

Behavioural insights to empower social media users

Testing tools to help users control what
they see

Behavioural Insights Discussion Paper

Published 21 May 2024



The discussion paper series

Ofcom is committed to encouraging debate on all aspects of media and communications regulation and to create rigorous evidence to inform that debate. One of the ways we do this is through publishing a series of discussion papers, extending across behavioural insights, economics and other disciplines. The research aims to make substantial contributions to our knowledge and to generate a wider debate on the themes covered.

Acknowledgements

Ofcom would like to thank the Behavioural Insights Team for their work in helping us develop and run the online randomised controlled trials, and for the detailed analysis set out in the Technical Reports.

Disclaimer

Discussion papers contribute to the work of Ofcom by providing rigorous research and encouraging debate in areas of Ofcom's remit. Discussion papers are one source that Ofcom may refer to, and use to inform its views, in discharging its statutory functions. However, they do not necessarily represent the concluded position of Ofcom on particular matters.

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

Section

Overview	4
1. Introduction	7
2. Sign-up Trial Interventions	10
3. Sign-up Trial Experiment Design.....	14
4. Sign-up Trial Findings	15
5. Check and Update Trial Interventions	19
6. Check and Update Trial Experiment Design	22
7. Check and Update Trial Findings	23
8. Discussion and Conclusion	26

Annex

A1	Long List of Barriers and Prioritisation	28
A2	Check and Update Trial: Initial Message Ideas	30

Overview

Social media provides many benefits, but almost 3 in 10 adults (27%) say they [have recently been exposed](#) to potentially harmful content.¹ Policy responses to empower users include improving media literacy so that users know how to take control of their online experience, and ensuring user-empowerment tools are effective and meet user expectations.

Content controls, which allow users to choose whether to reduce the amount of sensitive content they see, are one important tool offered by social media platforms. However, only 26% of people say they [have ever used](#) them.² Reasons not to use controls, even if aware of them, include not having enough time (26% of those aware but not using controls), controls being too complex to understand (14%) and not being able to find them (10%).

At the same time, there is a growing body of evidence demonstrating that the way platforms design and present their services ([‘the choice architecture’](#)) shapes how users respond.³ Ofcom’s Behavioural Insight specialists partnered with the Behavioural Insights Team (BIT) to run two experiments to build evidence on how platform choice architecture affects engagement with content controls among adult users. We used a mock-up of a typical social media platform, called WeConnect, to run two online randomised controlled trials, targeting different stages in the user journey.

- **The Sign-up trial varied how information and choices related to content controls are presented to users when they set up a new social media account.** The behavioural insight literature demonstrates that behaviour is particularly open to [change at key moments](#), such as first engagement.⁴ Equally, the UK Online Safety Act 2023 (‘the Act’) identifies the ‘earliest possible opportunity’ for users to make a choice on content controls as a key moment for platforms to empower users.
- **The Check and Update trial tested the effect of prompting users who have already set up an account to check and update their content controls while browsing.** This trial examined a different but overlapping challenge – how to encourage users to engage with content controls when they are browsing and are not obliged to do so.

In the **Sign-up trial**, users were given the choice between seeing “All content types” and “Reduced sensitive content”. Figure 1 provides an overview of how the information and choice options were presented in different trial arms. We tested the extent to which user choice was shaped by two types of behavioural interventions:

1. **The platform providing a pre-selected option (‘Default’).** This is a widespread practice on social media platforms.

¹ Ofcom, 2024. [Terms and conditions and content controls](#). Participants were asked about seeing potentially harmful content in the past 3 months on social media or VSPs, and provided the following examples “content related to violence, abuse, hatred, self-harm, unhealthy diets or eating disorders, or any other content that can be considered offensive, inappropriate, and cause serious distress”.

² Ofcom, 2024. [Terms and conditions and content controls](#).

³ CMA, 2022. [Online Choice Architecture: How digital design can harm competition and consumers](#).

⁴ BIT, 2014. [EAST Four simple ways to apply behavioural insights](#) [accessed January 25, 2024].

2. **The way information about the options is presented to users** by i) varying its prominence ('Information saliency' or 'Info saliency'); and ii) presenting it via a short tutorial, with or without an option to skip it ('Skippable microtutorial' and 'Non-skippable microtutorial').

Figure 1: Sign-up trial interventions overview


Basic presentation (Control)	Default	Info saliency	Microtutorial (Skippable or Non-skippable)
<p>Choose how much sensitive content appears in your own feed.</p> <p><input type="radio"/> All content types You may see some posts with sensitive content</p> <p><input type="radio"/> Reduced sensitive content You will see fewer posts with sensitive content</p> <p>Sensitive content doesn't go against our Community Guidelines, but refers to topics some people don't want to see. Learn more.</p> <p>Sensitive content examples hidden Next</p>	<p>Choose how much sensitive content appears in your own feed.</p> <p>Option pre-selected</p> <p><input checked="" type="radio"/> All content types You may see some posts with sensitive content</p> <p><input type="radio"/> Reduced sensitive content You will see fewer posts with sensitive content</p> <p>Sensitive content doesn't go against our Community Guidelines, but refers to topics some people don't want to see. Learn more.</p> <p>Next</p>	<p>Choose how much sensitive content appears in your own feed.</p> <p><input type="radio"/> All content types You may see some posts with sensitive content</p> <p><input type="radio"/> Reduced sensitive content You will see fewer posts with sensitive content</p> <p>Sensitive content doesn't go against our Community Guidelines, but refers to topics some people don't want to see, such as:</p> <ul style="list-style-type: none"> • 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 labeled as false or partly false by impartial third-party fact-checkers such as false news. <p>Sensitive content examples salient Next</p>	<p>Sensitive content examples presented via a brief training module.</p> <p>Click on each category to see some examples... then the "Next" button will appear.</p> <ul style="list-style-type: none"> • Violence: Content showing violence involving humans or animals, such as people fighting. <p>Hate speech</p> <p>Skip tutorial</p>

A concern with online choices is that they might be made swiftly, without adequate attention, as users click through to reach the content they are looking for. To assess whether users were satisfied with their initial choice, we asked them to review their choice after a period of browsing and decide if they wanted to change it. A high degree of change could indicate that the choice architecture at account set-up is distorting users' initial choices.

In the **Check and Update trial**, we explored whether prompts could encourage users to check and update their content settings while browsing. Figure 2 provides an overview of the messages tested. We tested how the following factors influence the effectiveness of prompts.

1. **Different motivations.** Prompts either emphasised that users could take control of their feed ('Empowerment') or highlighted the ease of updating settings ('Process').
2. **The timing of prompts.** Users were prompted either at the start of their browsing ('Pre-engagement') or after they had disliked a sensitive post ('Post-engagement').⁵ We tested whether users were more likely to check settings when they were encountering sensitive content and might be experiencing a reaction against it.

Figure 2: Check and Update trial interventions overview

No prompt (control)	Pre-engagement (start of the feed) Message: Empowerment	Post-engagement (after dislike a sensitive post) Message: Empowerment	Pre-engagement (start of the feed) Message: Process	Post-engagement (after dislike a sensitive post) Message: Process
<p>Need to click on </p>	<p>Your feed, your choice – you can choose the amount of sensitive content that you see.</p>	<p>We noticed you just disliked a post. Your feed, your choice – you can choose the amount of sensitive content that you see.</p>	<p>It takes just two steps to check and update your content settings.</p>	<p>We noticed you just disliked a post. It takes just two steps to check and update your content settings.</p>

⁵ Some users did not dislike any sensitive posts and received a prompt after the last sensitive post.

What we found

Sign-up trial

- **When “All content types” was pre-selected at sign-up, only 15% opted for “Reduced sensitive content”.** The figure was 24% with no pre-selected option.
- **Presenting information with examples of sensitive content on the decision page (‘Info saliency’) increased the selection of “Reduced sensitive content” at sign-up by 5 percentage points, to 29%.** A small increase in salience on what is considered sensitive content significantly increases the number of users who choose to avoid it.
- **Users have a strong tendency to continue with their initial choice – 88% of participants did not change setting after seeing the feed.** This ‘stickiness’ of initial choice was observed regardless of whether the initial choice was shaped by a default, and whether users had selected “All content” or “Reduced sensitive content”.

Check and update trial

- **Prompts can be effective in encouraging users to check settings:** without a prompt, only 4% of participants checked their content settings. 17%-23% checked when prompted.
- **The initial setting was ‘sticky’.** All participants started with “All content types”. Only 7% to 13% of users prompted to check their settings finished with “Reduced sensitive content”.
- **Prompts emphasising the ease of changing settings proved more effective in encouraging users to check content settings than prompts aiming to provide a sense of control (23% vs 17% checked).**⁶ Users seem to have an expectation that changing controls is burdensome and respond to information debunking that concern.⁷
- **Prompts after engagement with sensitive content encouraged more participants to check their settings than prompts before engagement (21% vs 18%).** Responsiveness to prompts can be increased by linking prompt timing to interaction with content and a user’s potentially heightened emotional state when they see sensitive content.

Conclusion: User decisions on content controls are heavily susceptible to the way that choice is presented. Defaults, salience, prompts, timing, and motivational messages all shaped user choices. Yet across both trials, most users stuck with the initial setting, even when provided with low-effort opportunities to revise it.

This points to two lessons for media literacy and user empowerment. Firstly, the initial choice at account set-up is important. Efforts to ensure this choice is informed and active will be particularly worthwhile. Secondly, however, users’ willingness to stick with the initial setting suggests that they are not motivated to optimise their content controls (at least not via the binary options offered in these trials – they were not offered fine-grained choices about which content they did and did not want to avoid). This may indicate that users have relatively weak preferences about their content settings and prefer to manage their content directly, by skipping or blocking content or providers, for example.

⁶ Our study did not test the effect of multiple prompts. To avoid overwhelming users, it is advisable for platforms to consider the number of prompts. Also, note that the percentage of participants who checked their settings was 21.45% for the Post-engagement prompt and 22.51% for the Process prompt. These are rounded to 21% and 23% in this report and to 21.5% and 22.5% in the Technical Report.

⁷ It is important that platform design ensures accessing content controls is easy, without too many steps.

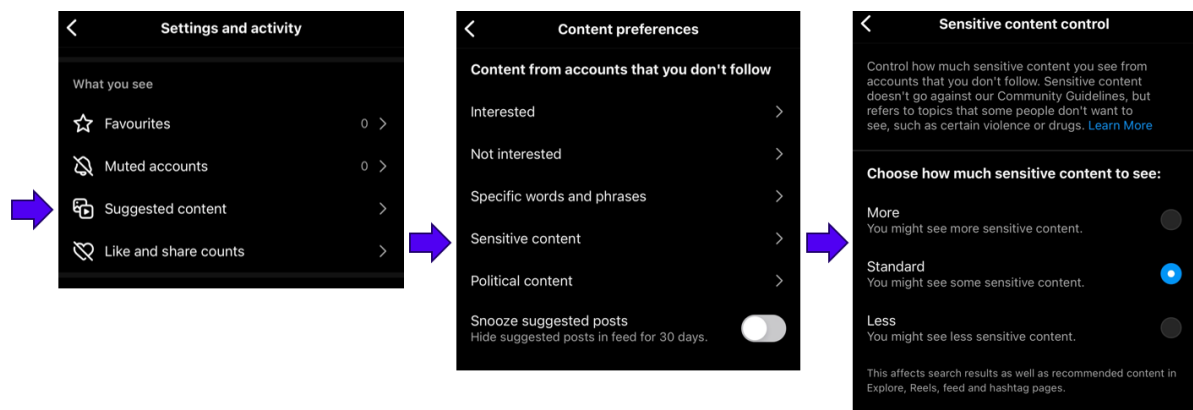
1. Introduction

Enabling social media users to control their feed

In the UK, almost everyone is online. Social media is used by [89% of adults](#), and has many benefits but may also be a source of negative experiences and harm.⁸ A third of users (32%) believe that people are cruel or unkind to one another on online communications platforms all or most of the time.⁹ Moreover, almost 3 in 10 adults (27%) say they [have recently been exposed](#) to potentially harmful content on social media and video-sharing platforms (VSP).¹⁰

Many social media platforms offer users tools to control what appears in their feed and to avoid encountering sensitive content (referred to as ‘content controls’ or ‘content settings’ in this paper). These include blocking content from certain accounts, muting words or hashtags, and indicating that you are not interested in seeing similar content. See Figure 3 for an example of typical content controls on a social media platform.

Figure 3: Example of existing content controls



But [only about a quarter](#) of social media users say they have ever used content controls.¹¹ Lack of awareness is one of the barriers, but almost half of social media users say they are aware of content controls but have never used them. Reasons for this include not having time (26% of those aware but not using controls), controls being too complex to understand (14%) and not being able to find them (10%).

There are many responses to this. Ofcom, with the support and engagement of the academics, platforms and [one of our media literacy external working groups](#), has created a suite of [Best Practice Design Principles for Media Literacy](#). The principles have been created to encourage platforms to incorporate media literacy considerations into the different stages of their work.

⁸ Ofcom, 2024. [Adults' Media Use and Attitudes Report](#).

⁹ Ofcom, 2024. [Adults' Media Use and Attitudes Report](#).

¹⁰ Ofcom, 2024. [Terms and conditions and content controls](#). Participants were asked about seeing potentially harmful content in the past 3 months on social media or VSPs, and provided the following examples “content related to violence, abuse, hatred, self-harm, unhealthy diets or eating disorders, or any other content that can be considered offensive, inappropriate, and cause serious distress”.

¹¹ Ofcom, 2024. [Terms and conditions and content controls](#).

Similarly, the Act recognises the value of content controls in enabling users to manage their exposure to certain types of content. Under the Act, certain services¹² will have to provide adult users with tools to control what content they see, and these will have to be easy to access, and offered at the earliest possible opportunity.¹³

Behavioural insights and online decision-making

The behavioural insight literature shows that access to information is not enough for active, informed choice.¹⁴ Users need to understand the consequences of each option, reflect on what is right for them, and have the opportunity and motivation to take appropriate action.

Moreover, users do not make decisions in a neutral environment, and small changes in [choice architecture](#) can have an impact.¹⁵ For example, defaults which do not match user preferences, complicated layout, and use of technical language can hinder users' ability to effectively engage with online controls.^{16, 17}

To build evidence on how choice architecture influences users' ability to make an informed, active choice about their social media content, Ofcom's Behavioural Insight specialists partnered with the Behavioural Insights Team (BIT) to run two randomised controlled trials (RCTs). We offered users options to choose the amount of sensitive content they received on a simulated social media platform. Importantly, our aim was not to steer users towards a particular choice. We did not, for example, aim to increase the number of users who reduce the amount of sensitive content on their feed. Rather, the goal was to empower users to make a choice that is right for them, reflecting their own preferences for content.

The trials focused on two different stages of the user journey.

The Sign-up trial focused on the earliest choice users make – when they sign up to a platform.

As mentioned above, under the Act, category 1 services have a duty to offer adult users features to control their engagement with certain types of content *at the earliest possible opportunity*.¹⁸ This stage is important from a behavioural perspective for two reasons. Firstly, people are more open to developing new habits and behaviours when they are doing something new.¹⁹ Secondly, the choices that users make may have a long-lasting impact on their experience. Evidence on what supports or hinders active informed decision-making at this stage is particularly important to understanding how platforms can empower users.

In this trial, we explored the impact of two types of behavioural intervention: pre-selecting one setting (Default) and varying how information is presented to users. The latter included presenting information about sensitive content examples on the initial choice page (Info saliency), or in small chunks via a short tutorial, with or without an option to skip (Skippable and

¹² Category 1 services that meet thresholds set out in secondary legislation.

¹³ Sections 15(3) to (5) of the Act.

¹⁴ CMA, 2022. [Online Choice Architecture: How digital design can harm competition and consumers](#) [accessed April 4, 2024].

¹⁵ CMA, 2022. [Online Choice Architecture: How digital design can harm competition and consumers](#).

¹⁶ BIT, 2020. Active Online Choices: [Summary of desk research](#) [accessed April 4, 2024].

¹⁷ CDEI, 2020. [Online targeting: Final report and recommendations](#) [accessed April 4, 2024].

¹⁸ Section 15(5) of the Act.

¹⁹ Kirkman, E., 2019. Free riding or discounted riding? How the framing of a bike share offer impacts offer-redemption. *Journal of Behavioral Public Administration*, 2(2).

Non-skippable microtutorials). These interventions were compared against a simple interface where information about sensitive content examples was hidden behind a ‘Learn more’ hyperlink (Basic presentation). Figure 1, above, shows the choices presented to users.

The Check and Update trial focused on prompting users to review their controls during browsing, assuming the initial choice happened some time ago. Under the Communications Act 2003 and the Online Safety Act 2023, Ofcom has a statutory duty to promote media literacy and to carry out research into media literacy. As part of fulfilling these duties, Ofcom is conducting research to establish what works when it comes to the promotion of media literacy online. This trial aimed to broaden our understanding of what influences users’ behaviour at a different stage of their user journey, ‘normal browsing’. At this stage, they may have no particular stimulus to think about their content controls but may be encountering content they would prefer not to see. We wanted to explore the effectiveness of different prompts to get users to engage with their content control settings at this stage, particularly given mounting evidence that users are resistant to frictions in their browsing, routinely clicking away interruptions.²⁰

We tested how prompt effectiveness varies when prompts appeal to different motivations. Prompts either emphasised the ease of updating settings (Process) or highlighted that users could take control of their feed (Empowerment). We also tested whether users were more likely to check their settings when they were encountering sensitive content and might be experiencing a reaction against it. Users were prompted either at the start of their browsing (Pre-engagement) or after they had disliked a sensitive post (Post-engagement).²¹ Figure 2, above, shows the prompts presented to users.

Together these trials give us a set of insights into how users make decisions about their content settings and the way those choices are shaped by platform design. This enhances the evidence base for Ofcom’s forthcoming codes of practice on user empowerment and our [Best Practice Design Principles for Media Literacy](#).

This report is structured as follows. The following section describes the Sign-up trial in more detail, including the rationale behind the interventions, the experiment design, and the findings. Following a similar structure, we then delve into the Check and Update trial. Finally, we discuss the key findings and conclusions across both trials, and outline potential avenues for future research.

²⁰ For example, Bahr, G.S. and Ford, R.A., 2011. [How and why pop-ups don’t work](#): Pop-up prompted eye movements, user affect and decision making. *Computers in Human Behavior*, 27(2), pp.776-783. Bravo-Lillo, C., Cranor, L., Komanduri, S., Schechter, S. and Sleeper, M., 2014. Harder to ignore? revisiting {Pop-Up} fatigue and approaches to prevent it. In 10th Symposium On Usable Privacy and Security (SOUPS 2014) (pp. 105-111).

²¹ Some users did not dislike any sensitive posts. They still received a prompt after they saw the last sensitive post.

2. Sign-up Trial Interventions

Why users might struggle to make an informed choice at sign-up

When users make a choice about the amount of sensitive content on their feed as part of the sign-up process, they may not have seen the platform’s feed. Therefore, they may not know what they might encounter if they choose “All content types” nor what they might miss out on if they choose “Reduced sensitive content”. They may not know the standards that platforms employ to define sensitive content. As a result, their decision-making is likely to be influenced by what information is provided (if any) and how that information is packaged up and presented (for example, whether salient, hidden, short or lengthy, in technical or plain English).

Some users may not be motivated to engage with the choice offered, preferring to click through quickly to start browsing. They may just go with the flow, leaving them susceptible to following the default settings selected by the platform.

To make an informed, active choice, users need to read and understand the information, reflect on the available options, and select the option that suits them best. We conducted initial desk research and ran internal workshops with Ofcom’s behavioural insights and online safety specialists to identify barriers that could interfere in this process (see Annex 1). Following a prioritisation exercise (see Annex 1), we selected the following barriers for intervention development.

- Lack of *attention* to the information and choices; skimming through to get to the feed.
- Lack of *understanding* of the information, and the different options.
- *Friction* in the form of extra clicks to get more detailed information.
- Tendency to stay with the *status quo*, such as a pre-selected option.

Addressing the barriers

To develop interventions aimed at the identified barriers, we took inspiration from [the CMA’s taxonomy of online choice architecture practices](#).²² We took into account i) relevance to choices made at the sign-up stage; ii) existing evidence and policy context; and iii) practical constraints. The main focus areas considered for intervention development included:

Choice information

- **How the information is presented:**²³ salience of information, ease of access, visual and design elements.

²² CMA, 2022. [Online Choice Architecture: How digital design can harm competition and consumers](#).

²³ We departed from the CMA’s taxonomy in this instance to better align with our specific context. The practices we considered here have some similarity to ‘Dark nudge’ and ‘Sensory manipulation’, but they are broader and relate to choice information rather than choice structure.

- **Framing of information:** how content control options are labelled (e.g. ‘restricted mode’ vs ‘safe mode’) and the terms used to describe them, whether broad (e.g. ‘sensitive’) or specific (e.g. violence, self-harm, etc.).

Choice structure

- **Defaults:** whether an option is pre-selected (and if so, which option).
- **Choice overload and bundling:** granularity of options (how many options are offered and what each option entails), and whether options are bundled into a ‘package’.

Ultimately, we prioritised **how the information is presented and defaults**, as we wanted to build on the most common practices and well-evidenced behavioural levers. Additionally, our focus was to start with generating high-level insights with broad relevance rather than delving into specific nuances. We hope to build evidence on the impact of the full list of factors listed above in the future and would be interested in any related evidence. Please contact us at Behavioural.insights@ofcom.org.uk if you have evidence you can share.

We developed a 5-arm trial design. The information provided to users and the choice options were the same across all arms, but we 1) introduced a default choice; and 2) varied how the information was presented. Figure 4 summarises the trial arms. The details and rationale for each intervention are explained below.

Figure 4: Overview of trial arms

Basic presentation Sensitive content examples hidden behind 'Learn more' hyperlink	Default A variation of the control with "All content types" pre-selected
<p>Choose how much sensitive content appears in your own feed.</p> <p><input type="radio"/> All content types You may see some posts with sensitive content</p> <p><input type="radio"/> Reduced sensitive content You will see fewer posts with sensitive content</p> <p>Sensitive content doesn't go against our Community Guidelines, but refers to topics some people don't want to see. Learn more.</p> <p><input type="button" value="Next"/></p>	<p>Choose how much sensitive content appears in your own feed.</p> <p><input checked="" type="radio"/> All content types You may see some posts with sensitive content</p> <p><input type="radio"/> Reduced sensitive content You will see fewer posts with sensitive content</p> <p>Sensitive content doesn't go against our Community Guidelines, but refers to topics some people don't want to see. Learn more.</p> <p><input type="button" value="Next"/></p>
Info saliency Sensitive content examples presented on the decision page itself	Microtutorials: Sensitive content examples presented via a brief training module. Non-skippable: cannot skip microtutorial Skippable: button to skip microtutorial
<p>Choose how much sensitive content appears in your own feed.</p> <p><input type="radio"/> All content types You may see some posts with sensitive content</p> <p><input type="radio"/> Reduced sensitive content You will see fewer posts with sensitive content</p> <p>Sensitive content doesn't go against our Community Guidelines, but refers to topics some people don't want to see, such as:</p> <ul style="list-style-type: none"> • 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 labeled as false or partly false by impartial third-party fact-checkers such as false news. <p><input type="button" value="Next"/></p>	<p>Choose how much sensitive content appears in your own feed.</p> <p>What do we mean by sensitive content? Follow the tutorial to find out.</p> <p><input type="button" value="Next"/></p> <p>Sensitive content doesn't go against our Community Guidelines, but refers to topics some people don't want to see.</p> <p><input type="button" value="Skip tutorial"/></p>

To assess whether users were satisfied with their initial choice we asked them to review their choice after a period of browsing ('Review' stage). This provided users with a chance to change their initial choice if they felt it did not reflect their preferences. A high degree of change at this stage could indicate that the choice architecture at account set-up is distorting users' choices.²⁴

Basic presentation (Control)

We aimed to create a basic but clear design that would serve as a comparison for the interventions. The design was inspired by existing designs on common social media platforms but did not fully replicate any of them. We expected that few users would click the 'Learn more'

²⁴ There could be other reasons why users prefer to keep or update their controls. These are discussed in the Findings section. However, we expect that the proportion of participants changing controls for other reasons, such as curiosity, would be stable across different arms.

hyperlink to discover the examples of sensitive content and instead would base their choice on their own understanding of sensitive content.

Default

Defaults introduce a barrier to making an active choice because users can move on without changing the pre-selected option. Defaults are common online, including on social media platforms, as they are an easy-to-implement tool to guide users' decisions.

We expected that pre-selecting "All content types" would reduce users' engagement with the information provided and increase the likelihood of participants proceeding to the feed with this option compared to the Basic presentation. We expected that this choice would be less active than in the other trial arms, and therefore, after seeing the feed, participants would be more likely to change their choice to "Reduced sensitive content".

Information saliency

Reducing the number of steps to access information can increase engagement with it.²⁵ Engagement can also be improved by making information more noticeable to attract users' attention. For example, ease of navigation and salience of controls improved outcomes in [Ofcom's reporting trial](#) and [BIT's Active Online Choices](#) experiments.^{26,27}

Providing information with examples of sensitive content on the choice page is unobtrusive for users, removes an extra click, and should not be costly to implement for platforms. We expected it would improve user comprehension and help them make an informed choice.

Microtutorials: Non-skippable and Skippable

Microtutorials are short step-by-step online guides. Unlike nudges that steer decisions, microtutorials aim to [boost users' capabilities](#) to make their own choices.²⁸ In an earlier Ofcom trial, we found that microtutorials [increased reporting of potentially harmful content](#).²⁹ They are common in online environments and are a promising tool to rapidly educate and empower users. However, they also risk disrupting the user experience.

Our microtutorial chunked examples of sensitive content into small, digestible parts and had interactive elements to enhance engagement and prevent users from clicking through. We included Non-skippable and Skippable microtutorials to assess voluntary engagement levels, and their impact when users have to engage with them to progress. We expected that the microtutorials would prompt users to pause, read and reflect on the information. Ultimately, this improved understanding would empower users to make an informed initial choice.

²⁵ For example, reducing the number of click-throughs needed to access content can make a big difference to take-up, see for example Rosenkranz, S., Vringer, K., Dirkmaat, T., van den Broek, E., Abeelen, C. and Travaille, A., 2017. Using behavioral insights to make firms more energy efficient: A field experiment on the effects of improved communication. *Energy policy*, 108, pp.184-193.

²⁶ BIT, 2021. [Active Online Choices: Designing to Empower Users](#) [accessed April 4, 2024].

²⁷ Ofcom, 2023. [Behavioural insights for online safety: understanding the impact of video sharing platform \(VSP\) design on user behaviour](#).

²⁸ Hertwig, R. and Grüne-Yanoff, T., 2017. Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6), pp.973-986.

²⁹ Ofcom, 2023. [Boosting users' safety online: Microtutorials](#).

3. Sign-up Trial Experiment Design

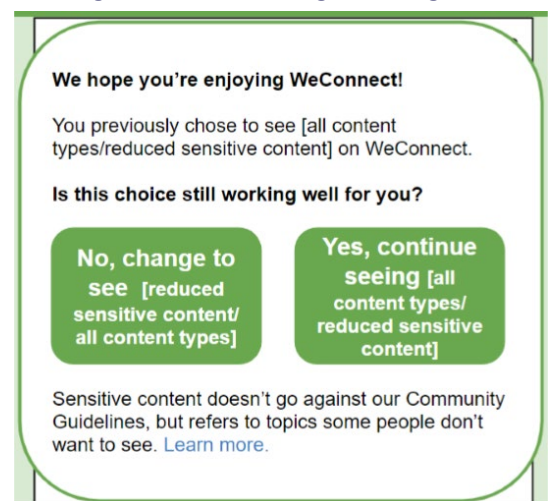
We tested these interventions on a simulated social media platform called WeConnect that mimicked real platforms. BIT ran the experiment with 3,500 adult participants in the UK in November – December 2023. Participants were randomly allocated to see one of the five designs of the initial content control interface.

Participants’ user journey included the following key components:

- a) **Sign-up** to WeConnect which included making their initial choice about the amount of sensitive content on their feed. This is where the interventions described in the previous section were incorporated.
- b) **Browsing the feed** which consisted of 24 content pieces, including short videos and text posts, some with accompanying images. The number of sensitive posts depended on the content settings choice during the sign-up process: 12 pieces of sensitive content for “All content types” and 2 pieces of sensitive content for “Reduced sensitive content”.³⁰ The sensitive content categories included hate, violence, and misinformation (see [Technical Report](#) for details on sourcing content, safeguarding procedures and ethical considerations).
- c) **Review stage** where participants were asked whether their pre-feed content settings choice was still working well for them as shown in Figure 5. The proportion of participants choosing “Yes, continue seeing...” was our main outcome of interest. If participants in one of the intervention arms were more likely to stay with their initial choice, our interpretation is that the intervention helped them make an active well-informed choice in the first place.³¹
- d) A **follow-up questionnaire** to understand participants’ behaviour, comprehension of what was classified as sensitive content in this trial and sentiment towards content controls. This included asking participants why they chose to change or continue with the initial content settings.

More details can be found in the [Technical Report](#).

Figure 5: Review stage message



³⁰ Users who selected “Reduced sensitive content” had 2 pieces of sensitive content in their feed to make it more realistic as we assumed that it could be difficult for platforms to filter out all sensitive content.

³¹ There could be various reasons driving users’ decisions to keep or change their controls. These are discussed in the Findings section.

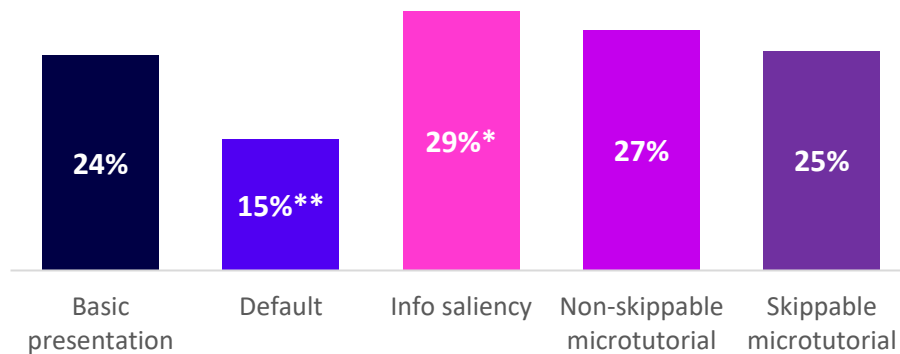
4. Sign-up Trial Findings

Pre-selection of “All content types” decreases choice of “Reduced sensitive content”; Info saliency increases it

We found that 1 in 4 participants (24%) chose to see “Reduced sensitive content” during sign-up in the Basic presentation arm. However, only 15% made this choice in the Default arm where “All content types” was pre-selected. This is in line with our expectations and the broader behavioural evidence showing that pre-selecting an option is a powerful way to affect user choice.³²

As expected, making the examples of sensitive content easier to access (Info saliency) increased the proportion of participants choosing “Reduced sensitive content” to 29% compared to 24% in Basic presentation. However, showing the examples via Non-skippable or Skippable microtutorials did not lead to significant changes compared to Basic presentation (see Figure 6).

Figure 6: Percentage of participants who chose “Reduced sensitive content” at sign-up



Note: ** statistically significant at the 1% level ($p < 0.01$); * statistically significant at the 5% level ($p < 0.05$) in comparison to Basic presentation.

Strong tendency to continue with initial choice

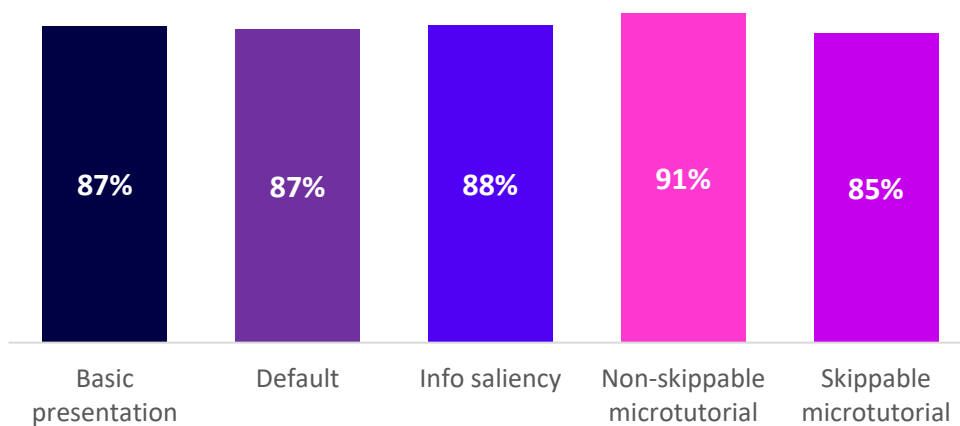
Differences in initial choice did not translate into different decisions at the review stage where participants were offered the chance to change or keep their initial choice having seen the feed.

After seeing the feed, 88% of participants on average chose to continue with their initial choice, regardless of what that initial choice was. The differences between the trial arms were not significant, although directionally the Non-skippable microtutorial performed best with 91% keeping their initial choice (87% in Basic presentation).

³² Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, 3(2), 159-186.

These results are surprising as we expected that, since more users started with “All content types” in the Default arm, more users would revise their choice after seeing the feed and realising it was not in line with their preferences. We also expected that Info saliency and Non-skippable and Skippable microtutorials would help users make an informed initial choice, and they would be more likely to stay with that choice than those who saw the Basic presentation.

Figure 7: Percentage of participants who kept initial choice at the review stage



Note: None of the interventions showed statistically significant differences at $p < 0.1$ compared to Basic presentation.

Additional mini-experiment with “Reduced sensitive content” as Default

Following the completion of the trial, BIT ran an exploratory mini-experiment to investigate the effect of pre-selecting a different option on the initial choice page. An additional 700 participants were recruited, and all were allocated to have “Reduced sensitive content” pre-selected on the initial choice page.

Around 42% of users proceeded with “Reduced sensitive content” as their initial choice. At the review stage, around 85% of users stayed with the initial choice.

While exploratory, these results support our main findings about the strong effect of pre-selecting an option on the initial choice page and the ‘stickiness’ of that choice.

Why do so many users stick with their initial choice?

One potential explanation is that underlying preferences for these choice options are relatively weak. Most users may simply not be motivated to change their initial choice, regardless of what it was. In a separate survey with YouGov, Ofcom found that 66% of users who were aware of content controls but never used them said they [just don’t think they need any content controls](#).³³

It may also be that the available controls do not meet their needs. Participants were offered a binary choice – “All content types” or “Reduced sensitive content” while preferences may be more granular. If neither option reflects the browsing experience they are looking for, they may care less which of these settings they end up with. For example, some users may prefer not to

³³ Ofcom, 2024. [Terms and conditions and content controls](#).

see specific categories of sensitive content, but do not want to choose “Reduced sensitive content” so they do not miss out on content they are interested in.

[Lack of trust](#) towards how platforms categorise content can also be an important factor for not engaging with content controls.³⁴

These hypotheses are in line with the reasons provided by participants for continuing with the initial choices in the follow-up questionnaire. Almost half (48%) thought it was the right option for them but only a quarter (26%) said they continued with their initial choice because they liked the content. Moreover, 18% said they were worried about missing out on content that they would like to see.

Why did the microtutorials not lead to more users keeping their settings?

Among the users who saw the Non-skippable microtutorial, 91% kept their initial choice at the review stage, but the difference with 87% in Basic presentation was not statistically significant. This is notable given the strong impact microtutorials had on [reporting behaviour](#) in Ofcom’s previous research on VSPs.³⁵ One important factor could be that the ‘baseline’ proportion of participants keeping their settings in Basic presentation was already high in this trial (87%). It may be disproportionately harder to get a significant improvement when the baseline is high.

Notably, among participants who could skip the microtutorial, 73% did so. The biggest reasons for skipping were not needing a tutorial (58%) and already knowing about sensitive content (43%). On the other hand, 66% who received the Non-skippable microtutorial found it easy to follow, and 35% said it helped them learn more about sensitive content.³⁶ Only 9% wished they could skip it and 4% found it annoying. When given the chance to skip, only 27% of users in our trial persist with a microtutorial, yet when no opportunity to skip is given, users do not seem to object.

Reasons for changing initial choice

We asked participants who made a change to their initial choice after seeing the feed about the reasons for change (participants could select multiple reasons). The most popular reason was being curious to see what would change (38%). 32% changed because they saw content that upset them, followed by 30% who did not like the content and 21% who said content did not match the expectations based on the original choice.

Different ways of presenting sensitive content examples did not affect comprehension

The interventions did not affect participants’ comprehension of what content is considered sensitive. Only five participants in the Basic presentation arm clicked ‘Learn more’ and thus had an opportunity to read the sensitive content examples. Given that the Info saliency intervention

³⁴ Ofcom, 2024. [Terms and conditions and content controls](#).

³⁵ Ofcom, 2023. [Boosting users’ safety online: Microtutorials](#).

³⁶ Participants could select all applicable options.

and the microtutorials were expected to drive participants' attention towards these examples, the lack of differences in comprehension is surprising.

These results may be driven by participants having their own pre-existing understanding of what constitutes sensitive content.³⁷ Making the platform's definitions more salient or presenting them in small chunks may not override users' pre-existing understanding. Even though comprehension and the review stage outcomes were unaffected, the prominence and clarity of important information did influence users' initial choices.

Positive sentiment, small differences across interventions

The aggregated sentiment towards the content settings page was positive and similar across the trial arms.³⁸ There were some differences between the questions comprising the sentiment score. Perceptions that the settings page is easy to understand and presented in a fair way were higher in the Info saliency arm and both Skippable and Non-skippable microtutorials compared to Basic presentation. These improvements did not affect comprehension and alignment with preferences.

Interestingly, there was no major reduction in positive sentiment among participants who saw "All content types" pre-selected (Default arm). The only exception was trust that the choices were presented with their best interests in mind (73% in Default vs 78% in Basic presentation). This indicates that users may have understood that pre-selected option may not be in their best interest but did not react strongly to it.

Limitations

The key limitations relate to the simulated nature of the environment which may not fully replicate the incentives and motivations that guide users' behaviour on social media. Moreover, real-world sensitive content may include content that is more harmful and more personalised than the content shown in our research. Finally, 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. On this basis, we have more confidence in the relative impact of interventions, than the precise measures of the magnitude of impact.

³⁷ We note that the wording of the question "Which of the following would you describe as sensitive content?" could have given the impression that participants were asked about their own interpretation of sensitive content rather than the platform's definition.

³⁸ Composed of questions about whether participants thought the content settings page was 1) easy to understand; 2) made them feel in control of the content they saw; 3) was presented in a fair way, and 4) whether they trusted that the choices were presented with their best interests in mind.

5. Check and Update Trial Interventions

Barriers to checking and updating content controls

In the Check and Update trial, we focused on a different part of the user journey – checking content controls *when browsing the feed* – and different behavioural barriers, namely motivation barriers. We explored a scenario where sign-up happened some time ago and users may not remember what their setting is.

To support our work in establishing Best Practice Design Principles for Media Literacy as well as our wider media literacy work, Ofcom commissioned YouGov to conduct qualitative research on user attitudes towards [on-platform media literacy interventions](#) and what can improve their efficacy.³⁹ Such interventions include but are not limited to labels, overlays, pop-ups, notifications, and resources, and can shape user behaviour on social media platforms. In the context of content controls, these interventions could be used to provide additional context or information to support users to make informed decisions, and reflect on their behaviour. While platform settings may be extensive and [difficult to navigate](#),⁴⁰ well-designed prompts in the form of pop-ups could provide an opportunity for users to reflect on their experience and easily change their settings.

Nevertheless, even when users are prompted, they may not have the motivation to engage and simply click past the prompt. More specifically users may:⁴¹

- Not be motivated to pay *attention* to the prompt in order to continue browsing;
- *Believe that they do not have the ability* to manage the feed; or
- *Have an expectation* that checking and updating content control settings is *onerous*.

Addressing the barriers

We focused on one type of on-platform intervention – prompts, in the form of pop-ups. We wanted to explore how we might mitigate the motivational barriers above and increase engagement with prompts. We hypothesised that prompt *timing* and *messaging* may address these barriers.

We developed a 5-arm trial design to test the following hypotheses.

- Prompts in the form of pop-ups encourage users to check their controls.
- Targeted timing of the prompt can improve its effectiveness.
- Messages targeting beliefs and expectations can address motivational barriers and improve prompt effectiveness.

³⁹ Ofcom, 2023. [User attitudes to on-platform interventions](#).

⁴⁰ CDEI, 2020. [Online targeting: Final report and recommendations](#).

⁴¹ We used the COM-B model presented in Annex 1 to check that these barriers were expected to be important, but not already explored in the Sign-up trial.

Prompts

The control arm did not contain any prompts. If users chose to click on the gear icon, they were presented with a pop-up allowing them to check their settings (Figure 8).

Timing: prompting before or after engagement

Prompts at timely moments can be an effective tool for behaviour change as users may be more receptive to a message.⁴² Moreover, emotional states can shape our actions.⁴³ Being in a 'hot' or 'cold' emotional state may impact users' willingness to check their content controls.⁴⁴ We decided to explore prompts at two different points linked to user engagement with the content and their potential emotional states.⁴⁵

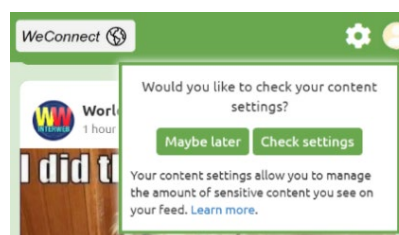
- **Pre-engagement: prompt at the start of the feed.** Participants received a prompt after the first post on the simulated feed (the first post was always non-sensitive). We expected that this would correspond to a higher likelihood of users being in a 'cold' state, as they had not yet been exposed to any sensitive content in the feed.⁴⁶
- **Post-engagement: prompt after disliking a sensitive post.** We expected that for a user browsing social media feed, a timely moment to check their content controls could be when they see something they do not like and have an emotional reaction to it. In our experimental setting, this translated into a prompt after clicking the 'dislike' icon on a sensitive post. Those who did not dislike any sensitive posts received the prompt after viewing the last sensitive post.

Messages: ease of process vs empowerment

Decisions are influenced by how information is worded and what aspects are emphasised. Varying the wording of messages can influence behaviour and resonate with different emotions. To develop the messages, we workshopped a long list of ideas (see Annex 2). We then prioritised them based on their expected impact and relevance to the trial. We aimed for each message to target a particular motivational barrier as we expected motivation to be an important factor driving engagement with content controls.

We developed two messages: one emphasising that the user is in control of the content in their feed and one emphasising the simplicity of checking and updating controls. These are referred to as 'Empowerment' and 'Process' messages (participants did not see these labels).

Figure 8: No prompt (Control). Pop-up appears if users click on the gear icon.



⁴² BIT, 2014. [EAST Four simple ways to apply behavioural insights](#) [accessed January 25, 2024].

⁴³ Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R. and Vlaev, I., 2012. Influencing behaviour: The mindspace way. *Journal of economic psychology*, 33(1), pp.264-277.

⁴⁴ We define the 'cold' state as being more rational and logical, and not influenced by emotions, while the 'hot' mental state reflects being influenced by emotions.

⁴⁵ We can only hypothesise about our participants emotional state after seeing each post as this would be challenging to measure reliably.

⁴⁶ The experiment does not take place in a vacuum: participants may have entered the experiment in a 'hot' emotional state driven by something outside of the experiment. However, the advantage of an RCT is that, by randomly assigning participants to trial arms, we are more likely to have a balance of unobservable characteristics (such as emotional states) across trial arms.

- **Empowerment message: *Your feed, your choice – you can choose the amount of sensitive content that you see.***

Users value control over their online experiences, but often feel that they [lack control](#).⁴⁷ We expected that mitigating the perceived powerlessness would increase users’ sense of autonomy and increase engagement.

- **Process message: *It takes just two steps to check and update your content settings.***

People often find user controls [difficult to find and navigate](#), and some may have concerns that they are purposefully designed that way.⁴⁸ In a survey conducted for Ofcom, 23% of users mentioned that not having time was one of the reasons [not to use content controls](#) on social media platforms and VSPs.⁴⁹ Updating content controls on our simulated platform was quick and easy but participants could still be concerned the process would be onerous. The Process message was designed to mitigate such concerns.

Figure 9: Summary of interventions

	Empowerment message	Process message
Pre-engagement	<p>Your feed, your choice – you can choose the amount of sensitive content that you see.</p> <p>Maybe later Check settings</p> <p>Sensitive content doesn't go against our Community Guidelines, but refers to topics some people don't want to see. Learn more.</p> <p>You can check your settings at any time by clicking the gear icon ⚙️ in the top right.</p>	<p>It takes just two steps to check and update your content settings.</p> <p>Maybe later Check settings</p> <p>Sensitive content doesn't go against our Community Guidelines, but refers to topics some people don't want to see. Learn more.</p> <p>You can check your settings at any time by clicking the gear icon ⚙️ in the top right.</p>
Post-engagement	<p>We noticed you just disliked a post. Your feed, you choice – you can choose the amount of sensitive content that you see.</p> <p>Maybe later Check settings</p> <p>Sensitive content doesn't go against our Community Guidelines, but refers to topics some people don't want to see. Learn more.</p> <p>You can check your settings at any time by clicking the gear icon ⚙️ in the top right.</p>	<p>We noticed you just disliked a post. It takes just two steps to check and update your content settings.</p> <p>Maybe later Check settings</p> <p>Sensitive content doesn't go against our Community Guidelines, but refers to topics some people don't want to see. Learn more.</p> <p>You can check your settings at any time by clicking the gear icon ⚙️ in the top right.</p>

⁴⁷ Centre for Data Ethics and Innovation, 2020. [Online targeting: Final report and recommendations](#) [accessed January 25, 2024].

⁴⁸ Centre for Data Ethics and Innovation, 2020. [Online targeting: Final report and recommendations](#) [accessed January 25, 2024].

⁴⁹ Ofcom, 2024. [Terms and conditions and content controls](#).

6. Check and Update Trial Experiment Design

We made small modifications to the simulated social media platform used in the Sign-up trial to fit the research purposes of this trial. BIT ran the experiment with 3,602 adult participants in the UK in December 2023 – January 2024. Participants were randomly allocated to receive one of the four prompts or no prompt at all.

Participants' user journey included the following key components.

- a) **Training task** where participants saw three non-sensitive posts and were asked to like, dislike and repost at least one post before they could proceed. The main aim of this stage was to prime participants to interact with the content. This was important because in one of the trial arms participants got a prompt after disliking a sensitive post. There was **no sign-up stage** because we were focusing on users updating their existing settings.
- b) **Browsing the feed** which consisted of 24 content pieces, including short videos and text posts, some with accompanying images. All participants started with "All content types" setting enabled, which meant 12 content pieces (50%) were sensitive. Similarly to the Sign-up trial, the sensitive content categories included hate, violence, and misinformation (see [Technical Report](#) for details).

If participants changed their setting to "Reduced sensitive content", the sensitive content in the remainder of the feed was replaced with non-sensitive posts. Sensitive content they had already seen was removed from their feed.

- c) A **follow-up questionnaire** included questions on recall, reasons for checking or not checking their settings, sentiment towards the prompt, their past experiences with content controls and content preferences.

We also asked participants about their risk preferences, their mood and energy level before the experiment. We speculated that these factors may be related to users' willingness to check their controls and wanted to gather evidence to inform future research.

7. Check and Update Trial Findings

Prompts encouraged checking of settings; timing and message mattered

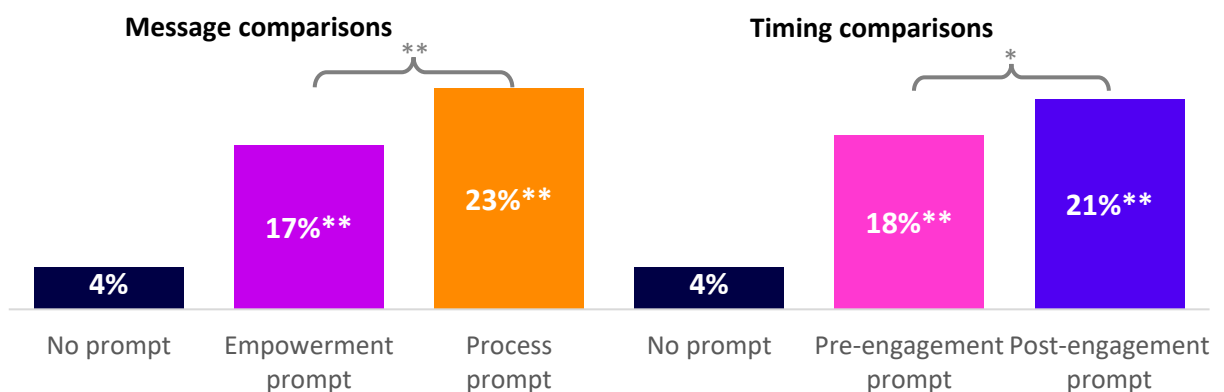
Without a prompt, only 4% of participants checked their content settings. The Process prompt, highlighting that content controls could be updated in two steps, encouraged 23% of participants to check their controls. This prompt was more effective than the Empowerment prompt highlighting that they were in control of their feed (17% checked). Users may have concerns that changing content settings is onerous and the Process message mitigated these.

Additionally, participants who saw the prompt after, rather than before they had engaged with the social media feed, were more likely to check their settings (21% vs 18%). Most participants allocated to this intervention disliked at least one sensitive post (88%) and received the prompt after disliking. The remaining 12% saw the prompt after the last sensitive post.⁵⁰

The effectiveness of the Post-engagement prompts supports our assumption that a timely moment to encourage users to check their content controls may arise when a user has just seen sensitive content, and disliked it. The feeling of dislike appears to increase users' motivation to change their feed when compared to a 'cold' state of standard browsing where users seem to have less drive to change their browsing experience. Overall, the Process post-engagement prompt was the most effective at encouraging users to check their controls (25% checked).

Note, though, that the differences are not very large. It looks like simply being prompted does most of the work. The timing and wording of the prompts do make a difference, but the main driver of behaviour appears to be the prompt itself.

Figure 10: Percentage who checked their content setting



Note: ** statistically significant at the 1% level ($p < 0.01$); * statistically significant at the 5% level ($p < 0.05$) in comparison to No prompt control.

⁵⁰ BIT ran sensitivity checks to assess whether excluding the group that saw the prompt after the last sensitive post would change the findings. The findings remained consistent (see the Technical Report).

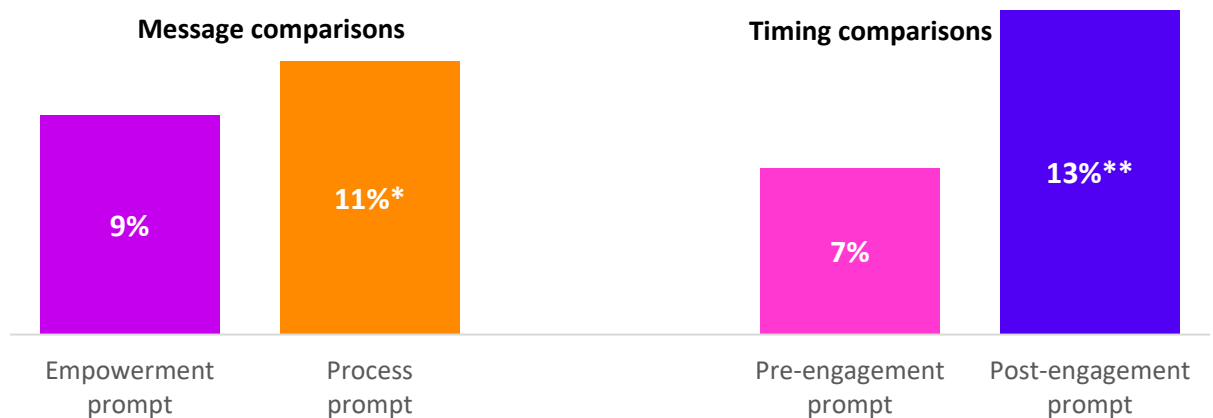
Timing and message also influenced the final setting

Overall, 8% of participants ended the experiment with the “Reduced sensitive content” setting (0% at the start of the experiment).

Participants were more likely to set their final choice to “Reduced sensitive content” if they

- saw the Process prompt rather than the Empowerment prompt (11% vs 9%), or
- saw the prompt after engaging with the feed rather than before (13% vs 7%).

Figure 11: Percentage who had “Reduced sensitive content” at the end of the experiment



Note: ** statistically significant at the 1% level ($p < 0.01$); * statistically significant at the 5% level ($p < 0.05$).

Mismatch between final choice and stated preferences

Even among the groups who saw the prompts, the vast majority still ended the experiment with the “All content types” setting (87%-93%). However, only 39% said they are usually comfortable seeing all content types.⁵¹ As a result, less than half of participants had their final setting match their stated preference. Although participants were not explicitly asked to match their setting with their usual preferences, this mismatch is notable.

Of all participants who did not check their settings, 41% said they were curious to see what content was available on the platform.⁵² Some participants did not pay attention to the prompt and did not know they could change the controls (only 33%-45% recalled the prompt). Finally, of those who did not check their settings, 17% said they do not care about it. As in the Sign-up trial, participants do not seem to have strong motivation to ensure their settings are optimised.

⁵¹ 49% said they are usually comfortable seeing reduced sensitive content on their feed, 39% said they are usually comfortable seeing all content types, while 12% said they do not know.

⁵² Participants could select multiple reasons. The data includes participants who did not get a prompt.

Higher risk preference associated with being more likely to check controls

Participants who indicated higher willingness to take risks with respect to content on social media platforms were significantly more likely to check their controls. We can hypothesise that these participants might be more concerned about content being restricted, while participants with lower willingness to take risks might be more concerned about seeing sensitive content. If so, concerns about content restrictions might be as strong a motivator to check content controls as concerns about seeing sensitive content. Another possible explanation is that higher risk preference might be related to higher willingness to explore platform functionality. Further research is needed to explore this.

Broadly positive sentiment towards prompts

Overall, 75% said the prompt was easy to understand, 65% said it made them feel in control of the content on WeConnect, and 62% said it was a useful reminder. This was not significantly different across different prompts.

Participants who saw the Process prompt were more likely to find the prompt annoying (22%) than those who saw the Empowerment prompt (18%).

Limitations

As in the Sign-up trial, the key limitations relate to the simulated nature of the environment and the content as well as the short-term nature of the experiment. In addition, this trial included a training task to prime participants to engage with the content. Engagement was high as 96% of participants engaged with at least one post in the feed compared to 60% in the Sign-up trial. This is likely to mean that the size of the impact of prompts are inflated relative to the Sign-up trial. However, we would not expect it to affect the relative size of impacts between prompts.

Finally, as with the Sign-up trial, the Check and Update trial did not test repeated exposure to prompts and the cumulative impact of prompts from different platforms that users are likely to encounter in the real-world context. Excessive use of prompts can be annoying, and users may [quickly dismiss](#) them.⁵³

⁵³ Bahr, G.S. and Ford, R.A., 2011. [How and why pop-ups don't work](#): Pop-up prompted eye movements, user affect and decision making. *Computers in Human Behavior*, 27(2), pp.776-783.

8. Discussion and Conclusion

For some social media users, seeing sensitive content can be distressing and cause harm. Enabling users to control what appears on their feed could contribute to fostering media literacy and keeping users safe online. Content settings are one control mechanism available to users. These trials provide valuable new insights on the significant ways that choice architecture can empower or disempower users to make an informed, active choice about the content they see.

Platform choice architecture influences users' initial decisions. In the Sign-up trial, **salient presentation of information about sensitive content** on the decision page resulted in more users choosing "Reduced sensitive content" at the sign-up stage (29% in Info saliency vs 24% in Basic presentation).

Pre-selecting an option also had a sizeable impact. Only 15% chose "Reduced sensitive content" in the Sign-up trial when "All content types" was pre-selected, while 42% did so in the additional mini-experiment when "Reduced sensitive content" was pre-selected.

Making an active, informed initial choice is particularly important because that choice is 'sticky'. Despite the differences in the initial choice, all participants were highly likely to stick with that initial choice when asked to review it (88% on average). None of the four interventions had a significant impact on this outcome compared to Basic presentation.

How choice architecture is used to influence the initial choice becomes particularly important in helping users control their online experience because users will likely stay with that choice (at least for some time).

Prompts encourage user action and messages focusing on simplicity delivered after engagement worked best. In the Check and Update trial, prompts emphasising process simplicity proved more effective in encouraging users to check content settings than prompts focusing on sense of control (23% vs 17% checked). Without a prompt, only 4% of participants checked their content settings. This finding underscores the importance of highlighting process simplicity to users. However, it is also important that the corresponding processes are designed to be simple and quick, so that the message accurately reflects this.

Prompting after engagement with sensitive content encouraged more participants to check their content settings than prompting before engagement (21% vs 18% checked). This suggests that the timing of prompts is important and can be made more impactful by linking to user interaction with the content and potentially their emotional state.

Sentiment towards the prompts was broadly positive. Overall, 75% said the prompt was easy to understand, 65% said it made them feel in control of the content on WeConnect, and 62% said it was a useful reminder. Although the Process prompt was perceived as more annoying (22%) than the Empowerment prompt (18%), this did not backfire on engagement as the Process prompt was found more effective. This provides reassurance about user attitudes to prompts.

Nevertheless, receiving too many prompts can be overwhelming and result in disengagement. Strategic use of prompts should focus on encouraging the most critical behaviours at the most impactful moments.

Defaults have a sizeable impact on final setting. The influence of defaults can be seen in the final setting participants had at the end of the experiment. In the Sign-up trial, where users had to make an active choice at sign-up, between 25% and 35% finished the experiment with the “Reduced sensitive content” setting. However, in the Check and Update trial, where all participants started with “All content types”, only between 7% and 13% of those who were prompted to check their settings finished with the “Reduced sensitive content” setting. This falls to 2% among participants who were not prompted.

Overall, we found that user choice with respect to sensitive content controls is heavily susceptible to the online choice architecture. Across both trials, we observed the stickiness of the initial setting, even when users were provided with low-effort opportunities to revise it, if needed. We hypothesise that users may prefer more tailored ways to reduce the likelihood of seeing certain types of content. Moreover, some users may have relatively weak underlying preferences regarding these choice options.

Future research

We are keen to explore the comparative efficacy of different control mechanisms when users can choose between them (e.g. hiding content, blocking accounts, blocking hashtags, and indicating they want to see less of similar content), and how this differs across user groups. Such insights could be pivotal in understanding which control mechanisms platforms should prioritise for prompting.

Moreover, we would like to explore the impact of prompts over time, prompt fatigue and what the optimal frequency is. Finally, we are interested in conducting research to understand how other online choice architecture elements, including but not limited to the framing of information and granularity of choice, can be used to further support users make choices that work for them.

A1 Long List of barriers and prioritisation

A1.1 Generating a long list of barriers

We conducted a quick review of the available evidence on awareness of, attitudes towards and engagement with user controls. We used the Capability – Opportunity – Motivation (COM-B)⁵⁴ model to map out the barriers that might be preventing users from aligning their settings with preferences.

Capability (Psychological, Physical)

Cognitive skills

- Limited mental capacity to engage with content controls
- Not understanding controls

Awareness

- Lack of awareness that content controls exist and what they do

Attention

- Limited attention span
- Users prioritise other actions (e.g. browsing)

Evaluating options

- Difficulty comparing different options and their impact

Memory

- Postpone, forget and never allocate time to engage with the content controls

Opportunity (Physical, Social)

Opportunities in the environment

- Difficultly navigating platforms, making user controls difficult to find
- Excessive friction (e.g. number of clicks to access controls)
- Specific/preferable controls unavailable

Prompts in the environment

- Lack of (salient) prompts
- Inconvenient timing of prompts
- Defaults not transparent
- Difficult to change
- Complex language

Resources & time

- No time to engage with controls

⁵⁴ Michie, S., Van Stralen, M.M. and West, R., 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science*, 6, pp.1-12.

Social norms

- Lack of explicit social norms and little awareness of what other users are doing

Role models

- Lack of role modelling of positive behaviour

Motivation (Reflective, Automatic)

Beliefs about consequences

- Don't believe anything will improve if they engage

Goals

- Lack of clear aim for engagement
- Automatic responses
- Automatically filter out or ignore user controls
- Users prone to inertia and status quo bias

Identity

- Engaging with online controls may feel out of character

Emotions

- Emotional response to both content and user controls may affect engagement

Habits

- No habit of looking for and using online controls

Belief in abilities

- Being under/over-confident about use of controls

A1.2 Prioritisation

We scored each barrier based on i) expected impact and centrality (considering positive and negative 'spillover' effects, and linkage with other barriers); and ii) ease of mitigation. Additionally, we considered the relevant policy context and how useful the insights on the barriers generated in this work would be.

Following this process, we prioritised the following barriers for focus in the Sign-up trial (the wording has been refined from the original list above for clarity).

- Lack of *attention* to the information and choices; skimming through to get to the feed.
- Lack of *understanding* of the information, and the different options.
- *Friction* in the form of extra clicks to get more detailed information.
- Tendency to stay with the *status quo*, such as a pre-selected option.

A2 Check and Update Trial: Initial message ideas

Table A2.1: List of initial message ideas

Barrier or enabler being targeted by message	Sample message text
Fear of distress	Keep yourself safe by reviewing your settings!
Sense of control	Curate the content that you want to see.
Perceived frictions	You can change your settings in just three clicks.
Timely moments to act	You have not reviewed your settings since signing up to {platform name}.
Lack of knowledge	You can learn more about what we mean by sensitive content in the settings.
Fear of negative outcomes	Changing your settings will not impact the other content you see on {platform name}.
Belief in ability to manage risks	Take control and manage risks from potentially distressing content.
Positive dimension / making the experience fun	Make your online experience better, more fun.
Trust	Do you trust this person?
Consequences of seeing harmful content	This post contains strong language, are you sure you want to post now or save to drafts for later?
Social considerations	Would you want your [X] to see this?
How the user is feeling	How are you feeling? / Is this a good time to respond?