Increasing engagement with Terms and Conditions

Technical report prepared by Behavioural Insights Team for Ofcom

30th July 2024

THE BEHAVIOURAL INSIGHTS TEAM

Contents

1. Background

1.1 Policy context

Ofcom has a duty to promote and research media literacy, which it defines as "the ability to use, understand and create media and communications in a variety of contexts". This includes user ability to understand social media platforms' Terms and Conditions (T&Cs) and give informed consent. Ofcom is also the regulator for video-sharing platforms (VSPs) and since November 2020, VSPs established in the UK must comply with measures designed to protect users. A number of these measures relate directly to T&Cs. For example, VSPs must include T&Cs that prohibit a user from uploading a video containing relevant harmful material.

Additionally, this research will build evidence for Ofcom's new duties under the Online Safety Act 2023 (OSA). For example, under the OSA, in-scope service providers must ensure that certain provisions within their terms of service[1] are ***'clear'*** and ***'accessible'***.[2] Having clear and accessible terms of service is important as they are typically how users can learn what is and is not allowed on a platform, what safety features and tools are available, and how to navigate and use the service to keep themselves and others safe.

However, despite their importance, users often do not engage with T&Cs. Recent Ofcom research suggests only 8% of social media and VSP users said they read T&Cs thoroughly before agreeing to them.[3] Therefore, Ofcom is looking to gather evidence about effective methods to encourage users to engage with T&Cs, and to read, understand and follow the service's rules.

Finally, this work adds to the research Ofcom's Behavioural Insights Hub is carrying out to explore if and how platform design changes can be used to reduce online harms, such as the previous online trials on content reporting.[4]

1.2 Research objectives

Together with Ofcom's Behavioural Insights (BI) Hub, the Behavioural Insights Team (BIT) conducted an online randomised control trial (RCT) to test different interventions that encourage users to engage with Community Guidelines, a category of T&Cs often found on social media platforms. More specifically, we tested different ways of encouraging users to click through to read the Community Guidelines at sign-up and mid-feed (i.e., when users are

[1] For the purposes of this report, T&Cs refer to all documents that set out the rules for using a platform. The OSA refers to "terms of service" which represent the same type of documents.
[2] UK Parliament, 2023. Online Safety Act 2023.
[3] Ofcom, 2024. Terms and conditions and content controls.
[4] Ofcom, 2023. Behavioural insight for online safety: understanding the impact of video sharing platform (VSP) design on user behaviour.

viewing a social media feed). We tested interventions with different framing and different timing.

1.2.1 Research questions

The trial aimed to answer the following research questions:

- **RQ1**: Does the framing of Community Guidelines [positive framing or relabelling them in a user-friendly way] increase participants' engagement?
- **RQ2:** Are participants more likely to engage with Community Guidelines when prompted while using an online platform?

The primary outcome measure was whether users click through to access Community Guidelines.

We also examined whether our interventions decreased reposting and/or increased reporting of content which violated the Community Guidelines. Furthermore, we included several exploratory outcomes measures, such as attitude towards and perception of the Community Guidelines, to help us to better understand the psychological mechanisms behind our primary and secondary analyses and generate hypotheses for future research (see section 3.5 for the full analytical framework).

2. Intervention development and hypotheses

2.1 Intervention development

2.1.1 Summary of research activities

Ofcom undertook several research activities to develop ideas for interventions. These interventions aimed to address key behavioural barriers and drivers to increase user engagement with T&Cs.

Initial research activities included:

1. A survey of consumer's views and use of T&Cs across social media and VSPs.[5]
2. Behavioural diagnosis using the 'COM-B' model to identify barriers and enablers to accessing T&Cs and checking rules on platforms.[6]
3. Workshopping intervention ideas and prioritisation, including identifying and shortlisting intervention options.

Further details on the development of the interventions are provided in Annex A.

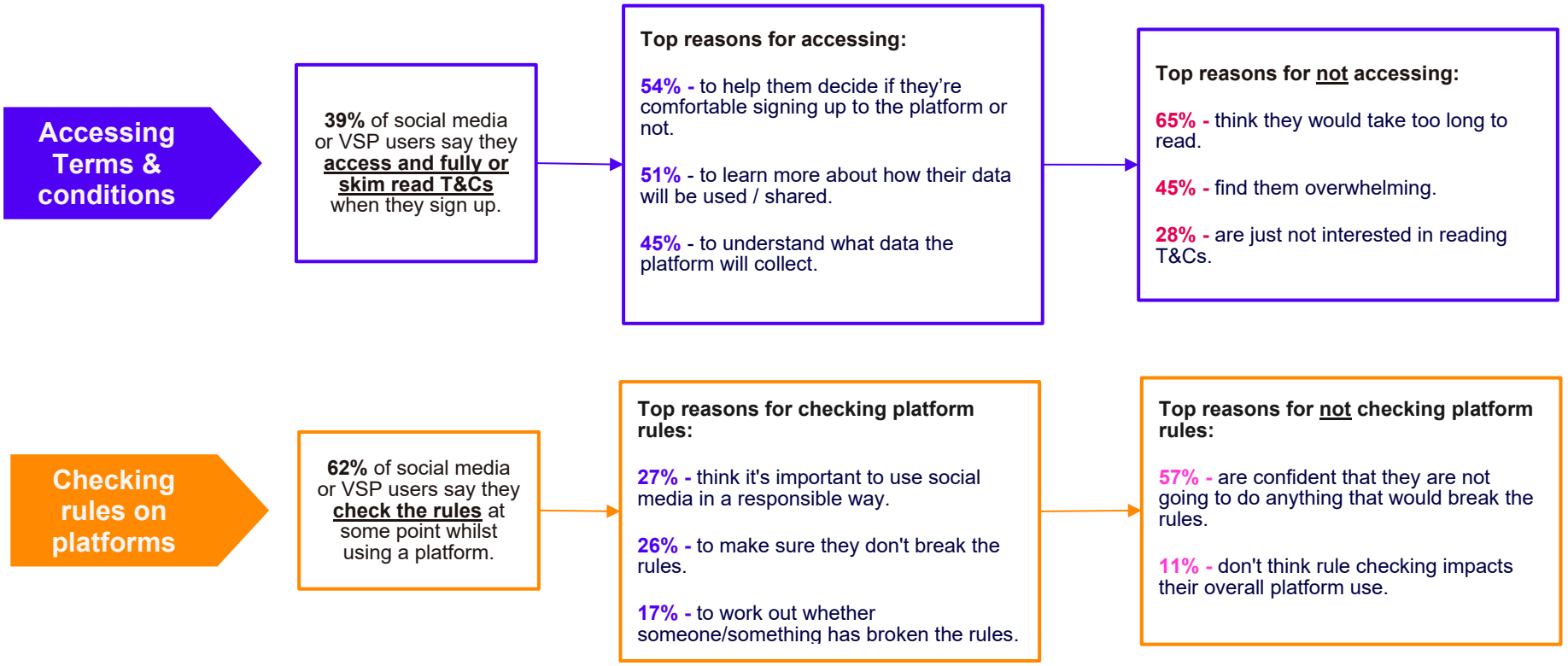
2.1.2 Survey findings

Ofcom surveyed consumers (n=2,149) to understand their views and use of T&Cs across social media platforms and VSPs. The survey focused on two key behaviours: accessing T&Cs while signing up and actively checking platform rules. Figure 1 provides an overview of the headline findings for each behaviour.

[5] Ofcom, 2024. Terms and conditions and content controls.

[6] Michie, S., van Stralen, M. M., & West, R. 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions.

Figure 1: A summary of key findings from the survey



2.1.3 Behavioural diagnosis

Ofcom's BI Hub undertook a behavioural diagnosis exercise (detailed in the Annex) informed by survey findings to identify capability, opportunity, and motivational factors relevant to users accessing T&Cs and checking rules on platforms.

2.1.4 Intervention design

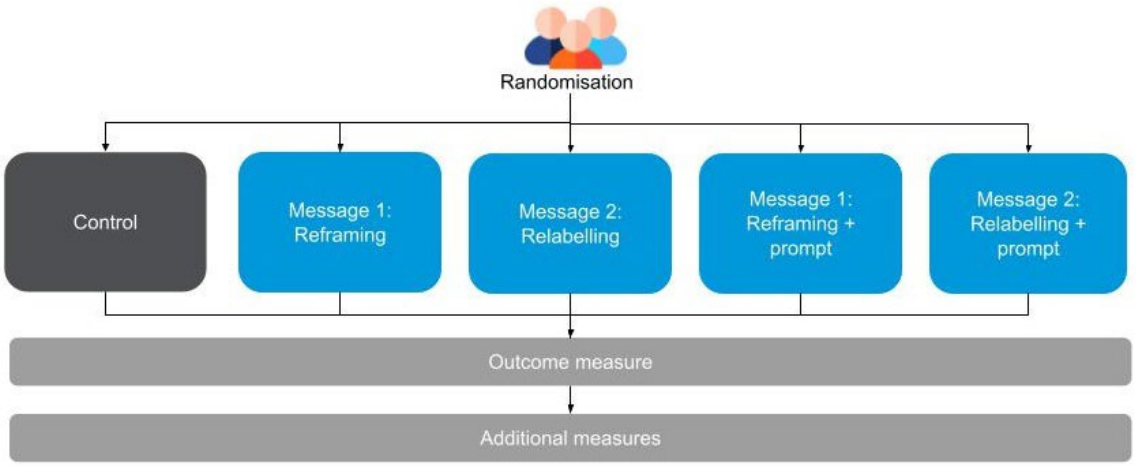
Based on the behavioural diagnosis exercise, BIT and Ofcom's BI Hub developed the interventions to be tested. The intervention design process was guided by three main strands of thought:

1. Relabelling the Community Guidelines in a user-friendly way could help overcome pre-existing biases and make the document's purpose more salient. This would increase users' capability and motivation to access them.
2. Reframing the purpose of the Community Guidelines could highlight their importance and increase users' motivation to access the documents.
3. Reminding users of the platform rules and how to access them while they are using the platform could increase the number of people accessing them.

2.2 Interventions and hypotheses

Based on the intervention design, two types of messages were tested (Reframing and Relabelling) at two time points (one with a message only at sign-up, and one with an additional prompt in the middle of the feed) in a 2x2 design with four treatment conditions. Figure 2 gives an overview of the Control arm and the four treatment arms into which participants were randomised.

Figure 2. Overview of trial arms.



2.2.1. Control arm

When designing the Control arm of the trial, we aimed to replicate what users encounter on social media platforms: a simple message and the ability to access the Community Guidelines through a link at sign-up (see Figure 3) or through the gear button (settings icon) in the main feed (see Figure 4).

Figure 3: Accessing Community Guidelines at sign up.

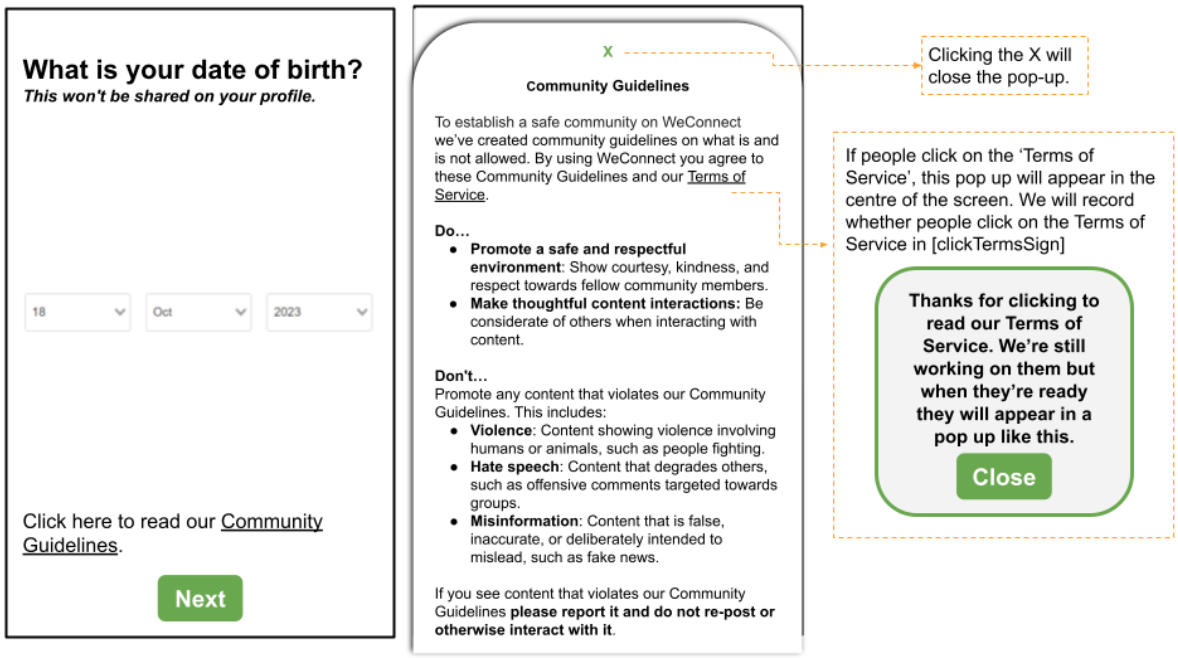
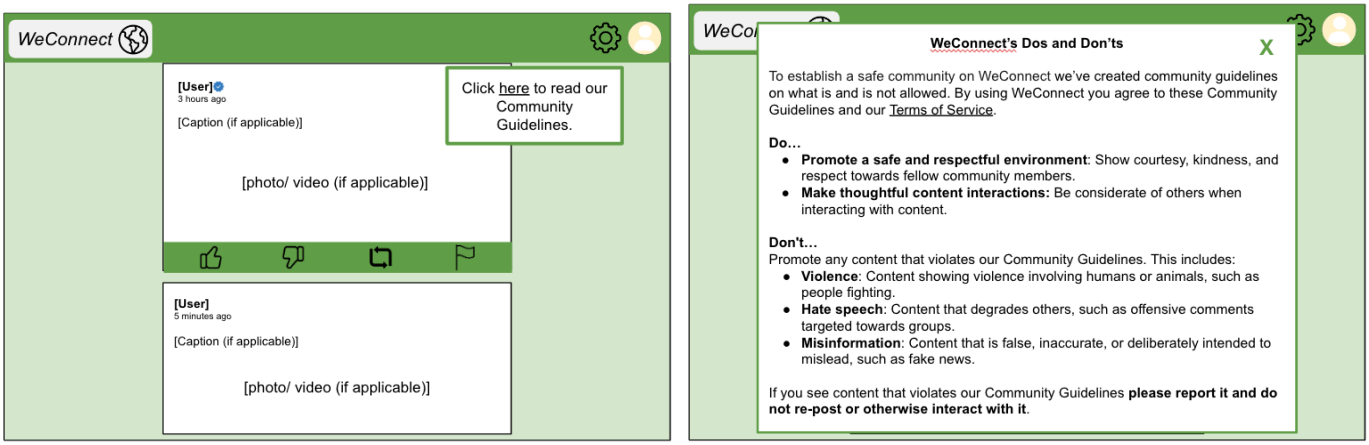


Figure 4: Accessing Community Guidelines in the main feed through the gear button.



2.2.2. Relabelling

In the relabelling arm, the Community Guidelines were relabelled as “Dos and Don’ts” to motivate users to click through to access them. The hypotheses for these arms were:

Primary outcome measure
H1a: Exposure to the relabelled title of the Community Guidelines **increases** the probability of participants clicking to view them compared to those in the Control group.

Secondary outcome measures
H1b: Exposure to the relabelled title of the Community Guidelines **increases (decreases)** the probability of participants **reporting (reposting)** violative content compared to those in the Control group.

2.2.3 Reframing

In the Reframing arm, the Community Guidelines were signposted with text explaining their importance and attempting to motivate users to click through to access them. The hypotheses for these arms were:

Primary outcome measure
H2a: Exposure to positive reframing **increases** the probability of participants clicking to view the Community Guidelines compared to those in the Control group.

Secondary outcome measures
H2b: Exposure to positive reframing **increases (decreases)** the probability of participants **reporting** (**reposting)** violative content compared to those in the Control group.

2.2.4. Prompting

As outlined above, we also tested whether prompting users after they had viewed some content would enhance the message’s impact, increasing engagement with the Community Guidelines (see Figure 5 for the message at sign-up and Figure 6 for prompts).

Figure 5: Reframing message at sign-up on the left, relabelling message at sign-up on the right.

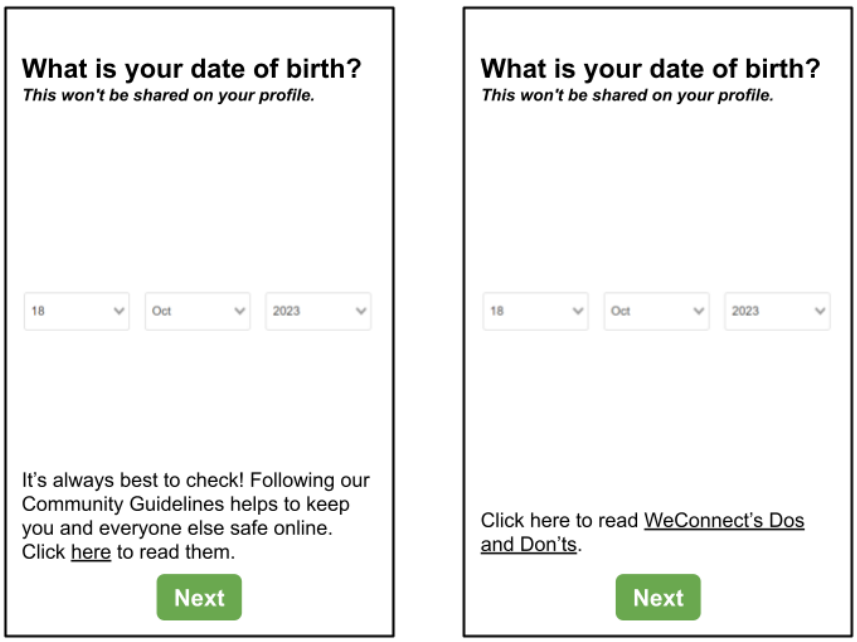


Figure 6: Reframing prompt on the left, relabelling prompt on the right.

2.2.4.1. Prompting and Reframing

The hypotheses for these arms were:

Primary outcome measure
H3a: Exposure to positive reframing alongside a mid-feed prompt **increases** the probability of participants clicking to view the Community Guidelines compared to those in the Control group.

H3b: Exposure to positive reframing alongside a mid-feed prompt **increases** the probability of participants clicking to view the Community Guidelines more than when they are only exposed to positive reframing.

Secondary outcome measures
H3c: Exposure to positive reframing alongside a mid-feed prompt **increases** (**decreases**) the probability of participants **reporting** (**reposting**) violative content compared to those in the Control group.

H3d: Exposure to positive reframing alongside a mid-feed prompt **increases** (**decreases**) the probability of participants **reporting (reposting)** violative content more than when they are only exposed to positive reframing.

2.2.4.2. Prompting and Relabelling

The hypotheses for these arms were:

Primary outcome measure
H4a: Exposure to the relabelled title of the Community Guidelines alongside a mid-feed prompt **increases** the probability of participants clicking to view the guidelines compared to those in the Control group.

H4b: Exposure to the relabelled title of the Community Guidelines alongside a mid-feed prompt **increases** the probability of participants clicking to view the guidelines more than when they are only exposed to the relabelled title.

Secondary outcome measures
H4c: Exposure to the relabelled title of the Community Guidelines alongside a mid-feed prompt **increases (decreases)** the probability of participants **reporting (reposting)** violative content compared to those in the Control group.

H4d: Exposure to the relabelled title of the Community Guidelines alongside a mid-feed prompt **increases (decreases)** the probability of participants **reporting** (**reposting**) violative content more than when they are only exposed to the relabelled title.

2.2.5. Treatment arms

The framing and timing of the message varied across treatment arms. See Table 1 for an overview of these arms.

Table 1: Overview of message content and delivery.

	Reframing message	Relabelling message
Message at sign-up	**Reframing**	**Relabelling**
	Arm 1: Positive reframing (only presented at sign-up) Message appeared at the age screen during sign up: *"It's always best to check! Following our Community Guidelines helps to keep you and everyone else safe online. Click here to read them."*	**Arm 2: Relabelling (only presented at sign-up)** Message appeared at the age screen during sign up: *"Click here to read WeConnect's Dos and Don'ts."*
Message at sign-up and mid-feed prompt	**Reframing + prompt**	**Relabelling + prompt**
	Arm 3: Positive reframing (presented at sign-up and mid-feed) Prompt appeared after participants had viewed a quarter of the feed with the following message embedded: *"Hey! We hope you like WeConnect! Since you're new here we just wanted to draw your attention to our Community Guidelines.* *It's always best to check! Following our Community Guidelines helps to keep you and everyone else safe online. Click here to read them."*	**Arm 4: Relabelling (presented at sign-up and mid-feed)** Prompt appeared after participants had viewed a quarter of the feed with the following message: *"Hey! We hope you like WeConnect!* *Since you're new here we just wanted to draw your attention to WeConnect's Dos and Don'ts."*

3. Methodology

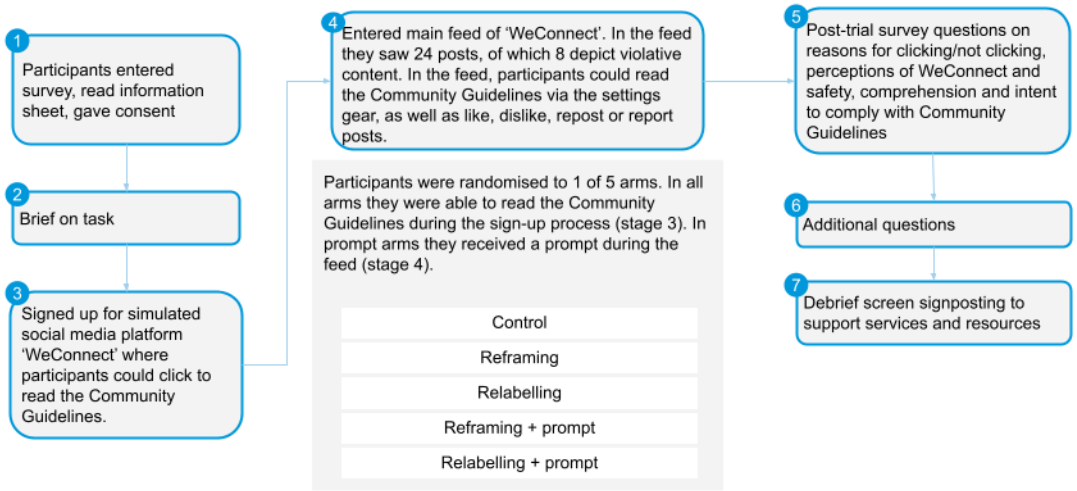
3.1 Trial design

To answer our research questions, we designed a simulated social media platform that mimicked real platforms. The simulated environment was embedded into an experimental survey with an RCT design. In an RCT, research participants are randomly divided into different groups and exposed to either an intervention or a control. Due to the random assignment into experimental arms, intergroup differences in outcome measures can be causally attributed to the interventions participants were exposed to. Our trial design allowed us to measure the causal impact the different interventions have on participants' behaviours and attitudes.

3.2 Simulated social media platform

Figure 7 illustrates the flow of the experiment.

Figure 7. Participant journey.



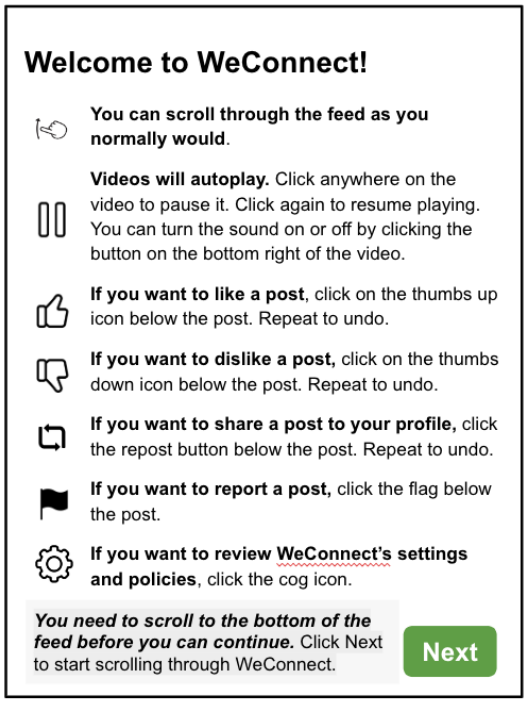
3.2.1 Platform design and functionality

We designed our platform, WeConnect, with the intention of facilitating a trial environment that mimics real experiences on social media as much as possible to increase the external validity of our findings. External validity refers to the extent to which the findings of a study can be generalised to, and are representative of, real-world populations, settings, and conditions beyond the specific context of the research. While WeConnect is not based on a single real-world platform, its design is inspired by popular platforms. By making participants' experiences on WeConnect as realistic as possible, we aimed to generate findings that indicate how our interventions would impact users' behaviours on actual platforms.

In a previous trial by Ofcom and BIT using a similar WeConnect platform, over 60% of participants said WeConnect felt similar or very similar to platforms they used before and 90% found WeConnect easy or very easy to use.[7]

The platform had two main components: (1) a sign-up process and (2) a content feed. During sign-up, participants went through a typical process where they were asked to allow push notifications, gave their date of birth, and were introduced to the platform functionalities (see Figure 8).

Figure 8. Example screen from the sign-up process.



After the sign-up process, participants entered the content feed on WeConnect. Figure 9 illustrates what the feed looked like. Participants had to scroll to the bottom of the feed before they could progress to the next stage of the experiment. Participants could engage with posts in the feed by liking, disliking, reposting, and reporting them. If participants reported content, they were asked to select why they reported the content (see Figure 10), and the reported content was then blurred. They could also click on the gear button to review the Community Guidelines. After participants scrolled through the feed and clicked 'Next' at the bottom, they progressed to a follow up survey (see section 3.2.4 for more detail).

[7] Ofcom, 2024. Testing content controls to tackle online harms.

Figure 9. WeConnect content feed after a participant clicks the gear icon (left figure is for the Control arm; middle figure is for the Reframing and Reframing + prompt arms, right figure is for the Relabelling and Relabelling + prompt arms).

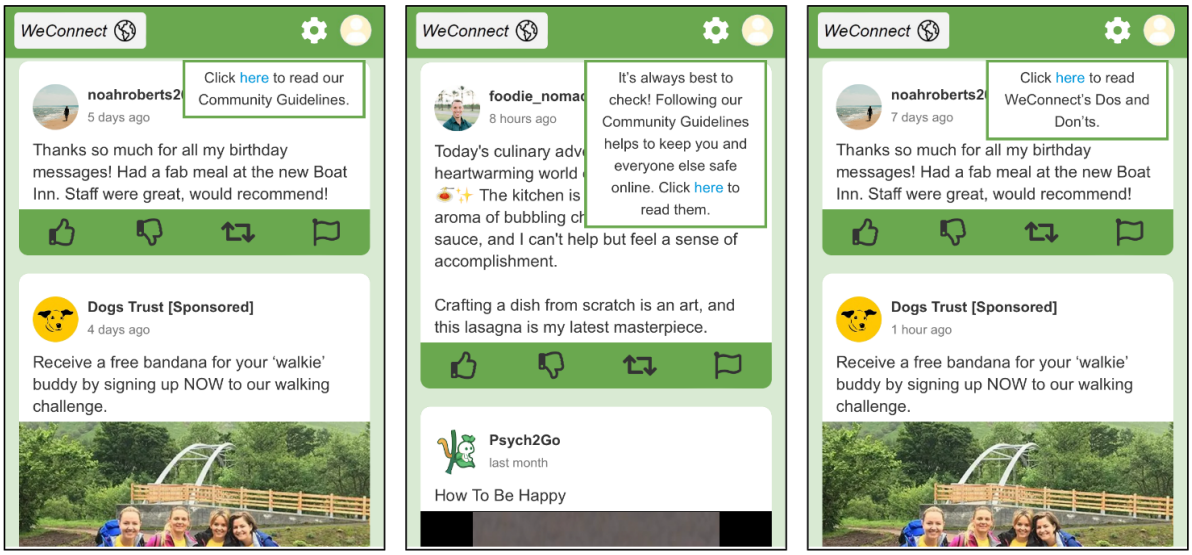


Figure 10. Report follow up question.

3.2.3 Stimuli

The content consisted of 6 short videos, 6 long videos and 12 short text posts. Most of the text posts were accompanied by images related to the content of the post. The amount of content was informed by previous social media trials BIT ran and aimed to keep participants engaged in the feed for 5 minutes.

Of the 24 pieces of content that participants saw in the main feed, 8 pieces (33%) violated WeConnect's Community Guidelines ('violative content'). The violative content categories included in the trial were hate, violence, and misinformation. The non-violative posts were made up of benign content that resembled the type of content users encounter on real social media platforms. The content was presented on the feed in a random order, apart from a few

restrictions. Participants could not see more than three pieces of violative content in a row, and their feed's first and last posts were always benign.

3.2.4 Post-feed survey

After interacting with the feed, participants completed a post-feed survey, which included questions on recall, their reasons for clicking or not clicking on the Community Guidelines, their perceptions of platform safety, their attitude towards the prompt (Reframing + prompt arm and Relabelling + prompt arm), and their comprehension of the Community Guidelines. Participants were also asked to provide additional demographic information including social grade, device typically used to access the internet, and social media platform use.

3.2.5 User testing

To ensure that our platform, the content, and the survey were understandable, easy to use and perceived as realistic, we conducted five user testing sessions with BIT employees not involved in the project. During these user tests, a BIT qualitative researcher worked closely with participants and had them think aloud (verbalise their thought processes) as they interacted with the experiment. Think-aloud protocols are a technique commonly used in product design.[8] Participants voiced their thoughts as they went through the platform and experiment, giving us insight into their comprehension and areas of confusion. The researcher who led these sessions used a facilitation guide that included observation prompts on crucial aspects of the experimental design (e.g., does the user understand what Community Guidelines are?).

Based on the researcher's observations, feedback on the platform, and content voiced by participants, BIT iteratively changed and updated the design of the platform, interventions, and survey questions.

3.3 Sampling and data collection

3.3.1 Sample criteria

We recruited a nationally representative sample of adults in the UK. Participants were required to:

- be aged 18 years or older,
- live in the UK,
- not taken part in the previous Making Sense of Media (MSOM) trial run by Ofcom and BIT on a similar platform[9].

[8] Ericsson K. A., Simon H. A.1993. Protocol Analysis: Verbal Reports as Data.

[9] We initially only invited people to take part in the experiment if they had not taken part in the previous two trials run by Ofcom and BIT (Ofcom User Controls trial and Ofcom Making Sense of Media (MSOM) Establish trial). However, during data collection we updated these criteria to facilitate recruitment and invited people who had taken part in the Ofcom User Controls trial. We controlled for previous participation in the analysis.

3.3.2 Power calculations

The sample size was based on power calculations for our primary outcome (whether participants clicked on the Community Guidelines; see section 3.5.1). In the absence of published online experiments looking at comparable outcomes, we conducted calculations for baseline proportions ranging from 5%-15% (see Table 2), assuming 80% statistical power and a significance level of α = 0.05 (correcting for 6 comparisons in primary analyses: 5% / 6 = 0.83%). A sample size of 3,500 participants (700 participants per arm) would allow us to detect a minimum detectable effect size of 4.81 - 7.21pp (percentage point difference) between arms where 5% - 15% of participants in the baseline arm clicked on the Community Guidelines respectively. We deemed this sufficient for an online experiment.

Table 2. Power calculations for a sample of 3,500 participants (700 per arm) assuming 80% statistical power and a significance level of α = 5% (0.83% after correcting for multiple hypotheses testing)

Outcome baseline	Minimum detectable effect size (% point difference)
5%	4.81pp
10%	6.24pp
15%	7.21pp

3.3.3 Data collection

All participants were recruited through the panel aggregator Lucid, with payments being administered by the panel providers they were registered with. Participants were only invited to take part in the experiment by Lucid if they met our sampling criteria (see section 3.3.1).

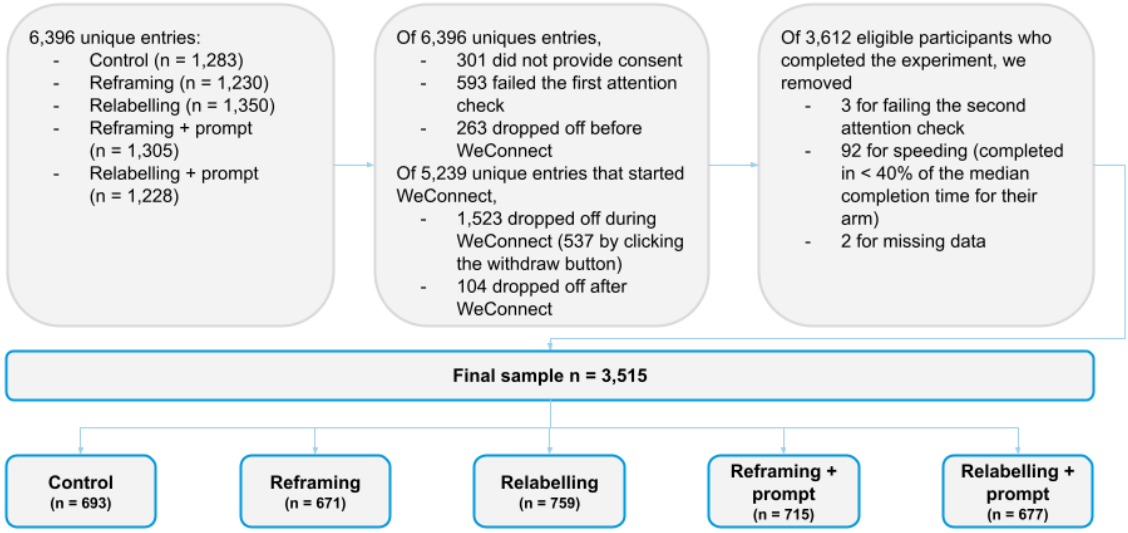
To identify and mitigate any data protection risks, Ofcom and BIT conducted a data protection impact assessment of the research, which was signed off by Ofcom's data protection officer and corporate secretary. As part of the trial, no personal data was collected from the participants. Participants were made aware of this fact through their panel providers before being redirected to our experiment.

To verify the online experiment worked as planned, we conducted a soft launch on ~100 participants. On this sample, we conducted diagnostic checks to ensure data capture proceeded as planned and participants were not reporting any issues. Given the absence of issues, we proceeded to full launch. During data collection, we continued to monitor the incoming sample against the quotas and flagged any criteria adjustments to the panel provider.

In the trial, we imposed additional pre-specified data quality measures in the form of attention and validation checks - only participants who passed these were retained for the analysis. The attention checks were brief questions near the beginning and the end of the trial, which asked people to choose a particular response item to confirm they were paying attention. As a validation check, we looked at the time participants spent working through the trial and

excluded those who were speeding through it (i.e., their survey completion time was less than 40% of the median completion time of that arm). Figure 11 shows the full participant flow with numbers on how many submissions were excluded at which part of the process.

Figure 11. Participant flow diagram.



3.4 Ethical considerations

The research went through BIT's and Ofcom's internal ethics review process and received full approval. The trial's main ethical and safeguarding concerns evolved around exposing participants, as well as BIT and Ofcom researchers, to sensitive content.

It was necessary to expose participants to sensitive content to generate evidence with high external validity, while ensuring we do not cause serious harm to participants. Content types displaying hate, violence and misinformation were included in the trial.

All text and imagery shown to participants in the trial were sourced from publicly available and freely reusable content (uploaded under a Creative Commons License) on platforms like YouTube and Unsplash. The age classification of all violative content was 18+, according to the BBFC content guidelines.

The following risk mitigation and safeguarding measures were implemented to ensure the research did not cause harm to participants and researchers.

1) All content shown to participants in the trial had been reviewed and approved by BIT's ethics reviewer.

2) Participants could only access the trial if they agreed to consent forms provided to them beforehand. The consent forms detailed the research purpose and themes of the reportable content. They outlined the potential risks involved in participating in the

trial, so that participants, particularly those with specific vulnerabilities that might be triggered by the content included, could make an informed choice as to whether to participate. The consent form also made clear to participants that they could leave the survey at any moment without giving a reason.

3) The simulated platform included a visible 'Withdraw' button in the interface that made it easy to leave the trial immediately. Leaving the trial through this emergency button did not impact participants' eligibility for compensation.

4) Regardless of whether the participants decided to complete the study, a debriefing screen was provided with telephone numbers and links signposting to immediate support resources such as the Mind Infoline or the Samaritans hotline.

5) BIT staff voluntarily joined the research after a risk briefing and were allowed to withdraw at any point without penalties. If team members became distressed, they were allowed to switch to lower-risk roles.

6) Mental health support from BIT was available to the researchers, including Mental Health First Aiders and an Employee Assistance Programme.

7) When sensitive content was shared with Ofcom (e.g., for test-link preview), sensitive content warnings were used to alert staff involved in the trial to potential risks.

8) Ofcom equally implemented internal safeguards to protect staff exposed to sensitive content as part of this research.

3.5 Analytical framework

We followed a pre-specified analysis framework which involved allocating our variables to primary, secondary, and exploratory outcomes based on an agreed upon hierarchy. We ran these analyses and made the following six comparisons for each:

- Control vs. Reframing
- Control vs. Relabelling
- Control vs. Reframing + prompt
- Control vs. Relabelling + prompt
- Reframing vs. Reframing + prompt
- Relabelling vs. Relabelling + prompt

3.5.1 Primary outcome

The primary outcome was whether the participant clicked on the Community Guidelines (called 'Dos and Don'ts' in the Relabelling and Relabelling + prompt arms) at least once at any point during the experiment, either through the sign-up page, gear icon, or prompt (clicks through the prompt were only available for the Reframing + prompt arm and the Relabelling + prompt arm). Clicking at least once was coded as 1 and not clicking at all was coded as 0.

3.5.2 Secondary outcomes

Secondary outcome 1: whether a participant reported at least one of the eight violative posts (coded as 1) or not (coded as 0).

Secondary outcome 2: whether a participant reposted at least one of the eight violative posts (coded as 1) or not (coded as 0).

3.5.3 Exploratory outcomes

Exploratory outcome 1: a numerical count of the number of violative posts reported.

Exploratory outcome 2: a numerical count of the number of violative posts reposted.

Exploratory outcome 3: whether participants recalled that WeConnect had rules/guidelines about appropriate behaviour or what content is/is not allowed (correct recall coded as 1; incorrect recall and 'don't know' responses coded as 0).

Exploratory outcome 4: whether participants reported at least one of the sixteen benign posts (coded as 1) or not (coded as 0).

3.5.4 Analytical strategy

We checked for differential attrition on the unique entries to the experiment who provided consent, passed the attention check and who made it to or past the WeConnect platform without dropping off (n = 5,239). We used a linear regression with the last page of the experiment they completed as the outcome variable and the arm as the predictor variable (See Table 15 in Annex B). We then checked that our final sample (n = 3,515) was balanced in terms of demographics (age, gender, ethnicity, annual household income (pre-tax), education, urbanicity, employment, region, social grade, and social media platform use) across arms using chi-squared tests for categorical variables and analysis of variance for continuous variables (See Table 16 and Table 17 in Annex B, respectively).

4. Results

4.1 Sample characteristics

We did not find evidence of differential attrition on the unique entries to the experiment (adjusted R^2 = 0.0005, $F(4,5234) = 1.644$, $p = .16$). The number of people who dropped off the experiment on or after the WeConnect feed are reported in Table 3. The demographics for our final sample (N = 3,515) are reported in Table 4. The sample was balanced across treatment arms for all variables (all $p > .05$), except for education, (X^2 (4) = 10.65, $p < .05$) and income (X^2 (4) = 9.62, $p < .05$). N per arm for unbalanced demographics are reported in Table 5. Despite this, by including education and income as covariate in all statistical models as planned, the effects of this imbalance were minimal. Since the sample was generally balanced on demographics, we continued with our prespecified analysis plan.

Table 3. Number of people in each arm who dropped off the experiment on or after the WeConnect feed.

Arm	Drop off number
Control	308
Reframing	303
Relabelling	317
Reframing + prompt	360
Relabelling + prompt	339

Table 4. Sample demographics for final sample (n = 3,515).

Age	
18-24	12%
25-54	64%
55 and over	23%
Gender	
Male	50%
Female	50%
Other (e.g. nonbinary)	1%
Ethnicity	
White	85%
Asian	6%
Black	5%

Mixed or other	4%
Annual pre-tax income	
£40,000 or over	48%
Less than £40,000	52%
Education	
Degree	32%
No degree	65%
Prefer not to say	4%
Urbanicity	
Urban	30%
Suburban	48%
Rural	22%
Employed	
Employed	72%
Unemployed	3%
Inactive	26%
Location	
London	13%
Midlands	17%
North	25%
South & East	31%
Wales, Scotland & Northern Ireland	15%
Social grade	
High	37%
Medium	56%
Low	7%
Don’t know	< 1%

Note. Some variable totals do not sum to 100% due to rounding.

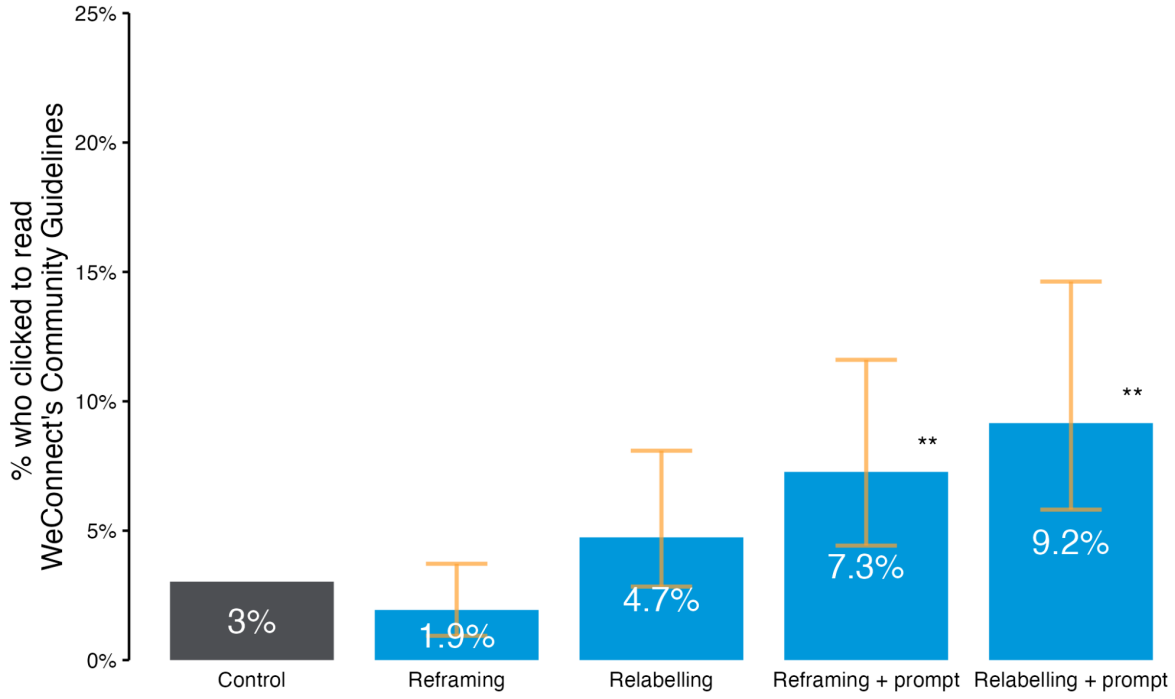
Table 5. N per arm for unbalanced demographics.

	Control	Reframing	Relabelling	Reframing + prompt	Relabelling + prompt
Education (N who are educated to degree level or higher)	197	226	224	230	239
Income (N who have above median income)	332	323	328	365	329

4.2 Primary analysis: Whether participants clicked to read WeConnect's Community Guidelines/Dos and Don'ts

Participants who saw the Reframing + prompt arm or Relabelling + prompt arm were significantly more likely to click to read WeConnect's Community Guidelines than those in the Control arm (7.3% and 9.2% compared to 3.0% respectively, both $p < .01$). Results are shown in Figure 12 (See Table 18 in Annex B for regression results).

Figure 12. The percentage of participants who clicked to read WeConnect's Community Guidelines (comparing each treatment arm to the Control arm).



** p < .01, * p < .05, + p < .1

Significance is corrected for multiple comparisons (Benjamini-Hochberg).

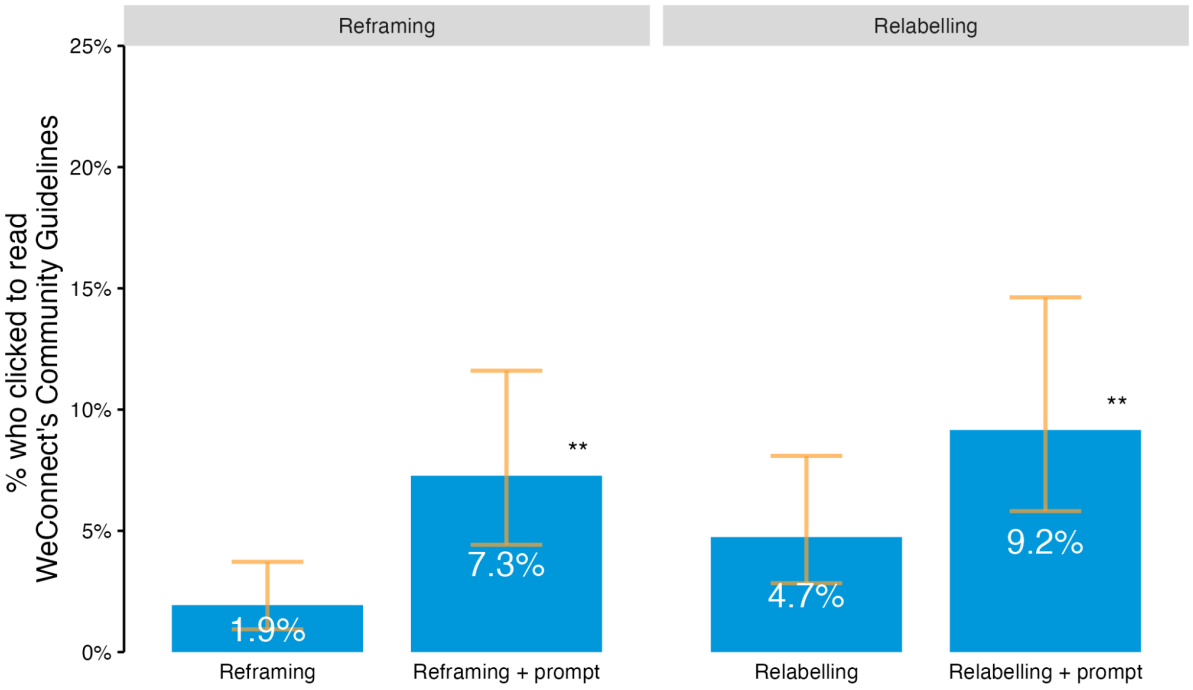
Confidence intervals (95%) on comparisons to the Control arm are not corrected for multiple comparisons.

Regression controls for age, gender, income, education, ethnicity, platform use and if they completed a previous similar trial.

Treatment bars show the mean for that group.

Participants who saw a prompt were significantly more likely to click to read WeConnect's Community Guidelines than those who saw the same message at sign-up only (7.3% vs. 1.9% for the Reframing arms; 9.2% vs. 4.7% for the Relabelling arms, both $p < .01$). Results are shown in Figure 13.

Figure 13. The percentage of participants who clicked to read WeConnect's Community Guidelines (comparing between Reframing arms and between Relabelling arms).



** p < .01, * p < .05, + p < .1

Significance is corrected for multiple comparisons (Benjamini-Hochberg).

Confidence intervals (95%) on comparisons to the Control arm are not corrected for multiple comparisons.

Regression controls for age, gender, income, education, ethnicity, platform use and if they completed a previous similar trial.

Treatment bars show the mean for that group.

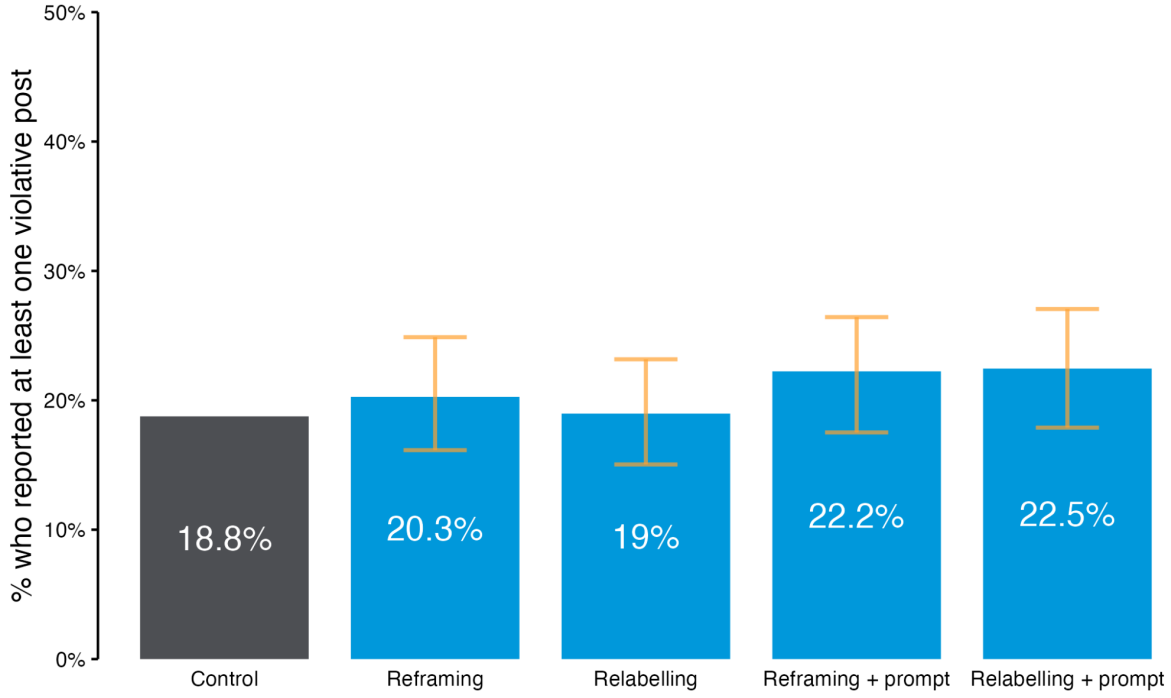
Post hoc, we also checked for significant differences between the best performing arms (both prompt arms). There were no significant differences between clicks to read WeConnect's Community Guidelines, between those in the Reframing + prompt arm and those in the Relabelling + prompt arm ($p > .05$).

4.3 Secondary analysis: Reporting / Reposting violative content

Reporting content

Overall, 20.5% of participants reported at least one violative post on WeConnect. This was not significantly different between the Control arm and any of the treatment arms. Over 90% of the reports of violative posts were accurate (i.e., reporting a violent video for violence. Results are shown in Figure 14 (See Table 19 in Annex B for regression results).

Figure 14. The percentage of participants who reported at least one violative post on WeConnect.



** p < .01, * p < .05, + p < .1

Significance is corrected for multiple comparisons (Benjamini-Hochberg).

Confidence intervals (95%) on comparisons to the Control arm are not corrected for multiple comparisons.

Regression controls for age, gender, income, education, ethnicity, platform use and if they completed a previous similar trial.

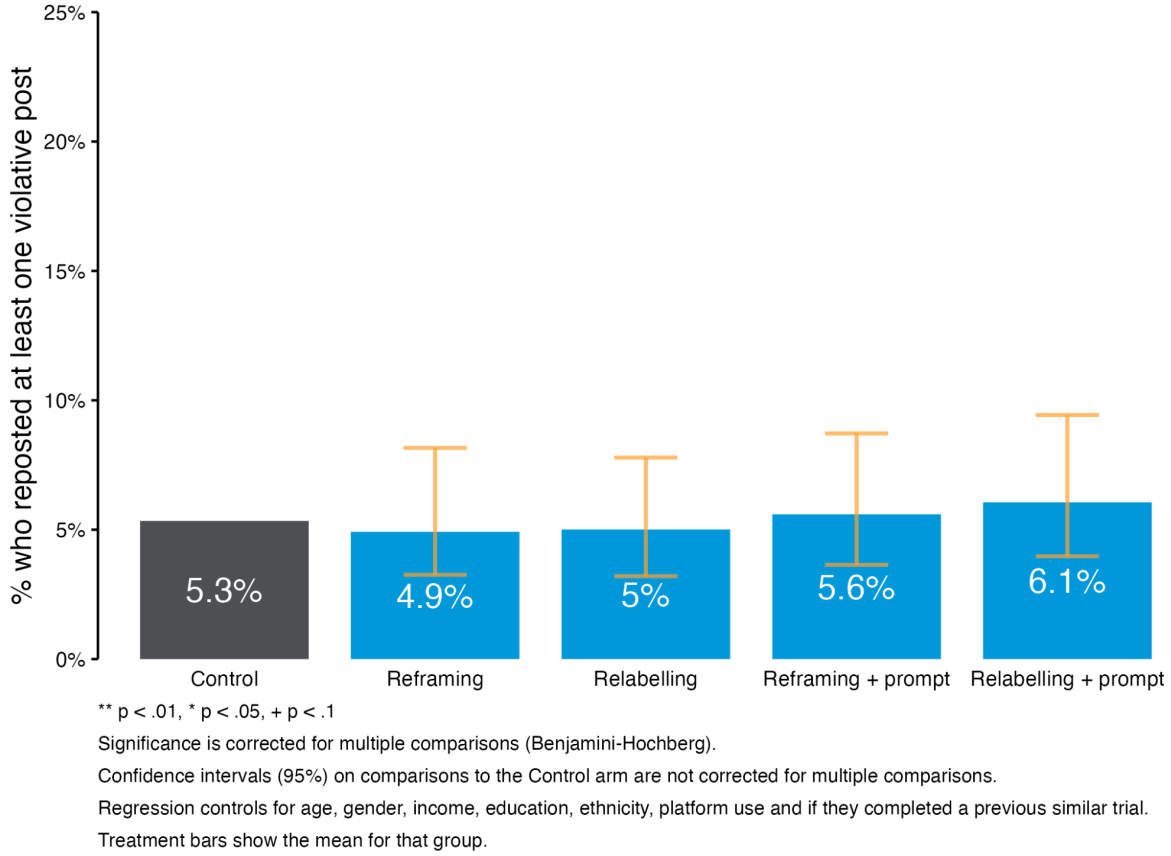
Treatment bars show the mean for that group.

There were also no differences between the Reframing and Reframing + prompt or between the Relabelling and Relabelling + prompt arms ($p > .05$).

Reposting content

Overall, 5.4% of participants reposted at least one violative post on WeConnect. This was not significantly different between the Control arm and any of the treatment arms (all $p > .05$). Results are shown in Figure 15 (See Table 20 in Annex B for regression results).

Figure 15. The percentage of participants who reposted at least one violative post on WeConnect.



** p < .01, * p < .05, + p < .1
Significance is corrected for multiple comparisons (Benjamini-Hochberg).
Confidence intervals (95%) on comparisons to the Control arm are not corrected for multiple comparisons.
Regression controls for age, gender, income, education, ethnicity, platform use and if they completed a previous similar trial.
Treatment bars show the mean for that group.

There were also no differences between the Reframing and Reframing + prompt or between the Relabelling and Relabelling + prompt arms (both $p > .05$).

4.4 Exploratory analysis

Number of violative posts reported

2,794 out of 3,515 participants (79%) did not report any violative content. Among those who reported at least one violative post (20.5%), the average number of pieces of content reported was 3.9. Therefore, these values were overdispersed (mean = 0.80, variance = 3.57). To account for the over dispersion, a zero-inflated negative binomial model was used to analyse the number of violative posts reported.

There were no significant differences between the Control arm and any of the treatment arms in the number of violative posts participants reported (all $p > .05$). There were also no differences between the Reframing and Reframing + prompt or between the Relabelling and Relabelling + prompt arms (both $p > .05$). Results are shown in Table 6.

Number of violative posts reposted

3,326 out of 3,515 participants (95%) did not repost any violative content. To account for the excess of zeros, a zero-inflated Poisson regression was used to analyse the number of violative posts reposted.

There were no significant differences between the Control arm and any of the treatment arms in the number of violative posts participants reposted (all $p > .05$).

Table 6. The mean number of posts reported or reposted (comparing each treatment arm to the Control arm).

All content types (Not tested for significant differences)					
Outcome	**Control**	**Reframing**	**Relabelling**	**Reframing + prompt**	**Relabelling + prompt**
Number of posts **reported**	0.75	0.84	0.72	0.93	0.92
Number of posts **reposted**	0.51	0.45	0.34	0.41	0.41
Violative posts only					
Outcome	**Control**	**Reframing**	**Relabelling**	**Reframing + prompt**	**Relabelling + prompt**
Number of violative posts **reported**	0.73	0.83 [0.60-0.94]	0.70 [0.47-0.81]	0.90 [0.58-0.90]	0.88 [0.55-0.88]
Number of violative posts **reposted**	0.07	0.08 [-0.31-1.02]	0.07 [-0.79-0.73]	0.07 [-1.40-0.28]	0.09 [-0.45-1.02]

** p < .01, * p < .05, + < .1
This table reports the means for each arm and results of regressions comparing each treatment arm against the Control arm.
Regressions control for age, gender, income, education, ethnicity, platform use, and if they completed a previous similar trial.
Significance and confidence intervals (95%; reported in brackets) are not corrected for multiple comparisons.

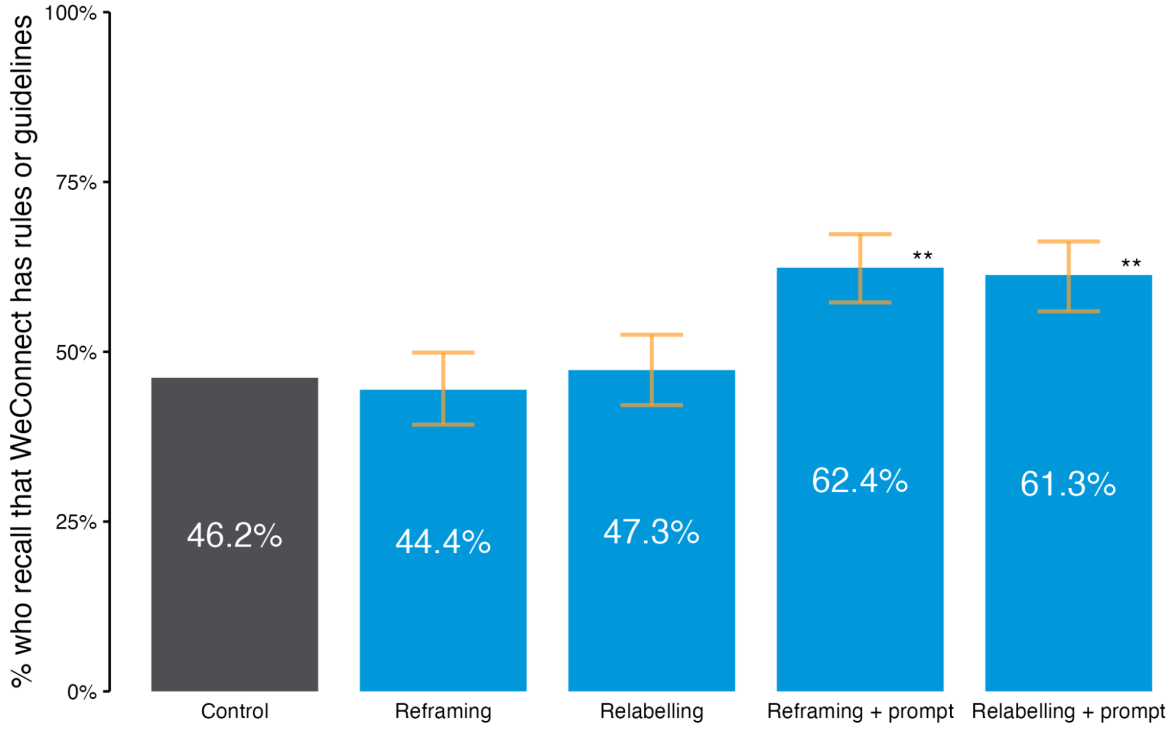
Recall

Overall, 1,838 of 3,515 participants (52.3%) recalled that there are rules or guidelines about how to use WeConnect, 710 of 3,515 participants (20.2%) said they did not recall, and 967 of 3,515 participants (27.5%) said they did not know. Significantly more people correctly

recalled that there are rules or guidelines about how to use WeConnect in the prompt arms than in the Control arm (62.4% in the Reframing + prompt arm and 61.3% in the Relabelling + prompt arm compared to 46.2% in the Control arm, $p < .01$). There were no significant differences between the Control arm and Reframing or Relabelling arms ($p > .05$). Results are shown in Figure 16.

Participants who saw a prompt were significantly more likely to correctly recall that WeConnect had rules or guidelines than those who saw the same message at sign up only (62.4% vs. 44.4% for the Reframing arms; 61.3% vs. 47.3% for the Relabelling arms, both p < .01). Results are shown in Figure 17.

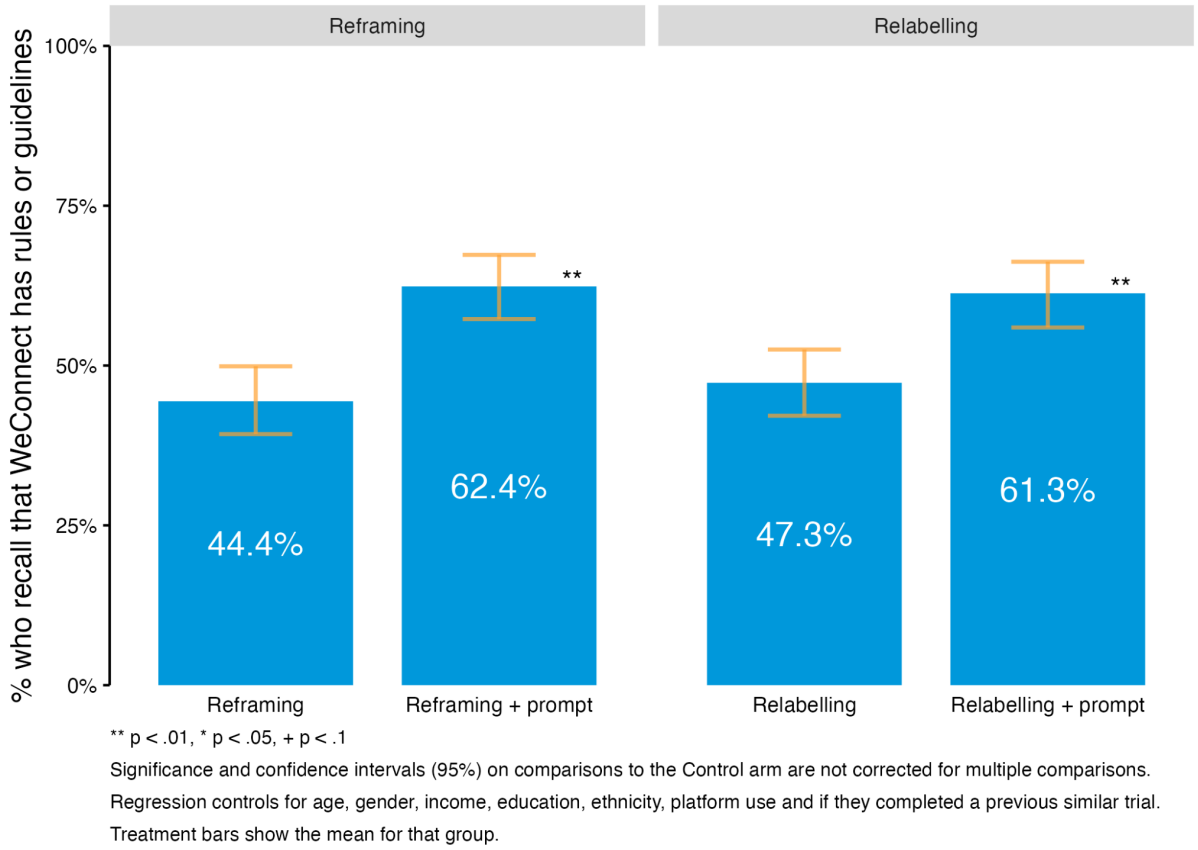
Figure 16. The percentage of participants who correctly recalled that WeConnect has rules or guidelines.



** p < .01, * p < .05, + p < .1

Significance and confidence intervals (95%) on comparisons to the Control arm are not corrected for multiple comparisons. Regression controls for age, gender, income, education, ethnicity, platform use and if they completed a previous similar trial. Treatment bars show the mean for that group.

Figure 17. The percentage of participants who correctly recalled that WeConnect has rules or guidelines (comparing between Reframing arms and between Relabelling arms).



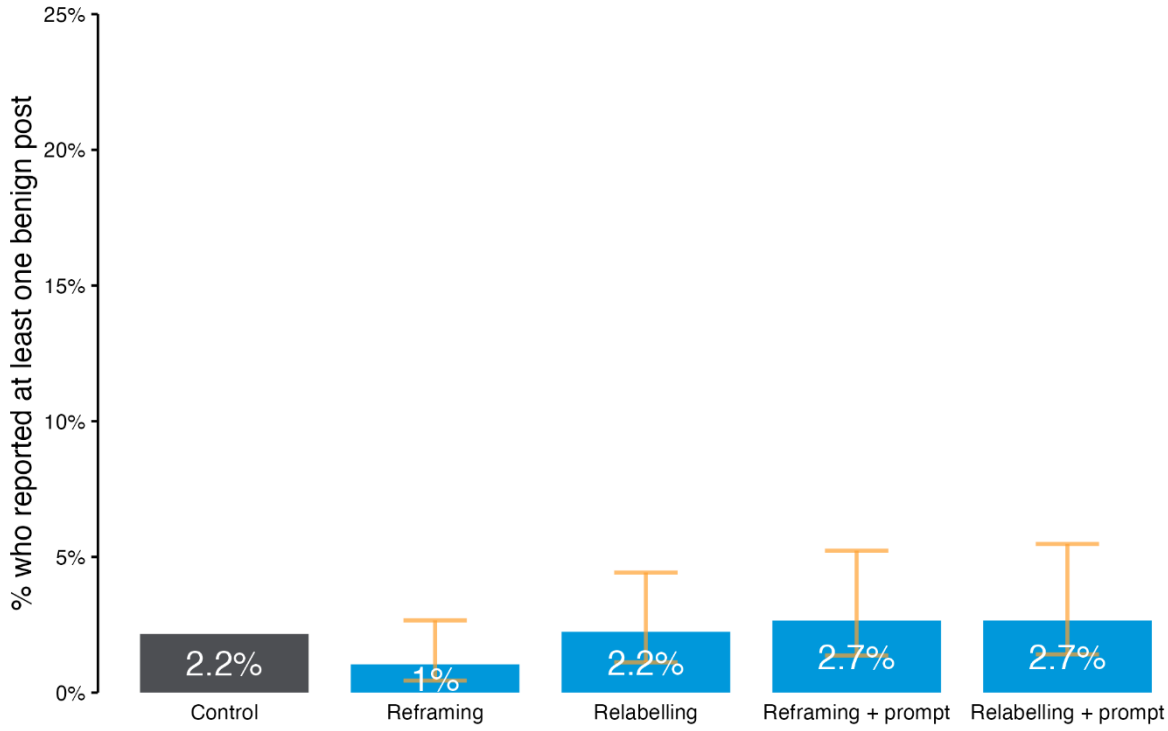
Post hoc, we also looked at the rate of recall for those who clicked to read the Community Guidelines and those who did not. 85.3% of those who clicked to read WeConnect's Community Guidelines recalled that WeConnect had rules or guidelines compared to 50.5% of those who did not click to read the Community Guidelines. This was not tested for significant differences.

Reports on benign posts

Overall, 2.2% of participants incorrectly reported at least one benign post when scrolling through the feed. There were no significant differences between any of the treatment arms and Control arm ($p > .05$). Results are shown in Figure 18.

Significantly more participants in the Reframing + prompt arm reported at least one benign post than in the Reframing arm ($p < .05$): 19 out of 715 participants (2.7%) vs. 7 out of 671 participants (1.0%. There was no significant difference between reports on benign posts in the Relabelling and Relabelling + prompt arms ($p > .05$). Results are shown in Figure 19.

Figure 18. The percentage of participants who reported at least one benign post.



** p < .01, * p < .05, + p < .1

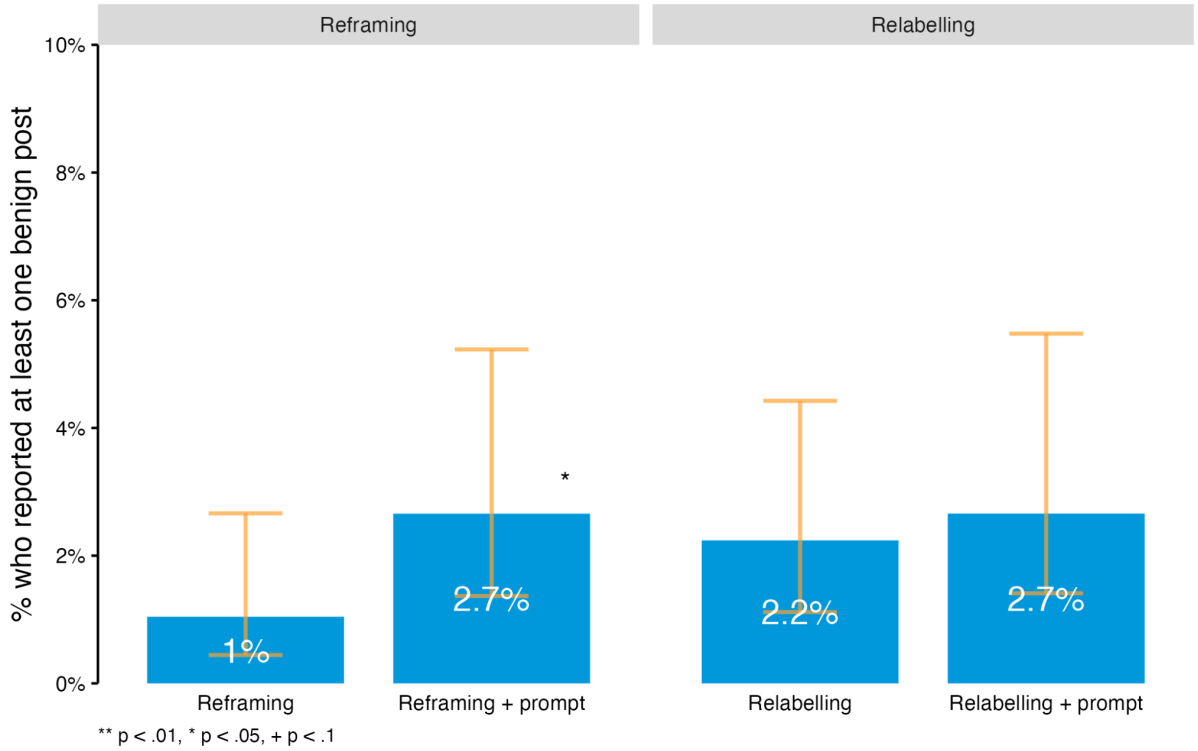
Significance and confidence intervals (95%) are not corrected for multiple comparisons.

Regression controls for age, gender, income, education, ethnicity, platform use and if they completed a previous similar trial.

Treatment bars show the mean for that group.

Figure 19. The percentage of participants who reported at least one benign post (comparing between Reframing arms and between Relabelling arms).

.



4.5 Exploratory descriptive analysis

When participants clicked to the Community Guidelines

Overall, 184 (5.2%) participants clicked to read WeConnect's Community Guidelines, through one of three routes (some participants clicked through more than once):

- **Sign-up:** Across all arms, 89 (2.5%) participants clicked through to Community Guidelines at the sign-up page. The rate of click through was fairly similar across treatment arms.
- **Through the gear icon during the main feed:** 64 (1.8%) participants clicked through the gear icon to the Community Guidelines. Of those who clicked through the gear icon, 43 (67%) participants were in the prompt arms compared to 21 (33%) in the no prompt arms (including the Control arm).
- **When prompted (only in the two arms where a mid-feed prompt was delivered):** 41 (3%) participants clicked through to the Community Guidelines across these two arms.

Responses by arm are shown in Table 7.

Table 7. The number of participants who clicked the Community Guidelines through each route compared by arm.

	Control	Reframing	Relabelling	Reframing + prompt	Relabelling + prompt
Clicked to read the Community Guidelines at least once	21	13	36	52	62
Clicked at sign up	15	8	27	13	26
Clicked through the gear icon	6	5	10	20	23
Clicked through the prompt	-	-	-	22	19
Total participants in arm	693	671	759	715	677

In the follow up survey, participants were asked why they clicked or did not click to read WeConnect's Community Guidelines. The top reasons participants clicked on WeConnect's Community Guidelines (n = 184) were that they wanted to make sure they were following the guidelines (55%) and that they wanted to find out more about WeConnect (53%). The full list of response results is in Table 8.

Table 8. Why participants clicked to read WeConnect's Community Guidelines.

In WeConnect, why did you choose to click on WeConnect's Community Guidelines/Dos and Don'ts? (Participants could select more than one option, n = 184)	
I wanted to make sure I was following WeConnect's guidelines	55%
I wanted to find out more information about WeConnect	53%
I want to keep myself and others safe on WeConnect	36%
I was reminded about it (Prompt arms only, n = 114)	18%
I thought I wouldn't be able to use the site unless I had	17%
I didn't mean to click on the link	2%

Other (e.g. “Tapped by accident”)	1%

The top reasons participants did not click on WeConnect’s Community Guidelines (n = 3,331) were that they assumed it would be the same as other platforms they have used (36%), they didn’t see the link (32%) and that they didn’t realise they could (30%). The full list of response results is in Table 9.

Table 9. Why participants did not click to read WeConnect’s Community Guidelines.

In WeConnect, why did you choose not to click on WeConnect’s Community Guidelines/Do’s and Dont’s? (Participants could select more than one option, n = 3,331)	
I assumed it would be the same as other platforms I have used	36%
I didn’t see the link	32%
I didn’t realise I could	30%
I feel I didn’t need to know all the details in order to use the platform	21%
I thought it would take too long	14%
I know how to use WeConnect	9%
I didn’t think it’d be useful	9%
I didn’t think I would be able to understand the information	4%
I don’t care about being safe online	2%
Other (e.g., “Prevents freedom of speech”, “Probably too wordy”)	2%

Attitude to prompts

Participants in the prompt arms (n = 1,392) were also asked whether they thought the prompt was annoying and whether they thought it was useful, and why. Full responses are shown in Table 10.

Table 10. Attitudes towards the prompts.

% who said the prompt was…	**Reframing + prompt**	**Relabelling + prompt**
Not at all annoying	27%	25%
A little annoying	25%	26%
Moderately annoying	17%	20%
Very annoying	13%	11%
I didn’t notice a prompt	19%	17%
% who said the prompt was…	**Reframing + prompt**	**Relabelling + prompt**
Not at all useful	18%	23%

A little useful	23%	22%
Moderately useful	23%	23%
Very useful	17%	15%
I didn't notice a prompt	19%	17%

Overall, 56% said the prompt was at least a little annoying (54% for the Reframing prompt and 57% for the Relabelling prompt). Of those who said the prompt was at least a little annoying (n = 776), the thing they found the most annoying was that it distracted them from scrolling on the platform (55%). The full list of response results is in Table 11.

Table 11. Why participants thought the prompt was annoying.

What did you find annoying about the pop-up message? (Participants could select more than one option, n = 776)	
The message distracted me from scrolling on the platform	55%
I didn't want to read WeConnect's Community Guidelines/ Dos and Don'ts	27%
I already knew how to check WeConnect's Community Guidelines/ Dos and Don'ts if I wanted to	20%
The message was not relevant to me	15%
I already read WeConnect's Community Guidelines/ Dos and Don'ts when I signed up	14%
Other (e.g. "I didn't know what I was meant to do about it", "Wasn't clear what it was for")	2%

Overall, 61% said the prompt was at least a little useful (63% for the Reframing prompt and 60% for the Relabelling prompt). Of those who thought it was at least a little useful (n = 855) the biggest reasons were that it was useful to know that WeConnect has Community Guidelines (42%) and that it appeared while they were scrolling through the feed (36%). The full list of responses is in Table 12.

Table 12. Why participants thought the prompt was useful.

What did you find useful about the pop-up message? (Participants could select more than one option, n = 855)	
It was useful to know that WeConnect have Community Guidelines /Dos and Don'ts	42%
It appeared while I was scrolling through the feed	36%
The pop-up made it easy to check WeConnect's Community Guidelines/Dos and Don'ts	33%
It made me think about the rules on WeConnect	32%

Other (e.g. “It alerted me to the existence of the Dos and Don'ts”, “It is good if you don’t know where to find them but do want to read them”)	2%

Why participants reported content

When reporting content, participants were asked whether they wanted to report it for hate speech, violence, misinformation, spam, another reason, or just that they did not like it. Participants who reported violative content (posts containing hate speech, violence, or misinformation) were generally accurate at categorising the type of violative content in the report form (Table 13).

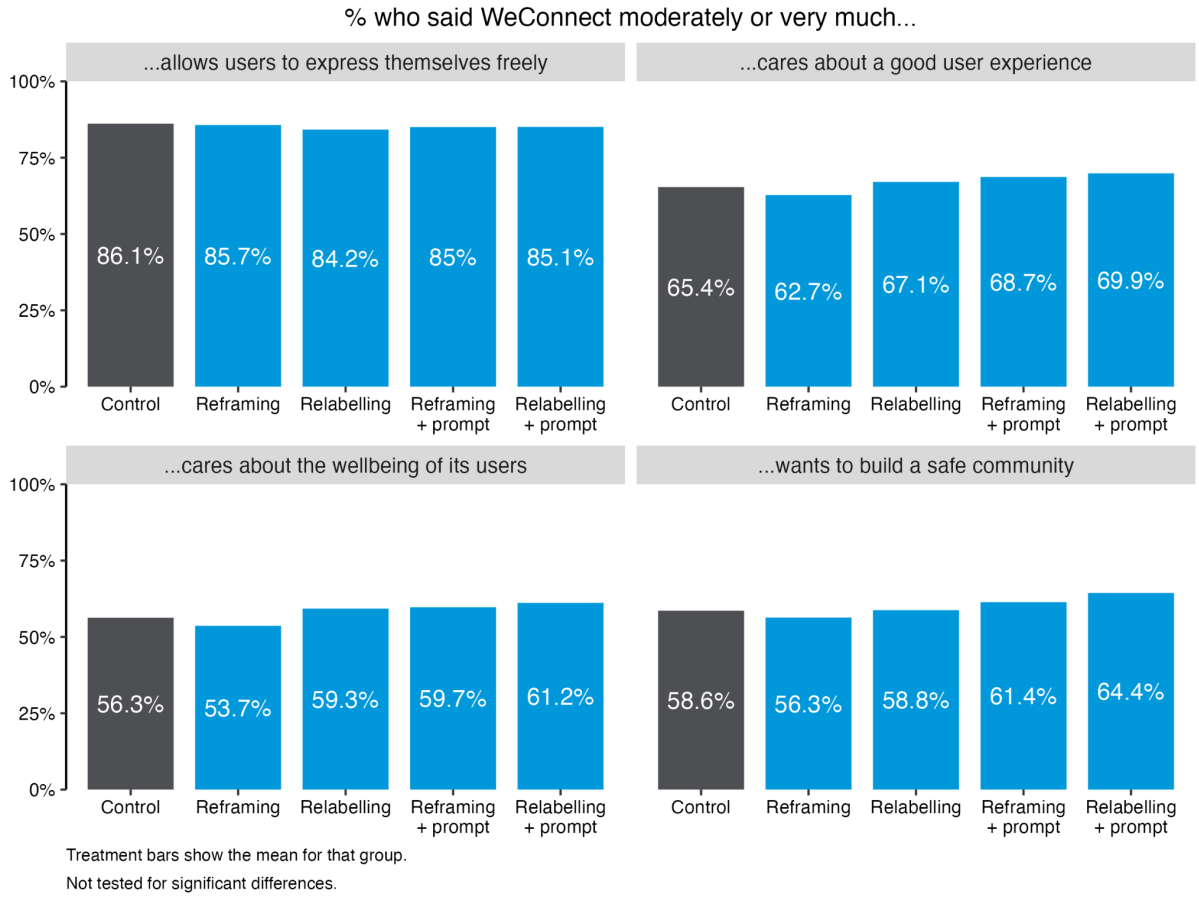
Table 13. The reports and reasons for reporting for each violative content post.

Content type	Description	% of participants who reported the post	% of participants who reported the post and correctly categorised the content type
Hate speech	A video expressing misogynistic views	9%	7%
	A short text post expressing hateful views about immigration	13%	12%
	A short text post expressing homophobic views	12%	11%
	A short text post expressing transphobic views	11%	10%
Violence	A video depicting a street fight	14%	13%
Misinformation	A short text post expressing misinformation about vaccines	11%	10%
	A long text post expressing misinformation about 15-minute cities	6%	5%
	A long text post expressing misinformation about the World Economic Forum	7%	6%

Perceptions of safety

In the post-trial survey, participants were also asked about their perceptions of WeConnect. Responses by arm are shown in Figure 20.

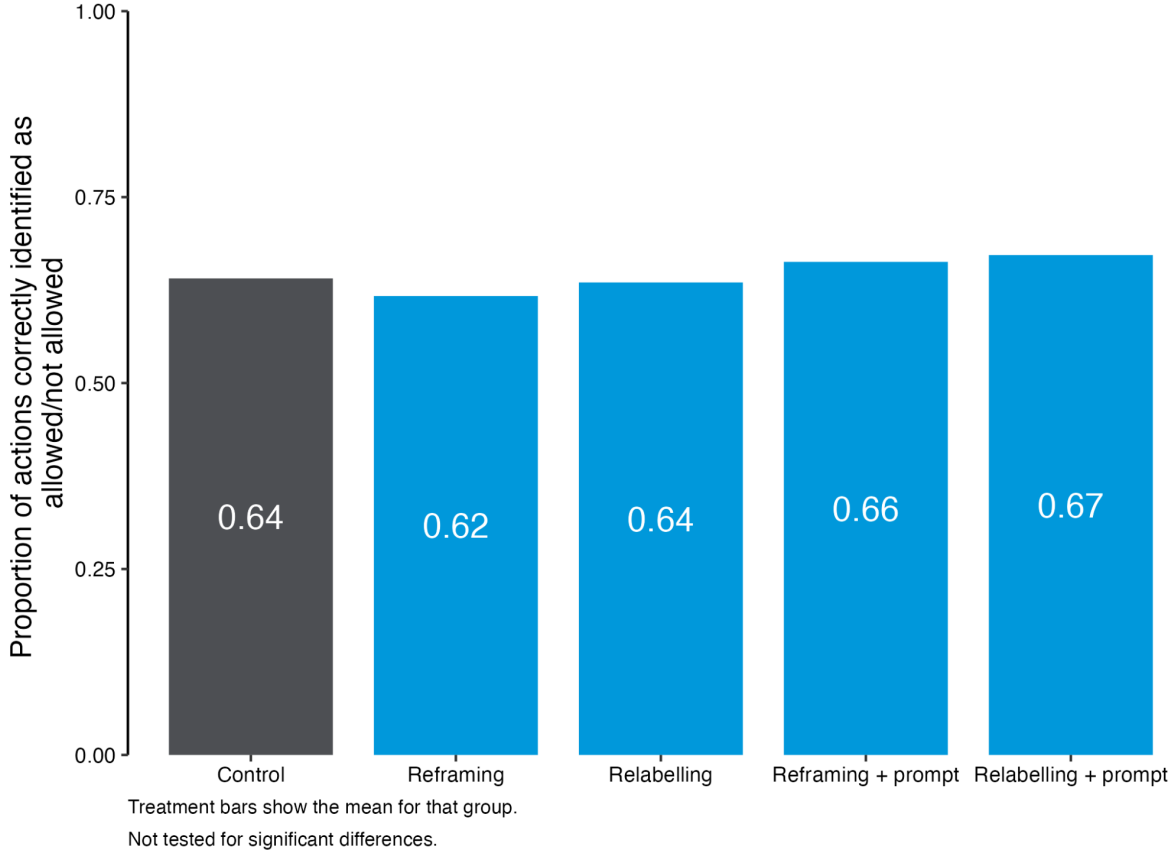
Figure 20. Perceptions of safety on WeConnect on different metrics by arm.



Comprehension

Participants were shown six actions that someone could take on WeConnect (e.g. specific comments, reposts, reports, and posts) and asked to identify whether they were or were not allowed under the Community Guidelines. On average, participants correctly identified 65% of the actions as allowed or not allowed. Results by treatment arm are shown in Figure 21. Overall, participants who clicked to read the Community Guidelines at least once correctly identified 81% of actions as allowed or not allowed, whereas participants who did not click only correctly recalled 64%.

Figure 21. Fraction of actions correctly identified as allowed or not allowed by treatment arm.



5. Summary and Limitations

The focus of this trial was to test different interventions that encouraged users to engage with Community Guidelines. We measured this by whether people clicked through to view the Community Guidelines. We tested interventions with different framing (relabelling or reframing) and different timing (one with a message only at sign-up, and one with an additional prompt in the middle of the feed) (see Table 1).

Prompting was an effective way to encourage users to click to read the Community Guidelines. Overall, participants who viewed a prompt were significantly more likely to click on WeConnect's Community Guidelines (7.3% for the Reframing + prompt and 9.2% for the Relabelling + prompt) than those who only saw a message at sign-up (1.9% for the Reframing arm, 4.7% for Relabelling arm), or the Control arm (3%).

Participants that did check the Community Guidelines did so because they were motivated to understand and follow them. When surveyed 55% said they wanted to ensure they were following the guidelines and 53% that they wanted to find out more.

Participants' attitudes towards prompts were mixed. Whilst some participants thought the prompts were annoying (30%), others thought they were useful (19%). The main reason participants thought the prompts were annoying was because it distracted them from scrolling through the platform (55%) and the main reason they found it useful was because they thought it was useful to know that WeConnect has Community Guidelines (42%).

Some exploratory results suggests that prompts increased the recall of rules or guidelines in WeConnect's Community Guidelines. Overall, 62.4% exposed to the Reframing + prompt arm and 61.3% in the Relabelling + prompt arm recalled the rules and guidelines compared to 46.2% in the Control arm, 44.4% in the Reframing arm and 47.3% in the Relabelling arm.

The interventions do not appear to have influenced user reporting or reposting of content. Across all arms, including the Control, there were no significant differences in user reporting or reposting of content. Overall, approximately 20% of participants reported at least one violative post, almost always for the correct reason, and 5% reposted one violative post.

None of the interventions increased participants' comprehension of what was allowed or disallowed on the platform. Across all the arms, participants correctly identified 65% of the actions as allowed or disallowed on the platform.

Limitations

Given the environment we ran our experiment in, several limitations apply to our findings. No matter how carefully designed, a simulated platform is not able to fully replicate the incentives and motivations that guide users' behaviours on social media. Importantly, real-life

violative content may include content that is more harmful and more personalised than the content shown in our research. Moreover, the short timescale at which our online experiment had to measure outcomes limits the conclusions that can be drawn with respect to the long-term effects of our interventions. In addition, we were limited to measuring engagement with the Community Guidelines solely through click-throughs, without any assurance that participants actually read or understood the content even if they clicked through. Despite these limitations, we believe online RCTs are a useful tool for building the evidence base.

6. Annexes

Annex A: Summary of barrier prioritisation and intervention development process

Ofcom's BI Hub conducted the following research activities to develop the interventions for the trial.

1. **Rapid evidence review.** Ofcom's BI Hub conducted a rapid evidence review of literature relevant to engagement and comprehension of T&Cs type documents and the relationship this has with relevant behaviours.
2. **Behavioural diagnosis 1.** Using the findings from the evidence review as well as team's expertise, Ofcom's BI Hub used the Capability – Opportunity – Motivation (COM-B) model to map out potential enablers and barriers to users accessing T&Cs and checking platform rules (see Table 14 for an overview).[10] The COM-B model proposes that behaviour is made up of three necessary components: capability, opportunity and motivation and was used in the design of a quantitative survey that explored potential influences on relevant behaviours and to inform an analysis of the results.
3. **Quantitative survey.** Ofcom used a quantitative survey, informed by the COM-B model, to test their assumptions across two key behaviours on a VSP or on social media: accessing T&Cs and checking platform rules.[11]
4. **Behavioural diagnosis 2.** The findings of the survey were used to update the assessment of enablers and barriers (see Table 14).
5. **BI Hub intervention development sprints.** Ofcom's BI Hub collated the evidence from their rapid evidence review, behavioural analysis and survey to develop a longlist of intervention ideas to increase engagement with platform T&Cs. Behavioural science frameworks (e.g., Behaviour Change Wheel) were also used to guide this process.
6. **Prioritisation of interventions.** To develop different types of messages for the trial, BIT ran a workshop with Ofcom's BI Hub and relevant Ofcom policy stakeholders to prioritise the intervention options. These ideas were then prioritised based on their relevance to the target behaviour, expected impact (i.e., would it change user behaviour?) and feasibility (i.e. would platforms do it?).

[10] Michie, S., van Stralen, M. M., & West, R. 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions.
[11] Ofcom, 2024. Terms and conditions and content controls.

Table 14: Overview of behavioural diagnosis.

COM-B Component	Sub-component	Assumption (behavioural diagnosis 1)	Updated assumption using survey evidence (behavioural diagnosis 2)
Capability	**Cognitive skills**	Users need to know where/how to find the platform's rules.	Many users do not know where to locate platforms rules. **Evidence**: 68% did not list either Community Guidelines or T&Cs as places they would check if they were unsure about the platform rules.
	Awareness	Users need to be aware that platforms have rules, why they are important and where they can find them.	Most users have heard of T&Cs, but many do not know that they contain platform rules. **Evidence**: Only 5% said they had never come across T&Cs but 68% did not list either Community Guidelines or T&Cs as places they would check if they were unsure about the platform rules.
	Attention	Users need to sustain attention long enough to check platform rules, in an environment with a lot of other attentional cues.	**Not tested within the survey**
	Evaluating options	There are lots of different policy documents that a user might need to check to familiarise themselves with the rules. These also differ across platforms.	**Not tested within the survey**
	Memory	Users need to remember how to familiarise themselves with platform rules.	**Not tested within the survey**
Opportunity	**Prompts in the environment**	Users are often not encouraged to familiarise themselves will platform rules (e.g., through the use of 'click wrap' agreements).	Users are more likely to check platform rules when prompted by the environment. **Evidence:** 23% check platform rules when they see something they do not think should be there; 19% check platform rules when prompted by the platform.
	Resources & time	The resources provided to users to help them check the rules (e.g. T&Cs) are often inadequate and do not facilitate user understanding.	Users perceive T&Cs as long and complex and do not think they have the time or ability to understand them. **Evidence**: Of those that do not read T&Cs while signing up, 65% said it was because it would take too long; 45% said they find them overwhelming.
	Opportunities in the Environment	There are opportunities to check platform rules (e.g., via T&Cs) but the design of platforms often does not facilitate this	Some users do not access T&Cs/platform rules because there is no obligation or encouragement to.

COM-B Component	Sub-component	Assumption (behavioural diagnosis 1)	Updated assumption using survey evidence (behavioural diagnosis 2)
		behaviour.	**Evidence:** Of those that do not read T&Cs while signing up, 20% said it was because platforms do not force them to.
	Social norms	There is not a strong culture of active rule checking within social media users.	**Not tested within survey**
Motivation	**Beliefs about consequences**	Users might believe there are no direct consequences of not checking platform rules.	Users perceive limited value to accessing and reading T&Cs or platform rules. This acts as a strong barrier to accessing. **Evidence:** Of those participants that did not read T&Cs at sign up, 28% thought they all tend to be similar so don't need to read them; 24% trust that platforms are not allowed to do anything illegal; 21% said it would not affect their decision to use the platform. Of those participants that did not check platform rules, 57% are confident they are not going to go anything that would break the rules.
	Beliefs about capabilities	No prior assumption	Some users do not think they have the capability to understand T&Cs, which puts them off accessing. **Evidence:** Of those that did not read T&Cs at sign up, 25% said they would not be able to understand them.
	Identity	Users might not see themselves as people who consistently and actively check the rules on social media.	**Not tested within the survey**
	Emotions	Social media can be a highly emotive environment which does not facilitate actively checking platform rules.	**Not tested within the survey**
	Habits	Many users will not be in the habit of checking platform rules when they are unclear.	**Not tested within the survey**

Annex B: Full results tables

Table 15. Differential attrition (Results from a linear regression)

Predictor	Coefficient	Standard error	t-value	p-value
Intercept	14.21918	0.18520	76.776	< 0.001*
Reframing	-0.08564	0.26362	-0.325	0.745
Relabelling	0.13360	0.25740	-0.519	0.604
Reframing + prompt	-0.38265	0.25752	-1.486	0.137
Relabelling + prompt	-0.38342	0.26147	-1.466	0.143

** p < .01, * p < .05, + < .1
This table shows the results of an OLS regression on unique entries to the experiment who provided consent, passed the attention check and who made it to or past the WeConnect platform without dropping off (n = 5,239), with last page of the experiment they completed as the outcome variable and the treatment arm as the predictor variable (adjusted R^2 = 0.0005).

Table 16. Balance checks (Results from chi-squared tests)

Variable	Test statistic	Degrees of freedom	p-value	n
Age	5.0943	8	0.7474	3,515
Gender	5.3068	4	0.2572	3,490
Education	10.65	4	0.03079*	3,390
Ethnicity	14.539	12	0.2676	3,515
Income	9.6163	4	0.04741*	3,515

Urbanicity	13.409	8	0.09854	3,515
Location	23.588	16	0.09889	3,515
Device type	2.6632	4	0.6157	3,515
Social grade	12.568	12	0.4012	3,515
Employment	6.0431	8	0.6424	3,515

** p < .01, * p < .05, + < .1
This table shows the results of chi-squared tests conducted to check whether categorical demographic variables were balanced across treatment arms, excluding small subgroups.

Table 17. Balance checks (Results from ANOVA on platform use)

	Degrees of freedom	Sum of squares	Mean square	F-value	p-value
Treatment	4	252	63.09	0.701	0.591
Residual	3510	315982	90.02		

** p < .01, * p < .05, + < .1
This table shows the results of an ANOVA test conducted to check whether continuous demographic variables (platform use) was balanced across treatment arms (n = 3,515).

Table 18. Primary analysis (Results from logistic regression on clicking through to read Community Guidelines)

Comparison	Coefficient	Standard error	z-value	p-value
Control - Reframing	0.490	0.358	1.367	0.172
Control - Relabelling	-0.485	0.281	-1.724	0.102
Control - Reframing + prompt	-0.914	0.266	-3.438	0.001**
Control - Relabelling + prompt	-1.191	0.260	-4.574	< 0.001**
Reframing - Reframing + prompt	-1.404	0.316	-4.436	< 0.001**
Relabelling - Relabelling + prompt	-0.707	0.218	-3.234	0.002**

** p < .01, * p < .05, + < .1
Regressions control for age, gender, income, education, ethnicity, platform use and if they completed a similar previous trial.
Significance corrected for multiple comparisons (Benjamini-Hochberg).

Table 19. Secondary analysis 1 (Results from logistic regression on reporting at least one violative post)

Comparison	Coefficient	Standard error	z-value	p-value
Control – Reframing	-0.090	0.138	-0.648	0.890
Control – Relabelling	-0.001	0.136	-0.005	0.996
Control – Reframing + prompt	-0.179	0.134	-1.336	0.727
Control – Relabelling + prompt	-0.208	0.136	-1.532	0.727
Reframing – Reframing + prompt	-0.089	0.133	-0.672	0.890
Relabelling – Relabelling + prompt	-0.207	0.132	-1.568	0.727

** p < .01, * p < .05, + < .1
Regressions control for age, gender, income, education, ethnicity, platform use and if they completed a similar previous trial.
Significance corrected for multiple comparisons (Benjamini-Hochberg).

Table 20. Secondary analysis 2 (Results from logistic regression on reposting at least one violative post)

Comparison	Coefficient	Standard error	z-value	p-value
Control - Reframing	0.0309	0.248	0.125	0.983
Control - Relabelling	0.0647	0.239	0.271	0.948
Control - Reframing + prompt	-0.0631	0.237	-0.267	0.948
Control - Relabelling + prompt	-0.1518	0.235	-0.664	0.890
Reframing - Reframing + prompt	-0.0940	0.243	-0.386	0.948
Relabelling - Relabelling + prompt	-0.2165	0.234	-0.927	0.890

** p < .01, * p < .05, + < .1
Regressions control for age, gender, income, education, ethnicity, platform use and if they completed a similar previous trial.
Significance corrected for multiple comparisons (Benjamini-Hochberg).