Promoting user engagement with Terms and Conditions

Testing behavioural techniques to improve engagement with social media Terms and Conditions (T&Cs)

Behavioural Insights Discussion Paper

Disclaimer

Regulatory Context

Ofcom is publishing this research under its Media Literacy duty. Ofcom has a duty to promote media literacy, including in respect of material available on the internet. Ofcom's approach to media literacy is multi-dimensional and considers a range of aspects including how the design of services can impact on users' ability to participate fully and safely online.

Ofcom also oversees the regulatory regime which requires UK-established Video Sharing Platforms to include measures and processes in their services that protect users from the risk of viewing harmful content.

Additionally, this research will build evidence with respect to Ofcom's new duties under the UK Online Safety Act 2023.

Contents

1. Overview

1.1   Social media Terms and Conditions ('T&Cs') provide the contract between service providers and their users.[1] They typically set out the rules for using a platform, including who can access the service, the content and behaviours permitted (often referred to as 'Community Guidelines'), and the consequences of breaking those rules. They aim to give clarity on what to expect when using a platform and explain any safety tools that are available. In short, a key objective of T&Cs is to act as a starting point for protecting users from harms and creating safe and trusted communities online.

1.2   The effectiveness of social media T&Cs is likely to be impacted by several factors such as - user engagement with and comprehension of T&Cs documents, user motivation to comply with the rules, and enforcement of those rules by the service. Encouraging users to engage with T&Cs and improve their understanding of key information is an important first step in allowing T&Cs to deliver on their objectives. However, previous research suggests that many users choose to accept T&Cs without fully reading them and many do not actively check the rules on social media platforms.[2]

1.3   In this research, we wanted to better understand the user relationship with social media T&Cs, specifically Community Guidelines. We tested behavioural techniques to encourage users to engage with T&Cs and measured the impact of this engagement on subsequent behaviour and interactions with content.

1.4   To do this, we commissioned the Behavioural Insights Team ('BIT') to run an online randomised controlled trial ('RCT') on a simulated social media platform ('WeConnect'). We tested the effectiveness of behavioural techniques in increasing the number of users who access Community Guidelines. We were also interested in the effect of reading Community Guidelines on reporting and reposting of violative content (i.e., content that violates the Community Guidelines) whilst using that platform.

1.5   The interventions we tested on the platform focused on how (i.e., changes to the user interface) and when (i.e., at sign-up or whilst scrolling) the link to Community Guidelines were presented to participants:

- Control: A link to Community Guidelines was presented to participants whilst they were signing up to the platform. This was intended to represent common practices by platforms.

- 'Reframing message': The link to Community Guidelines was presented on the sign-up page alongside a motivational message highlighting their importance.

- 'Relabelling': The name of the Community Guidelines document was changed to be more user-friendly ('Dos and Don'ts') and presented whilst signing up to the platform.

- 'Prompt': A prompt, either including the reframing message ('Reframing prompt') or the relabelled title ('Relabelling prompt'), was shown whilst participants were scrolling on the platform and had been exposed to some content.

---

[1] By 'Terms and Conditions', we mean any document setting out the rules for using a social media platform, including both Terms of Service and Community Guidelines.
[2] Ofcom, 2024. Terms and conditions and content controls

Key findings

Prompting was the only 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% for the Reframing prompt and 9% for the Relabelling prompt) than those who only saw a message at sign-up (2% for the Reframing message, 5% for Relabelling), or those in the Control (3%).

Neither the Reframing message nor the Relabelling had a statistically significant effect on the number of participants accessing the Community Guidelines at the sign-up stage.

The findings suggest that it can be useful to encourage users to engage with T&Cs documents at relevant points of time, outside of the sign-up stage. However, even after prompting, many users still choose not to access the documents. This suggests the context of how and when prompts are delivered will be important in determining their effectiveness.

The interventions did not influence user reporting or reposting of violative content.

Across all intervention groups, including the Control, there were no significant differences in user reporting or reposting of violative content.

This could suggest that other factors, outside of reading and understanding Community Guidelines, have a stronger influence on users' decisions to report or repost violative content. This suggests that further strategies may be necessary to encourage compliance with the platform's rules, in addition to increasing engagement with T&Cs.

Prompts increased recall of the platform rules, but not comprehension.

Overall, just over 60% of participants exposed to a prompt correctly recalled that WeConnect had rules and guidelines compared to, approximately, 45% in all the other groups.[3] However, nearly 40% of participants could not correctly recall the information included in the prompt.

This could be due to users not reading the information initially or quickly forgetting it. Some users may experience 'alert fatigue' from over-exposure to prompts across these types of platforms. It suggests the use of prompts needs to be selective and targeted to minimise this risk.

Being able to remember that there are rules did not necessarily translate into people understanding the rules better. Across all the groups, participants correctly identified 65% of actions that would be allowed or disallowed on the platform.

This suggests that, although prompts can remind users of the existence of rules, further strategies may be necessary to ensure a deeper understanding and adherence to platform rules and guidelines.

---

[3] 62% in the Reframing prompt arm, 61% in the Relabelling prompt arm, 46% in the Control arm, 44% in the Reframing arm and 47% in the Relabelling arm.

2. Introduction

To read or not to read...

2.1   T&Cs provide the contract between service providers and their users. Social media platforms' T&Cs typically include Community Guidelines. These set out standards for user behaviour and rules about content that cannot be shared.

2.2   T&Cs could achieve several aims:

- By explaining which safety tools and features are available on the service, empower users to make more informed choices about the types of services they use and how they use them.

- By being transparent about the circumstances in which a platform will remove a user's access, reduce unintentional or repeat violations.

- By educating users, encourage them to report content that may violate the T&Cs.

- Enable online communities to increase the accountability of service providers.

2.3   However, the effectiveness of these rules and guidelines to deliver greater safety outcomes is limited if users do not read or understand them. Ofcom research shows that between half and two-thirds of users sign up to online platforms without trying to access or read T&Cs.[4] Other research shows that if users do read T&Cs, they often do not understand them.[5] This lack of understanding is likely due to the complexity and length of the documents. Recent Ofcom research showed that Terms of Service documents of popular video sharing platforms ('VSPs') can take anywhere from eight minutes to over an hour to read, with most requiring advanced readings skills to understand.[6]

2.4   Previous research has demonstrated the potential impact of this lack of engagement with T&Cs and a lack of understanding of appropriate user behaviour. For example, one study found that one in three users who had been sanctioned by X (formerly known as Twitter) were unaware their content was not appropriate to post.[7] Similarly, another study found that over one in three sanctioned Reddit users did not understand why their post had been removed.[8] Both studies also showed that if users were well-informed about the platform rules, they were more likely to perceive the content moderation process as fair, to express positive sentiment towards the platform and were less likely to intend to re-offend in future. However, this research focused on users who had violated a platform's rules. It is unclear if these results would be replicated across a general user base.

2.5   We know that behavioural techniques can be effective in increasing the number of people accessing online T&Cs and their understanding of key information. For example, previous

---

[4] Ofcom, 2024. Terms and conditions and content controls; Ofcom, 2021 Online Nations Report, Page 40.
[5] Whitley, E., and Pujadas, R. 2018. Report on a study of how consumers currently consent to share their financial data with a third party.
[6] Ofcom, 2023. Regulating Video-Sharing Platforms (VSPs).
[7] Katsaros, M., Tyler, T., Kim, J., Meares, T. 2022. Procedural Justice and Self Governance on Twitter: Unpacking the Experience of Rule Breaking on Twitter
[8] Jhaver, S., Appling, D., Gilbert, E., Bruckman, A. 2019. "Did You Suspect the Post Would be Removed?": Understanding User Reactions to Content Removals on Reddit

research by BIT demonstrated that including a 'cost-cue' to inform users how long it would take to read a document doubled the number of people who accessed it. It also found that using icons to illustrate key information increased comprehension of that information by 34%.[9] However, in our analysis of VSP Terms of Service and Community Guidelines, we found that VSPs did not appear to use behavioural techniques to encourage people to access documents or improve their comprehension.[10]

2.6   The research in this report aims to address three evidence gaps. Firstly, much of the previous research focused on how to improve engagement and comprehension of online T&Cs outside of the social media context (e.g., retail shopping platforms). We wanted to understand if behavioural techniques could be effective in improving engagement with social media platforms' T&Cs.

2.7   Secondly, much of the previous research has relied on survey research and self-reported data. Although useful, we wanted to conduct experimental research with behavioural measures that could provide causal evidence about the impact of platform design.

2.8   Finally, to our knowledge, there is limited research that has explored the impact of reading social media T&Cs on subsequent decisions and behaviours while using that service. We wanted to understand if increasing the number of people who accessed social media T&Cs produced any changes in user behaviour and their interactions with content on a platform (e.g., reporting or reposting of content). This could indicate whether accessing T&Cs has an effect on compliance with a platform's rules.

2.9   To address these evidence gaps, we commissioned BIT to design a simulated social media platform that allowed us to test the effectiveness of behavioural techniques in increasing the number of people accessing social media platforms' T&Cs. We also measured whether these interventions had any effect on people's reporting or reposting of content that violated a platform's T&Cs (referred to as 'violative content'). In a social media context, rules for appropriate content and guidelines are often contained within Community Guidelines, and therefore these were the form of T&Cs document used in this research.

---

[9] The Behavioural Insights Team (BIT), 2019. Best practice guide: Improving consumer understanding of contractual terms and privacy policies: evidence-based actions for businesses
[10] Ofcom, 2023. Regulating Video-Sharing Platforms (VSPs)

3. Intervention development

3.1   We took a systematic and evidence-based approach to designing the interventions tested in this trial. Firstly, we used theoretical behavioural models to design and analyse a consumer survey to understand the influences on relevant behaviours. Secondly, we prioritised the strongest influences on these behaviours. Finally, alongside BIT, we used behavioural science frameworks to design interventions targeting the key influences identified.

Understanding why so few users 'read the fine print'

3.2   There is substantial evidence to indicate that users do not typically read online T&Cs.[11] However, there is less research that seeks to understand the reasons behind users' decisions to read or not to read T&Cs. We wanted to analyse the key influences on user behaviour before designing interventions and testing their impact.

3.3   We used the Capability, Opportunity and Motivation ('COM-B') model and Theoretical Domains Framework ('TDF') to design and analyse a consumer survey aimed at testing our assumptions regarding the key influences on two relevant behaviours. Namely: accessing T&Cs whilst signing up to a social media platform and checking the rules while using a platform (see Annex 1 for a full list of our initial assumptions).

3.4   The COM-B model suggests that behaviours are made up of three necessary components: Capability (e.g., physical skill or knowledge), Opportunity (e.g., prompts in the environment, time and resources) and Motivation (e.g., beliefs about what is good and bad, emotional reactions, impulses).[12] The TDF divides these components into subcomponents, enabling a more granular analysis of the main determinants of behaviour.[13]

3.5   The survey included questions relating to the reasons for accessing or not accessing T&Cs before agreeing to them, understanding of T&Cs, and the reasons for any lack of understanding. We also asked about participants' reasons for checking or not checking the rules on a platform, their confidence that their activity did not violate the rules, and the resources they would use if they were unsure about the rules.[14] We found:

- A third of online users said they scan T&Cs for key points but only 8% said they make an effort to read T&Cs fully before agreeing to them.

- In contrast, just over half of users (52%) said they tend to ignore T&Cs when signing up; this was because they think they will take too long to read (65%) or find them overwhelming (45%).

- The main reasons people gave for accessing T&Cs while signing up to a platform were: to help them decide if they were comfortable signing up (54%), to learn more about how their data will be used (51%), or to understand what data will be collected (45%).

[11] Ofcom, 2024. Terms and conditions and content controls; Ofcom, 2021 Online Nations Report, Page 40
[12] Michie, S., van Stralen, M.M. & West, R., 2021. The behaviour change wheel: A new method for characterising and designing behaviour change interventions.
[13] Atkins, L., Francis, J., Islam, R. *et al.* 2017. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems
[14] Ofcom, 2024. Terms and conditions and content controls

3.6   We were also interested in levels of engagement with the rules on what can be posted on social media and VSPs:

- More than a third of users (38%) said they never check the rules about what can be posted on their most-frequently used platform.

- The main reason for not checking platform rules was because users were confident they were not going to do anything that violated the rules (57% of respondents).

- However, two-thirds (66%) of people said they did check the rules at some point while using a platform. For example, when they see content they are unsure of (23%) or when they are prompted to (19%).

- The main reasons people gave for rule-checking were because they thought it was important to use social media in a responsible way (27%) or to make sure they didn't break any rules (26%).

3.7   We used the results of this survey to develop a list of the key behavioural influences we wanted to target with our interventions. See Table 1.

Table 1: A summary of the strongest behavioural influences on users reading T&Cs at sign up and checking social media platform rules

COM-B component	TDF component	Description	Survey evidence[15]
Capability	Awareness	Although most users were aware of T&Cs, many did not list T&Cs or Community Guidelines as places they would check the platform rules.	Only 5% of participants said they had never come across T&Cs *but* 68% of participants did not list either T&Cs or Community Guidelines as places they would check if they were unsure about what could be posted on a platform.
Opportunity	Prompts in the environment	Many users said that it was stimuli from the environment that made them want to check the platform rules.	23% of participants checked platform rules when they saw something they did not think should be there; 19% checked platform rules when prompted by the platform.

---

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

COM-B component	TDF component	Description	Survey evidence[15]
Motivation	Beliefs about consequences	Many users do not think that checking T&Cs or platform rules is valuable to their online experience.	*Of participants that did not read T&Cs at sign up*: 28% thought they all tend to be similar so did not need to read them; 24% trusted that platforms were not allowed to do anything illegal; 21% said it would not affect their decision to use the platform. *Of participants that did not check platform rules*: 57% were confident they were not going to do anything that would violate the rules.

Interventions

3.8 Working with BIT, we used our knowledge of the current literature, our analysis of the consumer survey findings and behavioural science frameworks (e.g., the Behaviour Change Wheel) to design interventions targeting the key influences identified (i.e., awareness, prompts in the environment and beliefs about consequences). We also used the CMA's taxonomy of online choice architecture to inform our thinking. The taxonomy groups commonly observed practices into three categories: choice structure, choice information and choice pressure. Choice structure practices involve changes to the type and order of choices presented to consumers and the effort involved in selecting different options. Choice information practices relate to the type and amount of information provided to consumers and how it is framed. Choice pressure practices use trusted messengers or try to create a sense of urgency to change consumer decisions.[16]

3.9 We used criteria such as impact (i.e., how likely is the intervention to be effective in increasing the number of people accessing Community Guidelines?) and feasibility (i.e., can we test it within our research? Would social media platforms be able to implement it?) to prioritise the intervention ideas. Through this process we decided on three types of interventions to test: Reframing message, Relabelling and Prompts.

Control

3.10 We created a standard design to act as a baseline for comparing the effects of the interventions. The standard design was inspired by current practices of social media platforms where links to T&Cs documents are often displayed as hyperlinks within the sign-up process and users can access Community Guidelines whilst using the platform but are not forced or encouraged to.

3.11 See Figure 1 (below) for screenshots of intervention screens in the five trial arms.

---

[16] CMA, 2022. Online Choice Architecture: How digital design can harm competition and consumers

Reframing message

3.12   Reframing relevant information to alter the way an individual perceives a stimulus or object has been shown to be an effective technique in changing behaviour across a range of contexts.[17] In the CMA's taxonomy, framing (or reframing) is categorised as a 'choice information' technique. Our consumer survey evidence suggests that users do not perceive T&Cs or platform rules as valuable to them or their online experience. To target these influences, we designed a reframing message that highlighted the importance of reading the Community Guidelines from a personal and societal point-of-view in user-friendly language. It also aimed to explicitly target users' potential overconfidence that their activity would not violate the platform rules by encouraging them to check anyway. *"It's always best to check! Following our Community Guidelines helps to keep you and everyone else safe online. Click here to read them."* This intervention targeted the 'motivation' component of behaviour and specifically users' beliefs about the consequences of accessing Community Guidelines.

Relabelling

3.13   Relabelling an object or stimulus is a form of reframing which attempts to change perceptions of the object by giving it a new label. The survey evidence suggested that only 1 in 5 (21%) users listed 'Community Guidelines' as a place they would go if they were unsure about posting something on a social media platform (an additional 1 in 10 (11%) said 'T&Cs'). Therefore, giving 'Community Guidelines' a more user-friendly name (that made their content and purpose clear) could be sufficient to encourage people to access it. We chose the label "Dos and Don'ts" as we considered this to clearly communicate the purpose and contents of the document. This intervention targeted the 'capability' component of behaviour, specifically users' awareness about the purpose of Community Guidelines.

Prompts

3.14   Prompts are cues or reminders designed to encourage individuals to engage in or avoid certain behaviours, at important points of time, and are used in a range of online and offline contexts.[18] In the CMA's taxonomy, prompts are categorised as a 'choice pressure' technique. Previous Ofcom research showed that prompts were effective in increasing reporting of potentially harmful content on a simulated VSP platform.[19]

3.15   To our knowledge, social media platforms often do not encourage users to read T&Cs while signing up, and many of them use 'click-wrap' agreements, where accepting the T&Cs is implicit in the act of signing up.[20]

3.16   Similarly, users are rarely encouraged to read the rules while using a platform. The consumer survey findings suggest that users may be more likely to access platform rules following a relevant stimulus in the online environment (e.g., after viewing distressing content or when prompted by the platform). Providing a prompt to users at a more salient timeline point (e.g., when they had started using the platform and been exposed to content) may encourage people to access Community Guidelines. This intervention targeted the

---

[17] CMA, 2022. Online Choice Architecture: How digital design can harm competition and consumers
[18] CMA, 2022. Online Choice Architecture: How digital design can harm competition and consumers
[19] Ofcom, 2022. Behavioural insights for online safety: understanding the impact of video sharing platform (VSP) design on user behaviour
[20] Ofcom, 2023. Regulating Video-Sharing Platforms (VSPs)

'opportunity' component of behaviour, specifically by providing a salient prompt that gave users a clear opportunity to access Community Guidelines.

Figure 1: Screenshots of the Control and four intervention arms
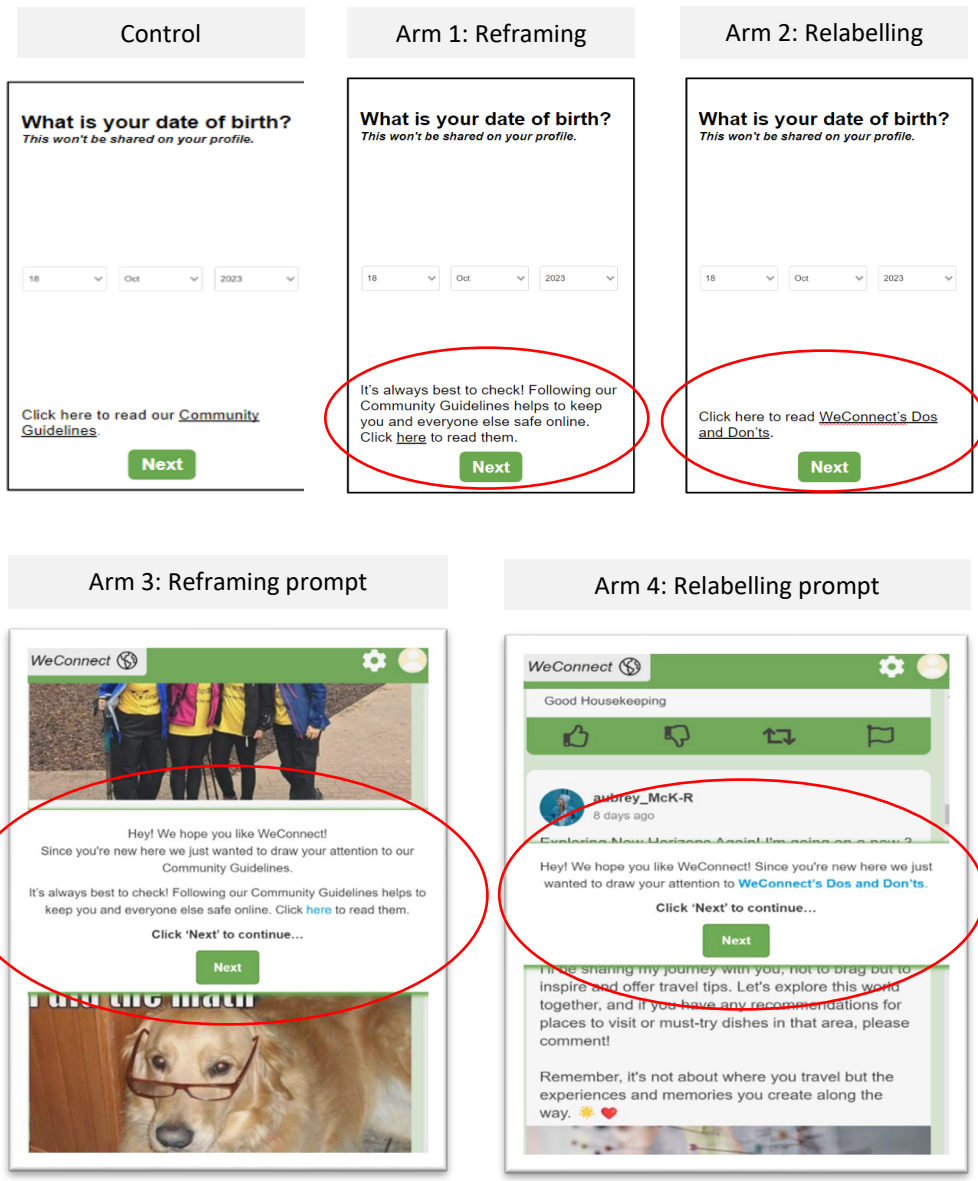
4. Experimental Design

Participant journey

4.1   We tested these interventions on a simulated social media platform called WeConnect, which closely mimicked real social media platforms. Our study included a nationally representative sample of 3,500 adults in the UK during January 2024. Participants were randomly assigned to one of the trial arms described above.

4.2   Participants' journey included the following key components:

- Sign-up to WeConnect: Participants made the decision to click through and sign-up to the platform. This is where the interventions described in the previous section were incorporated.

- Browsing the feed: The feed consisted of 24 pieces of content, including short videos and text posts, some accompanied by images. One-third of the feed contained content deemed to violate WeConnect's rules (such as hate, violence, and misinformation), while the remaining two-thirds were neutral. Participants could engage with the posts by liking, disliking, reposting, or reporting them.

- Prompts: In two of the intervention arms (Reframing prompt arm and Relabelling prompt arm), participants received prompts encouraging them to read the Community Guidelines after browsing a quarter of the content.

- Follow-up questionnaire: After participants browsed through the feed, we asked them questions to better understand their behaviour and comprehension of WeConnect's Community Guidelines. This included: their reasoning for clicking or not clicking to read the Community Guidelines, their recall of the platform's rules, their understanding of what was allowed or not allowed according to those rules, their perceptions of platform safety, and their attitude towards the prompt (only for participants in the prompt arms).

4.3   More details about the participants' journey through the experiment can be found in the Technical Report. See Annex 2 for WeConnect's Community Guidelines.

Outcome measures

4.4   The primary outcome of interest in this study was the percentage of participants who clicked-through to read the Community Guidelines. This could occur either during the sign-up stage, while browsing the feed (through a 'gear' icon), or in response to a prompt (for participants in the prompt arms).

4.5   Our hypothesis was that all our interventions would encourage participants to actively engage with the Community Guidelines. We anticipated an increase in the percentage of people clicking through to read the Community Guidelines compared to the Control arm.

4.6   As secondary outcome measures, we examined the percentage of participants who reported a violative post or reposted such content while browsing the feed. For a comprehensive set of hypotheses, please refer to the Technical Report.

5. Key Findings

Prompts increased accessing Community Guidelines

5.1   In the Control group, 3% of participants clicked through to read the Community Guidelines. This figure increased to 7% in the Reframing prompt arm, marking a statistically significant increase. The effect of the prompt was even more pronounced in the Relabelling prompt arm, where 9% of participants clicked through to read them. Figure 2 shows a comparison of the intervention arms compared to the control.

Figure 1: The percentage of people who clicked to read the Community Guidelines (Control arm compared to other study arms)
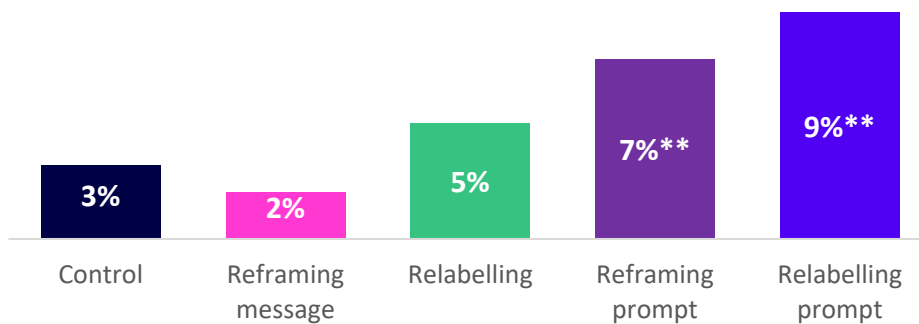
*Note: ** statistically significant at the 1% level (p<0.01) in comparison to the Control arm.*

5.2   Moreover, participants who received prompts while browsing the feed were significantly more likely to click through to the Community Guidelines, compared to those who encountered the same message only at the sign-up stage. Specifically, receiving a mid-feed prompt increased the proportion of people clicking through from 2% to 7% in the Reframing arms, and from 5% to 9% in the Relabelling arms. See Figure 3 for this comparison.

Figure 2: The percentage of people who clicked to read the Community Guidelines (intervention arms without prompts compared to intervention arms with prompts)
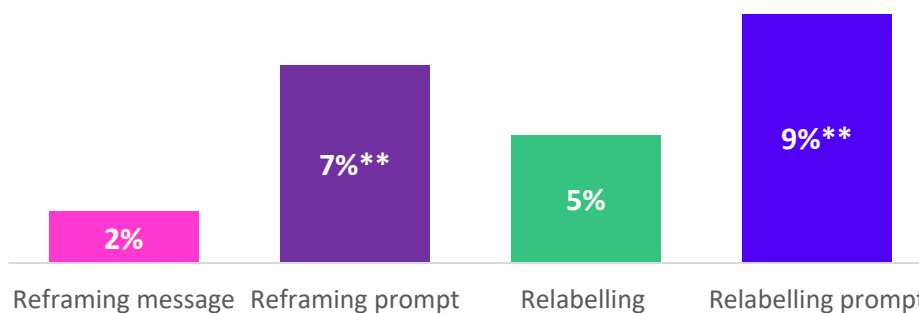
*Note: ** statistically significant at the 1% level (p<0.01) prompt arms compared to the 'without prompt' arms.*

5.3   The findings indicate a statistically significant increase in clicking through to read the Community Guidelines when prompts were employed. This is in line with our expectations

and the broader behavioural evidence demonstrating the effectiveness of prompts in shaping people's behaviour.[21]

5.4   Overall, only 5% participants, across all the trial arms, clicked to access the Community Guidelines. The main reasons participants clicked to read the Community Guidelines were to ensure compliance with the rules (55%) and to find more information about the platform (53%).[22] This suggests that users who clicked to read the Community Guidelines demonstrated an interest in understanding and adhering to platform rules. See Table 2 for the top three reasons participants gave for clicking to read the Community Guidelines.[23]

Table 2: Top three reasons why participants clicked to read WeConnect's Community Guidelines

Reason	%
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%

Note: Participants could select more than one option (n=184)

5.5   Conversely, across the trial, 95% of participants did not click to read the Community Guidelines. The top reasons were that people assumed the rules were similar to other platforms' (36%), people not seeing the link to the Community Guidelines (32%) or not realising they could click on it (30%). These insights highlight the range of factors that are influencing users' engagement with Community Guidelines. See Table 3 for the top three reasons participants gave for not clicking to read the Community Guidelines.[24]

Table 3: Top three reasons why participants did not click to read WeConnect's Community Guidelines

Reason	%
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%

Note: Participants could select more than one option (n=3,331)

---

[21] CMA, 2022. Online Choice Architecture: How digital design can harm competition and consumers; Ofcom, 2022. Behavioural insights for online safety: understanding the impact of video sharing platform (VSP) design on user behaviour
[22] Participants could select more than one option (n=184).
[23] A full breakdown of response options can be found in Table 8 of the Technical Report.
[24] A full breakdown of response options can be found in Table 9 of the Technical Report.
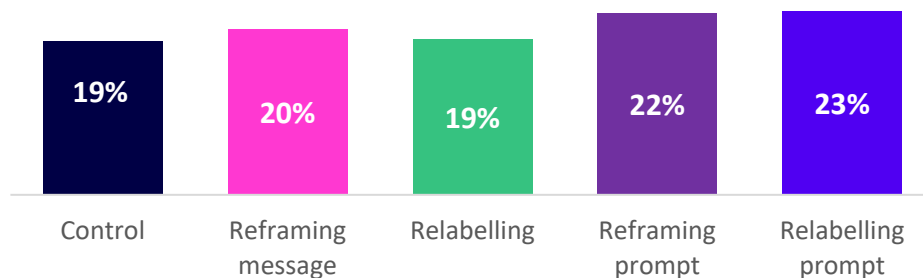
The interventions did not change reporting or reposting behaviour

5.6   Despite the statistically significant increase in clicking through to read the Community Guidelines in two of the intervention arms, none of the interventions significantly impacted reporting or reposting behaviour on the platform. Importantly, we could only measure the proportion of participants that clicked to access the Community Guidelines, but not the extent to which they read them.

Reporting

5.7   In the Control arm, 19% of participants reported at least one violative post. This rate was slightly higher for participants in the treatment arms, ranging from 19% to 23%, but none of the differences were statistically significant compared to the Control arm. Participants who received prompts were as likely to report violative posts as those who encountered the message at the sign-up stage only. See Figure 4 for a full comparison.

Figure 3: The percentage of people who reported at least one violative post (Control arm compared to the other study arms)

|				
19%	20%	19%	22%	23%
Control	Reframing message	Relabelling	Reframing prompt	Relabelling prompt

Reposting

5.8   Similarly, in terms of reposting behaviour, there were no significant differences between the Control arm and any of the treatment arms. In the Control arm, 5% of participants reposted at least one violative post. There was no statistically significant difference for participants in the treatment arms. Participants who received prompts were as likely to repost violative posts as those who encountered the message at the sign-up stage only. See Figure 5 for a full comparison.

5.9   Additionally, we did not find any statistically significant differences in the number of violative posts reposted across all study groups. On average, participants reposted a similar number of violative posts.[25]

---

[25] 3,326 out of 3,515 participants (95%) did not repost any violative content; see Table 18 and Table 19 in the Technical Report for the regression results.

Figure 4: The percentage of people who reposted at least one violative post (Control arm compared to the other study arms)

5%	5%	5%	6%	6%
Control	Reframing message	Relabelling	Reframing prompt	Relabelling prompt
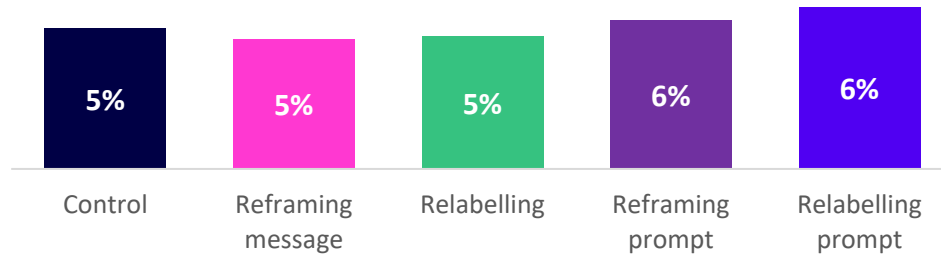
Prompts boosted recall of Community Guidelines

5.10   There was a statistically significant increase in the number of participants correctly recalling the existence of rules and guidelines on WeConnect in the prompt arms compared to the Control arm. Specifically, participants in the Reframing prompt arm were 16 percentage points more likely to recall the platform had rules, while those in the Relabelling prompt arm were 15 percentage points more likely to do so, compared to participants in the Control arm (where 46% correctly recalled the rules).[26] See Figure 6 for a full comparison.

Figure 5: The percentage of participants who correctly recalled that WeConnect had rules or guidelines (Control arm compared to other study arms)


46%	44%	47%	62%**	61%**
Control	Reframing message	Relabelling	Reframing prompt	Relabelling prompt
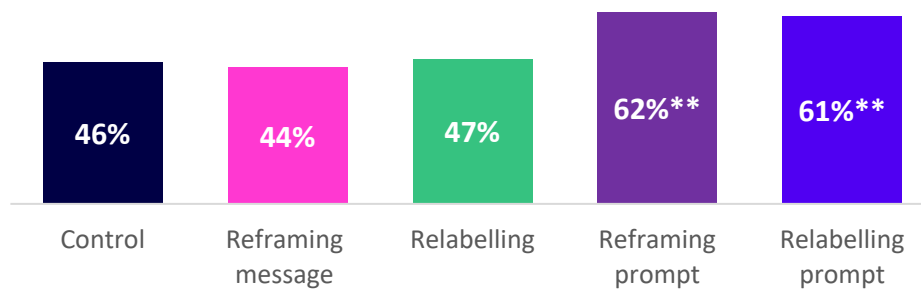
*Note: ** statistically significant at the 1% level (p<0.01) in comparison to the Control arm.*

5.11   However, this means that nearly 40% of participants who had been exposed to a prompt still did not correctly recall the platform had rules and guidelines.

5.12   Despite the increase in recall among participants in the prompt arms, their comprehension of rules remained unaffected. Participants were presented with six online behaviours and asked if these behaviours were allowed or not allowed according to the platform's rules. There was no statistically significant difference in terms of the number of questions answered correctly between any of the study arms. See Figure 7 for a full comparison.

---

[26] The recall rate for participants who clicked to read the Community Guidelines was substantially higher (85%) than for those who did not click (51%), indicating a potential mechanism for recall when prompts were employed.

Figure 6: The percentage of correct answers given to the comprehension questions (Control arm compared to other study arms)
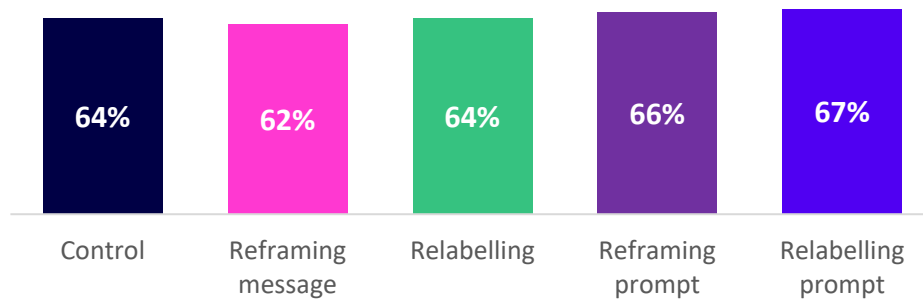
6. Discussion

6.1   We ran an online RCT on a simulated social media platform to test the effectiveness of behavioural techniques in increasing the number of users who access Community Guidelines. We were also interested in whether reading Community Guidelines had any effect on users reporting or reposting content while using the platform.

Prompting was the only effective way to encourage users to click to read the Community Guidelines

6.2   Overall, participants who were exposed to a prompt were significantly more likely to click on the Community Guidelines than those who saw a message about Community Guidelines at the sign-up stage only. This suggests that encouraging users to engage with important T&Cs documents at other points in the user journey, outside of the sign-up stage, could increase the number of people accessing and reading them.

6.3   However, the effect sizes we observed were smaller than in other studies. For example, another Ofcom online trial found that over 20% of participants checked their content controls when prompted, compared to 4% without a prompt.[27] One explanation could be that although prompting helps to address important opportunity barriers, users' motivation to engage with T&Cs is still very low and difficult to target. Therefore, even in a carefully designed online environment, there is likely to be an upper limit on user engagement with T&Cs.

6.4   Reframing or relabelling on their own were not as effective. Participants in the Reframing and Relabelling arms were no more likely than those in the Control arm to click to read the Community Guidelines, with only 2% and 5% making this decision, respectively. This further supports the idea that users' motivation to engage with online T&Cs is difficult to address and might act as an ongoing barrier for improved engagement. For instance, it appears that the reframing message used in this experiment may not have been sufficiently powerful to significantly increase participants' motivation to access the Community Guidelines. Additionally, Ofcom research shows that only 1 in 4 adult social media users said they often post, share or comment on the platforms they use.[28] The users who do not regularly post or share on platforms may choose not to read T&Cs because they believe they are highly unlikely to violate the rules.

6.5   We note that overall in the experiment, only 5% of participants clicked to read the platform's Community Guidelines. This finding contrasts with Ofcom's consumer survey in which 39% of users claimed to fully or skim read T&Cs and 66% said they check platform rules at some point in their user journey.[29] The discrepancy may be attributed to poor recall or 'social desirability' bias on the part of the respondent, i.e., participants reporting how they would like to or think they should behave rather than what they *actually* do. This

---

[27] Ofcom, 2024. Behavioural insights to empower social media users
[28] Ofcom, 2023. Adults' Online Behaviours and Attitudes 2023 Wave 1 and Wave 2 Combined Data Tables (Table 30)
[29] Ofcom, 2024. Terms and conditions and content controls

difference between survey and experimental results highlights a strength of conducting experimental research in the context of online safety.[30]

6.6   Our findings suggest that prompts delivered 'in-context' (i.e., when the behaviour being encouraged is more relevant) may be more effective than prompts at less relevant time points in the user journey. For example, prompts to engage with Community Guidelines could be more effective just after a user has encountered content they found distressing. Further research could investigate the optimal use of prompts to encourage online safety behaviours.

6.7   Additionally, services might need to consider employing a broad range of behavioural techniques to encourage users to access their T&Cs. Future research could look into further testing the effectiveness of 'choice structure' (e.g., using visual stimuli) or 'choice pressure'(e.g., commitment mechanisms) online choice architecture practices.[31]

The interventions did not influence user reporting or reposting of violative content

6.8   Our research suggests increasing the accessing of Community Guidelines, by itself, was not sufficient to change some online behaviours (e.g., reporting or reposting of violative content).

6.9   One explanation is that understanding of the rules on a platform is not the strongest determinant of whether a user reports or reposts violative content. There could be other factors having a stronger influence (e.g., belief in effectiveness of reporting or social norms regarding reporting of content).

6.10   Previous research had suggested those with a better understanding of the rules on a platform were less likely to say that they intended to violate platform rules in future.[32] However, this research used self-reported data and focused on uploading content (rather than reporting or reposting) with a population of users who had previously been sanctioned by a platform. Our findings suggest that accessing platform rules is not sufficient to change the reporting or reposting behaviour in a general (i.e., non-sanctioned) population of users. Future research could explore the effects of accessing platform rules on other relevant behaviours such as uploading content.

6.11   Our findings also suggest that it could be important for platforms to consider other methods of encouraging compliance with T&Cs, beyond just increasing engagement. For example, providing reminders of key information at the point of potential rule violation (such as uploading or reposting content).

6.12   There are some important caveats to note when interpreting these findings. Firstly, we do not have a baseline comparison of how many people usually report or repost content that

---

[30] We recognise that the duration of our experiment was approximately ten minutes and therefore only provides a snapshot of the total user journey and the opportunities users might have to engage with T&Cs on a real-world platform.

[31] CMA, 2022. Online Choice Architecture: How digital design can harm competition and consumers

[32] Katsaros, M., Tyler, T., Kim, J., Meares, T. 2022. Procedural Justice and Self Governance on Twitter: Unpacking the Experience of Rule Breaking on Twitter; Jhaver, S., Appling, D., Gilbert, E., Bruckman, A. 2019. "Did You Suspect the Post Would be Removed?": Understanding User Reactions to Content Removals on Reddit

violates social media T&Cs. In a previous Ofcom trial, using a simulated VSP, only 1% of participants reported potentially harmful content in the Control condition. This rose to 11% when participants received a prompt encouraging them to report.[33] This is substantially lower than the 19% to 23% of participants who reported at least one violative post across the trial arms in this research. These differences highlight the importance of context in reporting behaviour.

6.13 Secondly, we also do not know the extent to which participants actually read the Community Guidelines after they clicked to access them. Additionally, the messages within the guidelines that discouraged violative behaviour (i.e., "If you see content that violates our Community Guidelines, please report it and do not re-post or otherwise interact with it") may not in themselves have been prominent enough to prompt increased compliance, especially considering the baseline reporting rate in the experiment (see WeConnect's Community Guidelines in Annex 2).

6.14 Finally, reporting and reposting were secondary outcome measures in our trial and, therefore, we cannot attribute causality between interventions and observed outcomes.

Prompts increased recall of the platform rules, but not comprehension

6.15 We found that prompts were effective in improving the correct recall that WeConnect had rules and guidelines. This indicates that prompts can be a useful reminder about the existence of rules on a platform.

6.16 At the same time, nearly 40% of participants did not correctly recall the information included in the prompt and 32% of users, who did not click on the link, did not recall seeing the prompt in the first place. This could be due to users not reading the information included in the prompt or quickly forgetting it. Another explanation could be that users experience 'alert fatigue' from the constant use of prompts on these types of platforms, leading them to close pop-up screens without reading them. This suggests the use of prompts needs to be selective and targeted to minimise this risk. As indicated above, the effectiveness of prompts is likely to be heavily context dependent and influenced by factors such as previous exposure, the type of behaviour being encouraged and user base. Therefore, it is important for online platforms to conduct their own research to establish when, where and how prompts are most effective in encouraging safer online behaviours.

6.17 Additionally, while prompts effectively increased recall of Community Guidelines, this did not translate into improved comprehension. This suggests that although prompts can remind users of the existence of rules, further strategies may be necessary to ensure a deeper understanding and adherence to platform's rules and guidelines.

6.18 It is possible that participants' existing familiarity with typical social media platform rules allowed them to have a good understanding of behavioural standards and make informed decisions about the content they encounter. However, if a platform has standards or rules which differ from others, they may need to pro-actively draw this to users' attention.

---

[33] Ofcom, 2022. [Behavioural insights for online safety: understanding the impact of video sharing platform (VSP) design on user behaviour](#)

Limitations

6.19   Our research has limitations which should be considered when interpreting the findings. Firstly, 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 be more harmful and personalised than the content shown in our research. This could result in different levels of interaction than those observed in our study.

6.20   Secondly, the short timescale at which our online experiment had to measure outcomes limits the conclusions that can be drawn regarding the long-term effects of our interventions.

6.21   Thirdly, 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.

6.22   Despite these limitations, we believe online RCTs are a useful tool for building the evidence base regarding how users behave online.

A1. Key influences on relevant behaviour

We used our knowledge of the existing literature to develop a set of assumptions regarding the key influences on the two relevant behaviours: users reading T&Cs at sign up and checking social media platform rules. We used the consumer survey to test these assumptions and prioritise the influences to target with our intervention.
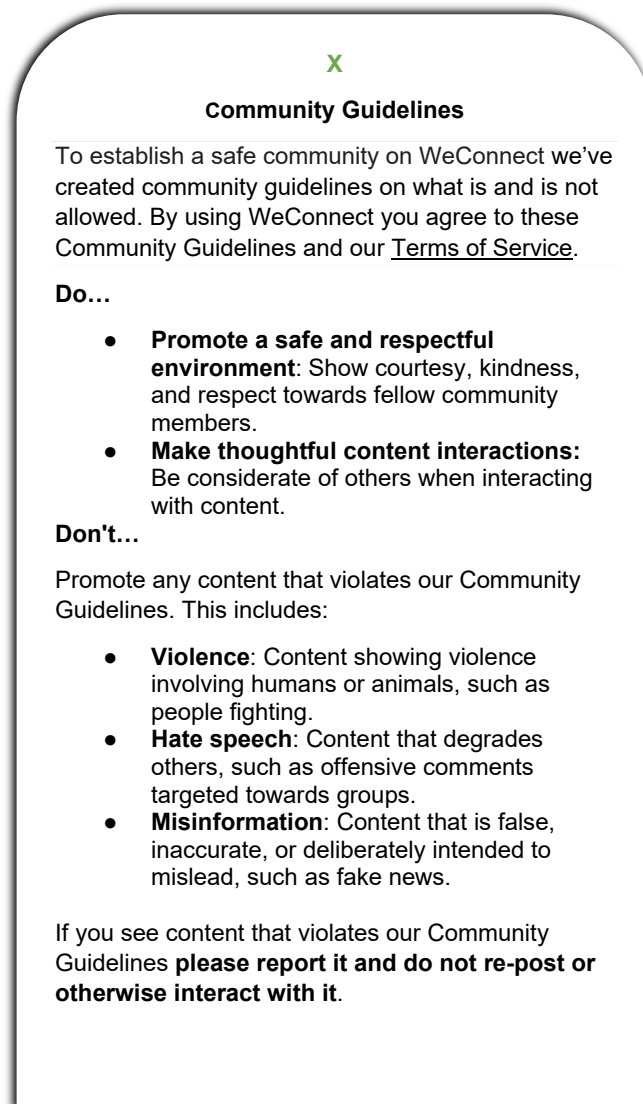
Figure 8: A full list of the COM-B and TDF influences on reading T&Cs and actively checking social media platform rules

Capability *(Psychological, Physical)*				
Cognitive skills	**Awareness**	**Attention**	**Evaluating options**	**Memory**
Users need to know where/how to find the T&Cs/ platform's rules.	Users need to be aware that platforms have T&Cs/rules, why they are important and where they can find them.	Users need to sustain attention long enough to check T&Cs/platform rules, in an environment with a lot of other attentional cues.	There are lots of different policy documents that a user might need to check in order to familiarise themselves with the T&Cs/rules. These also differ across platforms.	Users need to remember how to familiarise themselves with T&Cs/platform rules.

Opportunity *(Physical, Social)*			
Prompts in the environment	**Resources & time**	**Opportunities in the environment**	**Social norms**
Users are often not encouraged to familiarise themselves will T&Cs/platform rules (e.g., through the use of 'click wrap' agreements)	The resources provided to users to help them check the T&Cs/rules are often inadequate and do not facilitate user understanding.	There are opportunities to check T&Cs/platform rules but the design of platforms often does not facilitate this behaviour.	There isn't a strong culture of reading T&Cs or active rule checking within social media users.

Motivation *(Reflective, Automatic)*			
Beliefs about consequences	**Identity**	**Emotions**	**Habits**
Users might believe there are no direct consequences of not reading T&Cs/platform rules.	Users might not see themselves as people who consistently and actively read T&Cs/ rules on social media.	Social media can be a highly emotive environment which does not facilitate reading T&Cs/platform rules.	Many users will not be in the habit of checking T&Cs/ platform rules when they are unclear.

See slide 51 for relevant literature

23

A2. WeConnect's Community Guidelines

Figure 9: A screenshot of WeConnect's Community Guidelines (named "Dos and Don'ts" in the Relabelling and Relabelling prompt arms)

X

Community Guidelines

To establish a safe community on WeConnect we've created community guidelines on what is and is not allowed. By using WeConnect you agree to these Community Guidelines and our Terms of Service.

Do…

- **Promote a safe and respectful environment**: Show courtesy, kindness, and respect towards fellow community members.
- **Make thoughtful content interactions:** Be considerate of others when interacting with content.

Don't…

Promote any content that violates our Community Guidelines. This includes:

- **Violence**: Content showing violence involving humans or animals, such as people fighting.
- **Hate speech**: Content that degrades others, such as offensive comments targeted towards groups.
- **Misinformation**: Content that is false, inaccurate, or deliberately intended to mislead, such as fake news.

If you see content that violates our Community Guidelines **please report it and do not re-post or otherwise interact with it**.