self censorship people practiced in her country of origin, which she thought she would have left behind when moving to the UK as she thought it was a more liberal country.

"I think the biggest thing, that I took away is that I need to not share too much about myself. My opinions, what's important to me, what my beliefs are."

- Whilst she doesn't like the idea of censorship, she highlighted that more needs to be done in order to tackle online hate. This includes more robust legislation and more input from the media and the government.

Online abuse case study

Aanika, a female south-east Asian university student who identifies as lesbian

- She sent a tweet about the impact of colonialism in Asia and faced an 'unrelenting barrage' of hateful homophobic abuse from Tamil accounts (who saw information about her sexual orientation on her social media profile).
- She initially felt upset, as though her identity was being invalidated, then became scared and intimidated by threats of violence.

"It was generally saying stuff [which is] invalidating, your experiences in your own faith, and your queer identities are wrong. Everything you are saying is wrong."

- She blocked users who were making abusive comments, however they made new accounts and continued to be abusive towards her.
- She spoke with friends about her experience and updated her profile picture to her wearing a mask in order to conceal her identity.
- Her major frustration was platforms failing to suspend or delete accounts of repeat offenders. She suggested that, moving forward, platforms need to be able to detect more subtle forms of hate, whilst schools need to roll out more comprehensive learning around online safety.

Online hate case study

Kit, is a white transgender man in their twenties who is neurodivergent

- They consider themselves an activist which means they are regularly exposed to hateful content online.
- They described one incident in which they observed hate directed at the trans community, which they felt were exacerbated by posts from certain popular accounts.

"Following these instances it emboldens people to speak out about these issues. I reached a point with [a well known user] where I blocked and muted her because I would see people online who follow her and quoting her messages, pushing back against what she was saying but it's still exposure to that hateful content."

- The impacts they face are dependent on their current mental state, but they are often left feeling hopeless and sometimes self-hating.
- Viewing this hate online makes them more concerned about people's views in the real world and the negative or hateful feelings people might be harbouring.
- Due to their experiences they have now closed their social media account and regularly speaks to friends and family offline if they have had negative experiences.
- They don't believe platforms are currently doing enough to tackle online hate. They explained that the functionality to search for topics on social media accounts makes hate more accessible as other users will comment on their posts that aren't following them.

"Even when I share someone else's post speaking out against [a well known user] there'll be people who see I retweeted and then start replying to me. Even if I share something I get tagged in it and dragged into the hateful conversation."

Participant feedback on taking part in the research

Reflections on the approach

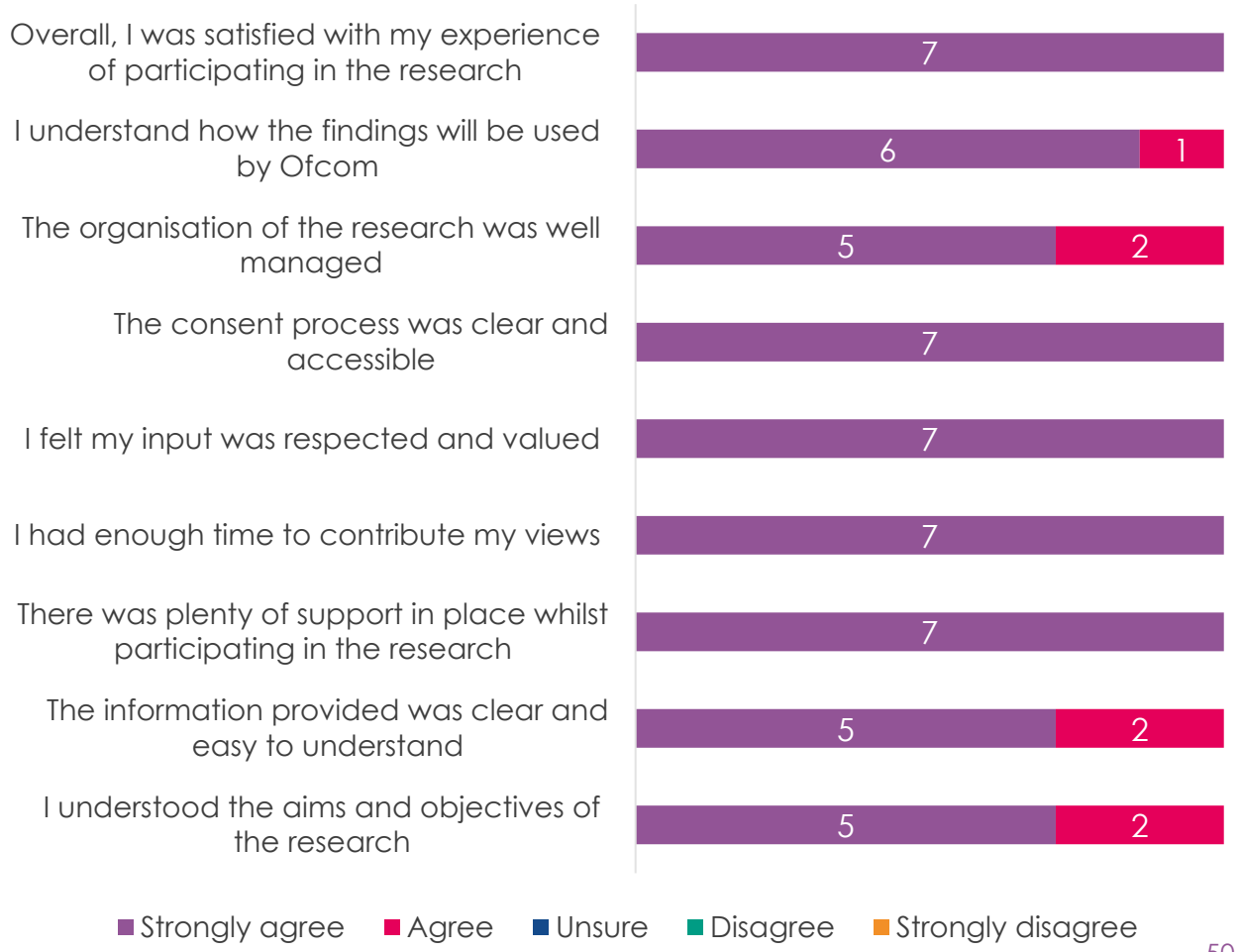
In light of the sensitive nature of the topic and the potential harms to participants' wellbeing was considered at each stage of the research.

As noted on Slide 8, steps taken included, offering sessions with a trained counsellor, sharing a summary of support organisations, an emphasis on rapport building with participants at the start of interview, avoiding sharing 'raw' examples of online hate, and setting clear ground rules for participation in the online workshop.

Participants were positive about the experience of taking part in the research. Towards the end of each interview, several mentioned that it had been helpful to have the time and space to reflect on their experiences. No participants took up the counselling offer, although a few noted that knowing it was available reassured them that the process was a supportive one.

Feedback collected from participants who participated in both an interview and the online discussion workshop indicates that all felt satisfied with the experience of taking part and well supported.

Online survey - participant feedback (base: 7)



Thank you

TRAVERSE

www.traverse.ltd