*Behavioural Public Policy* (2025), 1–18
doi:10.1017/bpp.2025.10016

CAMBRIDGE UNIVERSITY PRESS

FINDINGS FROM THE FIELD

# How does the design of social media content controls shape users' choice? Evidence from an online experiment

Natalia Shakhina[1], Pinelopi Skotida[1], Sujatha Krishnan-Barman[2], Martin Wessel[2], Elena Meyer Zu Brickwedde[2], Thea House[2] and Rupert Gill[1]

[1]Office of Communications (Ofcom), London, UK and [2]The Behavioural Insights Team, UK
**Corresponding author:** Pinelopi Skotida; Email: pinelopi.skotida@ofcom.org.uk

(Received 4 December 2024; revised 3 April 2025; accepted 5 August 2025)

**Abstract**
Social media offers many benefits but also carries risks, including exposure to distressing content. The UK's Online Safety Act requires certain platforms to empower users to control the content they see. Content controls can reduce users' exposure to sensitive content. However, there is little public data on how platform design shapes the use of these controls. In our online randomised controlled trial on a simulated social media platform, participants were given an initial choice between seeing 'All content types' or 'Reduced sensitive content'. After browsing, they were given the opportunity to change their choice. In the Control arm, none of the options were pre-selected. 24% chose 'Reduced sensitive content'. Pre-selecting 'All content types' reduced this proportion to 15%. Conversely, adding a description of 'sensitive content' on the choice page increased that figure to 29%. The initial choice proved to be 'sticky'. When invited to review after browsing, those defaulted away from 'Reduced sensitive content' did not switch any more than those whose choice was not influenced by a default. Overall, user choice was susceptible to choice architecture, and users' tendency to update their initial choice was weak. This highlights the importance of platform design to deliver genuine user empowerment.

**Keywords:** online choice architecture; online harms; randomised controlled trial; sensitive content controls; social media user behaviour

## Policy challenge and research aims

Social media provides many benefits, but almost 3 in 10 adults in the UK (27%) say they have recently been exposed to potentially harmful content (Ofcom, 2024). Many social media platforms offer users tools to control what appears in their feeds to avoid distressing or harmful content (see example in Figure 1). We refer to such content as

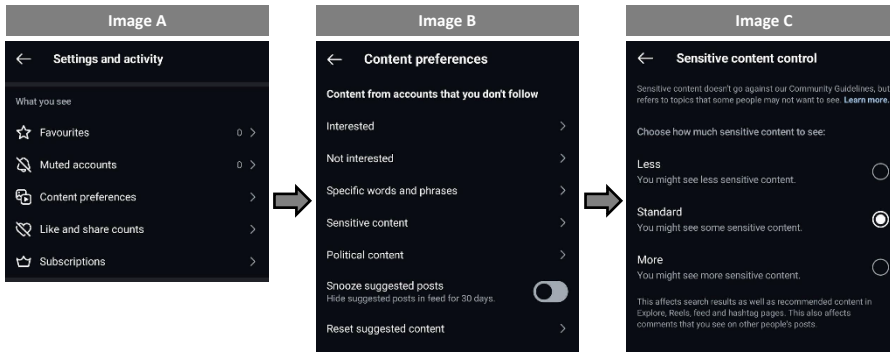


**Figure 1.** Example of existing content controls.
*Note:* Content controls are tools offered by social media platforms that allow users to control the content that appears in their feeds. In this trial we looked at sensitive content controls that users can select to avoid distressing or harmful content. Image A: Screenshot from social media settings and activity page [accessed 21 March 2025]. Image B: Screenshot from social media content preferences page [accessed 21 March 2025]. Image C: Screenshot from social media sensitive content control page [accessed 21 March 2025].

'sensitive content' and the tools to adjust these settings as 'content controls' or 'content settings'.

Only about a quarter of social media users say they have ever used them (Ofcom, 2024). Lack of awareness is one factor, but there are other barriers to engagement, such as perceived lack of time, complex settings, difficulty finding them and previous experiences (Ofcom, 2024).

The UK's Online Safety Act 2023 ('the Act') requires certain online services,[1] including some social media platforms, to offer adult users tools to control what content they see 'at the earliest possible opportunity'. Regulators are interested in how platform design influences user choices and whether it enables users to act in line with their preferences. This research contributes to the evidence base in this policy area.

Behaviour change literature shows that small changes in the environment – the choice architecture – can influence people's decisions (Thaler, Sunstein, & Balz, 2014). For example, defaults, complicated layouts, long disclosures and biased language can hinder users' engagement with online controls (Centre for Data Ethics and Innovation, 2020). Exploring the impact of online choice architecture addressing harmful practices is increasingly important for policymakers and regulators (Busch & Fletcher, 2024). Research on its impact on safety, privacy and data sharing is growing (Bauer, Bergstrøm, & Foss-Madsen, 2021). However, there is limited public research in the context of social media content controls. This research addresses that gap.

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

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

This experiment focused on one mechanism by which users can control their feed: users choosing between 'All content types' and 'Reduced sensitive content' when signing up for a new social media platform. This choice may have a long-lasting impact on their experience due to people's tendency to stick with the existing option, known as status quo bias (Samuelson & Zeckhauser, 1988).

We explored how choice architecture can help or hinder users' ability to make informed decisions about the amount of sensitive content on their feeds. Our goal was not to steer users towards a 'safer' choice but to help them understand, reflect on and select the option that suits them best.

## Methodology

We identified barriers, developed interventions and tested them on a simulated social media platform with UK adults using a randomised controlled trial (RCT).

### *Diagnosing barriers*

To identify barriers to user engagement with content controls, we conducted a literature review, and workshops using the Capability, Opportunity, Motivation and Behaviour model (Michie *et al.*, 2011) and the Theoretical Domains Framework (Atkins *et al.*, 2017). See Supplementary materials A for the list of barriers and the prioritisation approach. After prioritising, we selected the following barriers for intervention development:

- Lack of attention to the information and options as users skim through to get to the feed.
- Lack of understanding of the information on content controls or the different options.
- Key information about controls is often buried in submenus or requires additional navigation, making it less visible and accessible to users.
- Tendency to stay with the status quo, such as a pre-selected default option.

### *Intervention design*

To address the identified barriers, we developed interventions focusing on choice information and choice structure based on the Competition & Markets Authority's taxonomy of online choice architecture practices (Competition & Markets Authority, 2022). Under choice information, we considered how information is presented, such as its salience, ease of access, visual and design elements, as well as framing, such as how content control options are labelled and described. Under the choice structure, we considered pre-selecting a default option, and the granularity and bundling of options.

Ultimately, we prioritised testing (1) the use of defaults, as one of the strongest choice architecture interventions (Mertens *et al.*, 2022) and effective across different domains (Jachimowicz *et al.*, 2019), not yet explored in this context; and (2) how information is presented, building on research on video-sharing platforms (Ofcom, 2023).

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

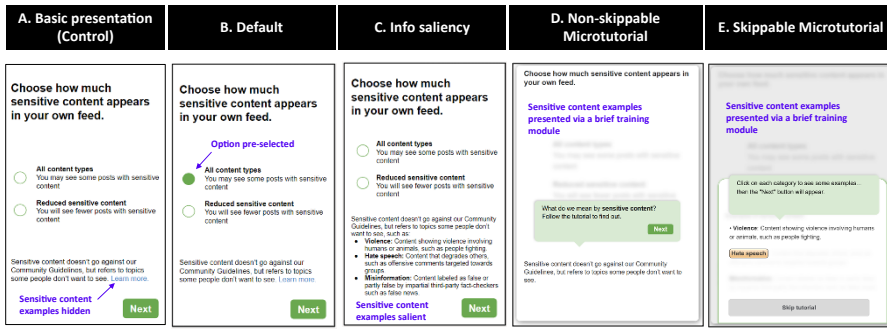


**Figure 2.** Trial interventions overview.
*Note:* Our five-arm RCT with one control and four intervention arms. Image A: Control arm. It reflected the design popular social media platforms currently use for their sensitive content settings. Sensitive content definition and examples are hidden under 'Learn more'. Image B: Intervention 'Default'. 'All content types' option was pre-selected. Image C: Intervention 'Info saliency'. Examples of sensitive content were more salient and easier to access. Image D: First screen of the intervention 'Non-skippable microtutorial'. Sensitive content examples were presented via a brief interactive training module where participants did not have the option to skip past it. Image E: Third screen of the intervention 'Skippable microtutorial'. Sensitive content examples were presented via a brief interactive training module where participants had the option to skip past it.

We developed a five-arm RCT with one control and four interventions. Figure 2 gives an overview of the trial arms.

To determine whether users' initial content control choices align with their preferences, we included a 'review' stage. We asked participants whether they would like to keep or change their choice after a period of browsing (Figure 3; details about the feed are under '*Experimental flow and simulated social media platform*'). If an intervention leads to a higher degree of change compared to the control, this could indicate that the choice architecture of that intervention is distorting users' choices at sign-up, leading them to make selections that do not reflect their true preferences (see Figure 4 for the full list of hypotheses).[2]

## *Trial arms and hypotheses*

### *Basic presentation (Control arm)*

The Control arm reflected the content control design used by many popular platforms. None of the options were pre-selected, so participants could make an active choice. Users could access a more detailed definition of sensitive content by clicking 'Learn more'. The information provided was the same as in all other trial arms (Figure 2).

[2]There could be other reasons why users prefer to keep or update their controls. These are discussed in the Results section and Supplementary materials F9 and F10. However, we expect that the proportion of participants changing controls for other reasons, such as curiosity, would be stable across different arms as the interventions did not target them.

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

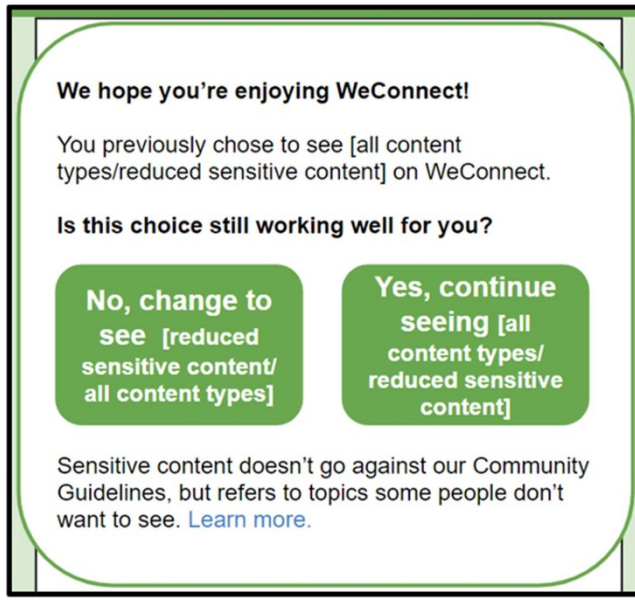


**Figure 3.** Review stage message.
*Note:* Prompt offering users the option to keep or change their initial content control choice at the end of the browsing stage.

| | |
|---|---|
| **1. Primary hypotheses** | |
| **Hypothesis 1a** | Participants in the Default arm would be significantly more likely to change their content settings to 'Reduced sensitive content' after viewing the feed compared to the Control. |
| **Hypothesis 1b** | Participants in the Information salience arm would be significantly less likely to change their content settings after viewing the feed compared to the Control. |
| **Hypotheses 1c and 1d** | Participants in the non-skippable and skippable microtutorials would be significantly less likely to change their content settings after viewing the feed compared to the Control. |
| **2. Secondary hypotheses** | |
| **Hypothesis 2a** | Default will significantly reduce the likelihood of participants correctly identifying content as sensitive compared to the Control. |
| **Hypothesis 2b** | Making information more salient would significantly increase the likelihood of participants correctly identifying content as sensitive compared to the Control. |
| **Hypotheses 2c and 2d** | Non-skippable and skippable microtutorials would significantly increase the likelihood of participants correctly identifying sensitive content compared to the Control. |
| **3. Exploratory hypothesis** | |
| **Hypothesis 3a** | Participants in the Default arm would be significantly less likely to initially select 'Reduced sensitive content' compared to the Control. |

**Figure 4.** Hypotheses.
*Note:* We formulated only one exploratory hypothesis about the impact of defaults based on literature highlighting their expected effects on user behaviour. Exploratory analyses have not been corrected for multiple comparisons. This means that exploratory results should be treated with caution (for hypothesis exploration) rather than formally concluding whether this hypothesis can be rejected or not (hypothesis confirming). For exploratory comparisons we focus more on the direction and magnitude of effects, rather than significance and power.

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

*Default arm*

Defaults introduce a barrier to making an active choice because users can proceed without changing the pre-selected option. Defaults are common on social media platforms. Information provided to participants in the Default and Control arms was the same. However, we expected that compared to the Control participants in the Default arm would be less motivated to engage with the information and more likely to proceed with 'All content types', as it was pre-selected. We expected that this choice would be less active, and therefore, after seeing the feed, participants would be more likely to change their initial choice. Following the completion of this research, we ran an exploratory mini-experiment to investigate the effect of pre-selecting 'Reduced sensitive content', that is reported in Supplementary materials G.[3]

*Information salience arm*

Reducing the number of steps to access information can increase engagement with it (Rosenkranz *et al.*, 2017). This intervention tested whether users made a more informed choice when the examples of sensitive content were more salient and easier to access.

*Non-skippable and skippable microtutorial arms*

Microtutorials are short step-by-step online guides. Unlike nudges that steer decisions, microtutorials aim to boost users' capabilities to make their own choices (Hertwig & Grüne-Yanoff, 2017).

The trial tested whether chunking examples of sensitive content into small segments within an interactive microtutorial could help users align their settings with preferences. We included both a non-skippable and skippable microtutorial arm, as both types are used by online platforms. All steps of the microtutorials are in Supplementary materials B1.

### *Experimental flow and simulated social media platform*

We tested these interventions on a simulated social media platform called WeConnect. To improve authenticity and external validity, WeConnect reflected the design of real social media platforms. The platform had two main components: (1) a sign-up process and (2) a content feed. During sign-up, participants chose their content settings, selecting between 'All content types' and 'Reduced sensitive content'. The sign-up included other steps that mirrored the real-world process but were not included in the analysis (Supplementary materials B2).

The feed contained 24 content pieces: six short videos, six long text posts and 12 short text posts. Depending on the setting chosen during sign-up, they saw either 12 ('All content types') or two ('Reduced sensitive content') sensitive pieces, covering hate, violence and misinformation (Figure 5). See Supplementary materials C for details on content selection (C1) and ethics and safeguarding (C2).

Figure 6 illustrates the flow of the experiment.

Participant interactions (e.g. liking a post) were not visible to others. After engaging with the feed, participants proceeded to the review stage (Figure 3), and the

[3]Unfortunately, it was not feasible to include this extra trial arm in the main experiment in the first place.

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

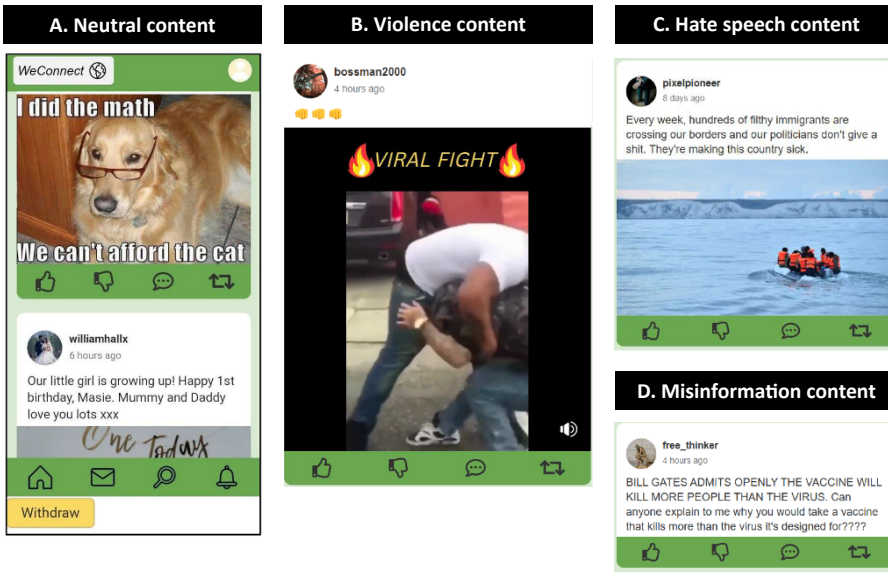


**Figure 5.** WeConnect content feed examples.
*Note:* The researchers developed the posts by drawing inspiration from existing social media content, without replicating it verbatim. These posts do not reflect the views of the researchers or the organisations involved in the research. See Supplementary materials C1 for further details on content sourcing and C2 on ethical considerations and safeguarding measures.

post-trial survey, which included a comprehension test (Figure 7) and further questions (Supplementary materials E6, E8 and E9).

To address potential issues concerning platform functionality, intervention design and data collection, we conducted user testing and a soft launch prior to fieldwork (Supplementary materials D).

### *Sample and data collection*

We recruited a nationally representative sample of adult internet users from the UK. Our final sample comprised 3,500 UK adults (18+) recruited through the panel aggregator Lucid between 24 November and 14 December 2023. Further details in Supplementary materials include power calculations[4] (E1), data collection (E2) and sample demographics (Table S.2).

### *Analysis strategy*

We followed a pre-specified analysis framework that was approved before trial data were collected.[5]

[4]Since participants are recruited from an online panel, we couldn't access individuals who are not online. However, this does not pose a problem for our research, as our target population consists of people who are present online and either currently use or will use social media in the future.

[5]Our pre-specified analysis framework was internal but followed the principles and sections outlined in official pre-registration guidelines. It was reviewed by researchers from outside of the project team and approved by stakeholders across the two organisations involved in the project.

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

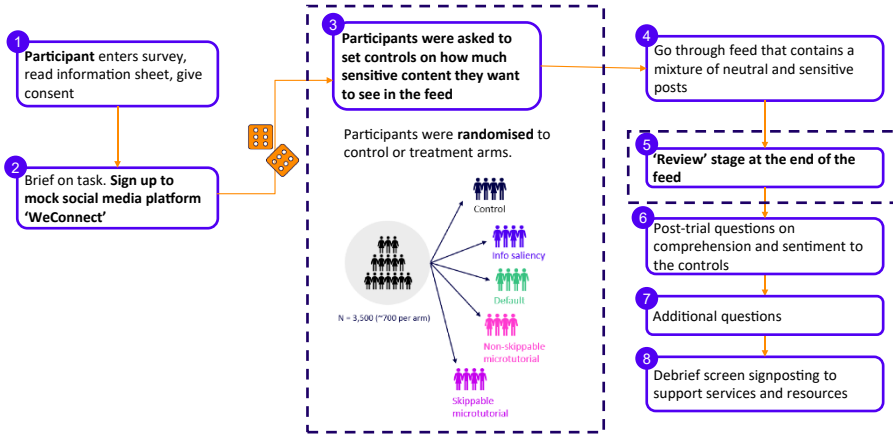


**Figure 6.** Experiment flow.
*Note:* Overview of the participant journey and flow of the experiment. At stage 1, participants enter the survey, review the information sheet and give their consent to proceed to the study. As part of the consent process, participants are informed that they might see sensitive content. At stage 2, participants are asked to sign up and interact with the fake social media platform 'WeConnect' as they would normally do on any social media platform they use. At stage 3, they are randomly allocated to the control or one of the four treatment arms. At this stage they are asked to set controls on how much sensitive content they want to see in their feed. At stage 4, they are instructed to go through 'WeConnect's' feed. In the feed they see 24 posts, a max of 12 of which are sensitive, reportable content. In the feed, participants can like, dislike, comment or reshare posts. At stage 5, when they reach the end of the feed, they are asked if they would like to continue with their current option or would like to change it. At stage 6, participants are asked to answer a series of post-trial questions related to comprehension, sentiment and experience of using content controls. As part of the comprehension test, participants are asked to classify eight items of content as either sensitive or not sensitive. At stage 7, participants are asked some additional questions related to their social media use and socioeconomic status, as part of the standard demographics collected. Finally, at stage 8, participants are provided with the debrief screen containing more details about the purpose of the trial and signposting to support services and resources.

Which of the following would you describe as sensitive content?

| | Not sensitive content | Sensitive content |
|---|---|---|
| A short text post in which a politician describes immigrants as 'violent animals' | ○ | ○ |
| A selfie video in which a man argues that women should not have the right to drive a car | ○ | ○ |
| A link to a news article covering the outcome a lawsuit about election fraud | ○ | ○ |
| A video of a drunk user chugging three beers in a row | ○ | ○ |
| A link to a false blogpost about side effects of vaccines | ○ | ○ |
| A photo of a dog chasing a cat | ○ | ○ |
| A video of teenagers fighting on a playground | ○ | ○ |
| A photo of a woman holding a Union Jack flag | ○ | ○ |

**Figure 7.** Comprehension test.
*Note:* After interacting with the feed, participants were asked a comprehension question which involved categorising 8 descriptions of posts as either sensitive or not sensitive content.

For outcomes with binary data (primary outcome and exploratory analyses), we conducted logit regressions. For outcomes with count data (secondary outcomes), we conducted Poisson regressions. For all models, our predictor variable was the treatment

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

variable with the Control arm as the baseline, and we included age, gender, income, education, ethnicity and platform use as covariates. We used a significance level of 5% throughout. To control the false discovery rate, we corrected for four comparisons across the primary outcome and eight comparisons across secondary outcomes using the Benjamini–Hochberg adjustment (no adjustments were made for exploratory analyses). Data analysis was conducted in R. Further details on the analytical strategy are presented in Supplementary materials E.

## Results

### *Primary analysis: whether participants continue with their initial choice*

In the Control arm (the 'Basic presentation'), 87% of participants maintained their initial choice after viewing the feed (Figure 8). After correcting for multiple comparisons, we found no significant differences between the Control and any of the intervention arms ($p > 0.05$; Table 1). This did not provide evidence to support our hypotheses as we expected that compared to Control, the proportion of participants continuing with their initial choice would be lower in the Default arm and higher in all other intervention arms (Hypotheses 1a–d).

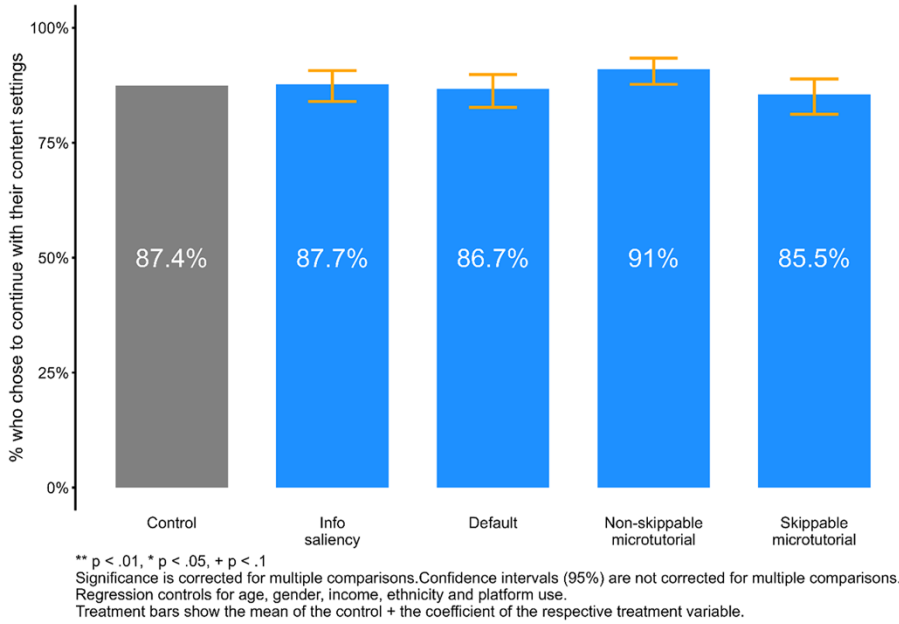


**Figure 8.** Primary analysis comparing the percentage of participants who chose to continue with their content settings in the Control arm to each intervention arm.
*Note:* Error bars indicate 95% confidence intervals.

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

**Table 1.** Primary analysis model output

| Predictor | Beta | SE | Z value | p value | 95% CI lower | 95% CI upper | Odds ratio (OR) | Benjamini–Hochberg adjusted p value |
|---|---|---|---|---|---|---|---|---|
| (Intercept) | 1.40 | 0.29 | 4.85 | **0.00** | 0.85 | 1.98 | 4.06 | NA |
| Info saliency | 0.03 | 0.16 | 0.21 | 0.83 | −0.28 | 0.34 | 1.03 | 0.83 |
| Default | −0.06 | 0.16 | −0.39 | 0.70 | −0.37 | 0.25 | 0.94 | 0.83 |
| Non-skippable microtutorial | 0.37 | 0.18 | 2.13 | **0.03** | 0.03 | 0.72 | 1.45 | 0.13 |
| Skippable microtutorial | −0.16 | 0.16 | −1.04 | 0.30 | −0.47 | 0.14 | 0.85 | 0.59 |
| Age 55 and over | 0.11 | 0.13 | 0.90 | 0.37 | −0.13 | 0.36 | 1.12 | NA |
| Age under 25 | 0.32 | 0.19 | 1.73 | 0.08 | −0.03 | 0.71 | 1.38 | NA |
| Male | 0.48 | 0.11 | 4.40 | **0.00** | 0.27 | 0.70 | 1.62 | NA |
| Other | 0.59 | 0.75 | 0.79 | 0.43 | −0.66 | 2.43 | 1.80 | NA |
| Income less than £40,000 | −0.10 | 0.11 | −0.89 | 0.37 | −0.31 | 0.12 | 0.91 | NA |
| Ethnicity Black | 0.42 | 0.34 | 1.23 | 0.22 | −0.24 | 1.11 | 1.52 | NA |
| Ethnicity other | −0.33 | 0.34 | −0.98 | 0.33 | −0.99 | 0.34 | 0.72 | NA |
| Ethnicity White | −0.07 | 0.23 | −0.33 | 0.74 | −0.55 | 0.36 | 0.93 | NA |
| Education no degree | 0.13 | 0.12 | 1.07 | 0.28 | −0.11 | 0.36 | 1.13 | NA |
| Education prefer not to say | 0.30 | 0.33 | 0.91 | 0.36 | −0.31 | 0.98 | 1.35 | NA |
| Social media use | 0.02 | 0.01 | 2.86 | **0.00** | 0.01 | 0.03 | 1.02 | NA |

### *Secondary analysis: comprehension of what constitutes sensitive content*

After viewing the feed, participants were asked to categorise eight items of content as either sensitive or not sensitive. On average, participants correctly categorised 5.87 pieces of content (Figure 9, Table 2).[6] None of the treatment arms resulted in significant changes from the Control ($p > 0.05$), even though only five participants in the Control arm clicked 'Learn more' and thus had an opportunity to read the sensitive content examples. This did not provide evidence to support our hypotheses as we expected that the probability of correctly classifying content would be lower in the Default arm and higher in all other intervention arms compared to the Control (Hypotheses 2a–d).

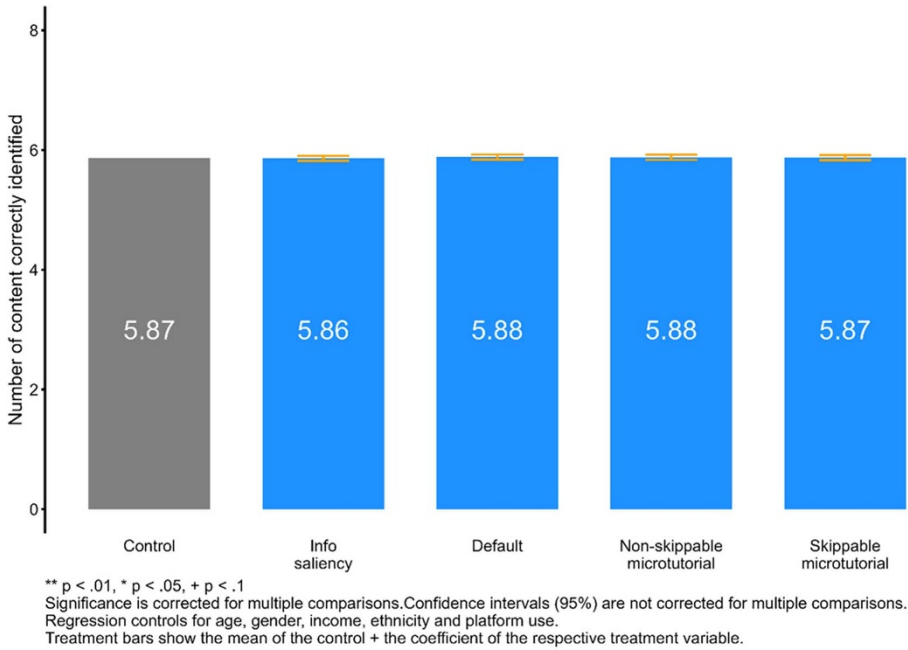


**Figure 9.** Secondary analysis, comparing the content participants correctly identified as sensitive or not sensitive in the Control arm to each intervention arm.
*Note:* Error bars indicate 95% confidence intervals.

[6]We acknowledge that for ethical reasons our trial did not include the more extreme types of sensitive content. Including such content would likely have resulted in a higher score. However, our main question of interest related to the impact of trial arms on comprehension rather than the absolute level.

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

**Table 2.** Secondary analysis (comprehension) model output

| Predictor | Beta | SE | Z value | P value | 95% CI lower | 95% CI upper | Rates ratio (RR) | Benjamini–Hochberg adjusted p value |
|---|---|---|---|---|---|---|---|---|
| (Intercept) | 1.70 | 0.04 | 43.12 | **0.00** | 1.63 | 1.78 | 5.50 | |
| Info saliency | −0.01 | 0.02 | −0.29 | 0.78 | −0.05 | 0.04 | 0.99 | 1 |
| Default | 0.01 | 0.02 | 0.66 | 0.51 | −0.03 | 0.06 | 1.01 | 1 |
| Non-skippable microtutorial | 0.01 | 0.02 | 0.49 | 0.63 | −0.03 | 0.05 | 1.01 | 1 |
| Skippable microtutorial | 0.00 | 0.02 | 0.23 | 0.82 | −0.04 | 0.05 | 1.01 | 1 |
| Age 55 and over | 0.02 | 0.02 | 0.94 | 0.35 | −0.02 | 0.05 | 1.02 | |
| Age under 25 | 0.00 | 0.02 | 0.17 | 0.87 | −0.04 | 0.05 | 1.00 | |
| Male | −0.03 | 0.01 | −2.07 | **0.04** | −0.06 | 0.00 | 0.97 | |
| Other | −0.01 | 0.09 | −0.13 | 0.90 | −0.19 | 0.15 | 0.99 | |
| Income less than £40,000 | −0.03 | 0.01 | −1.76 | 0.08 | −0.05 | 0.00 | 0.97 | |
| Ethnicity Black | 0.05 | 0.04 | 1.12 | 0.26 | −0.03 | 0.13 | 1.05 | |
| Ethnicity other | 0.03 | 0.05 | 0.52 | 0.61 | −0.07 | 0.12 | 1.03 | |
| Ethnicity White | 0.07 | 0.03 | 2.38 | **0.02** | 0.01 | 0.14 | 1.08 | |
| Education no degree | 0.01 | 0.02 | 0.52 | 0.60 | −0.02 | 0.04 | 1.01 | |
| Education prefer not to say | −0.02 | 0.04 | −0.54 | 0.59 | −0.11 | 0.06 | 0.98 | |
| Social media use | 0.00 | 0.00 | 1.24 | 0.22 | 0.00 | 0.00 | 1.00 | |

### *Exploratory analysis: initial choice*

Overall, 24% of participants across all trial arms chose 'Reduced sensitive content' at sign-up. The Information salience intervention significantly increased the proportion of participants making this choice compared to the Control (29.4% vs 24.4%, $p < 0.05$). Conversely, the Default significantly reduced the proportion making this choice compared to the Control (14.9% vs 24.4%, $p < 0.01$, Figure 10, Table 3), in line with Hypothesis 3a.

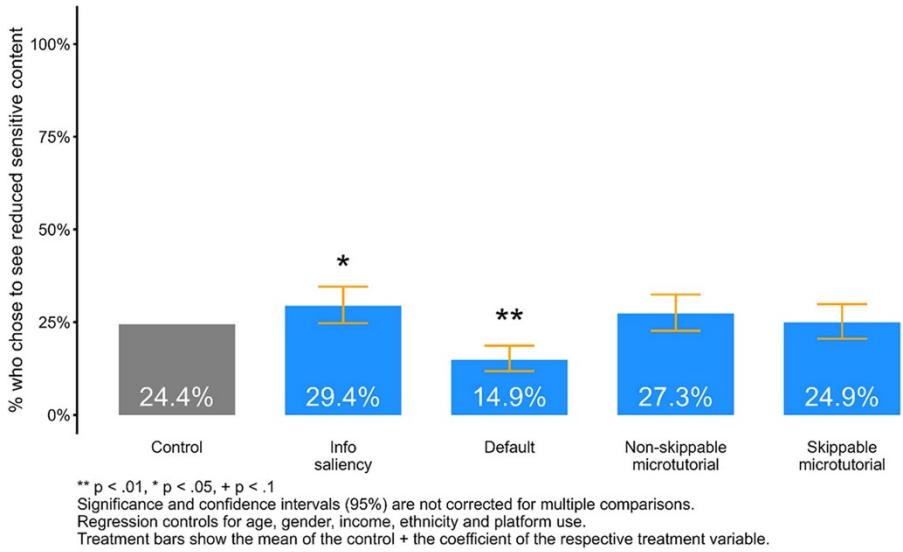


**Figure 10.** Exploratory analysis, comparing the percentage of participants who chose to see reduced sensitive content in the Control arm to each intervention arm.
*Note:* Error bars indicate 95% confidence intervals.

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

**Table 3.** Exploratory analysis (initial choice) model output

| Predictor | Beta | SE | Z value | P value | 95% CI lower | 95% CI upper | Odds ratio (OR) | Cohen's d effect size |
|---|---|---|---|---|---|---|---|---|
| (Intercept) | 0.27 | 0.22 | 1.20 | 0.23 | −0.17 | 0.70 | 1.30 | 0.15 |
| Info saliency | 0.25 | 0.12 | 2.10 | **0.04** | 0.02 | 0.49 | 1.29 | 0.14 |
| Default | −0.61 | 0.14 | −4.45 | **0.00** | −0.89 | −0.35 | 0.54 | −0.34 |
| Non-skippable microtutorial | 0.15 | 0.13 | 1.19 | 0.23 | −0.10 | 0.40 | 1.16 | 0.08 |
| Skippable microtutorial | 0.03 | 0.13 | 0.21 | 0.83 | −0.22 | 0.28 | 1.03 | 0.01 |
| Age 55 and over | 0.33 | 0.10 | 3.36 | **0.00** | 0.14 | 0.52 | 1.39 | 0.18 |
| Age under 25 | −0.08 | 0.14 | −0.59 | 0.55 | −0.37 | 0.19 | 0.92 | −0.05 |
| Male | −0.87 | 0.09 | −9.99 | **0.00** | −1.05 | −0.70 | 0.42 | −0.48 |
| Other | −0.09 | 0.48 | −0.20 | 0.84 | −1.10 | 0.81 | 0.91 | −0.05 |
| Income less than £40,000 | −0.07 | 0.09 | −0.84 | 0.40 | −0.24 | 0.10 | 0.93 | −0.04 |
| Ethnicity Black | 0.20 | 0.23 | 0.85 | 0.39 | −0.26 | 0.65 | 1.22 | 0.11 |
| Ethnicity other | −0.38 | 0.28 | −1.33 | 0.18 | −0.95 | 0.17 | 0.68 | −0.21 |
| Ethnicity White | −0.31 | 0.17 | −1.78 | 0.07 | −0.64 | 0.04 | 0.73 | −0.17 |
| Education no degree | −0.30 | 0.09 | −3.27 | **0.00** | −0.48 | −0.12 | 0.74 | −0.17 |
| Education prefer not to say | −0.15 | 0.24 | −0.63 | 0.53 | −0.64 | 0.31 | 0.86 | −0.08 |
| Social media use | −0.04 | 0.00 | −7.73 | **0.00** | −0.05 | −0.03 | 0.96 | −0.02 |

### *Exploratory descriptives: reasons for decision to change or keep setting*

Participants could select multiple response options from the list provided. Among participants who continued with the original choice ($n = 3{,}069$), the most popular reasons included thinking it was the right option for them (48%), the content matching their expectations (34%) and liking the content they saw (26%). These were the top three reasons regardless of the choice they decided to keep (Supplementary materials Tables S.5–S.7).[7]

The most popular reason for changing 'All content types' to 'Reduced sensitive content' ($n = 320$) was because they saw content that upset them (43%) (Supplementary materials Table S.9) whereas the top reason for changing to 'All content types' ($n = 111$) was because they were curious to see what would change (65%) (Supplementary materials Table S.10; aggregated data in Table S.8).

### *Exploratory descriptives: skipping the microtutorial*

Of the 664 participants in the Skippable microtutorial arm, 73% skipped the tutorial (reasons in Supplementary materials Table S.11).

Additional results in Supplementary materials F include goodness of fit test (F1), understanding of choice options (F3), sentiment (F2, F4–F7), behaviour on the platform (F8), sentiment to non-skippable microtutorial (F12), comprehension by post topic (F13), previous experience with content controls (F14), results of ordinal models (F15) and post-hoc mini-experiment (G).

## Discussion and conclusion

We aimed to determine whether changes to the online choice architecture could help or hinder users' ability to align content settings with their preferences.

We found that interventions affected the initial choice at the sign-up stage. In the Control arm, where participants made an active choice, 24% chose 'Reduced sensitive content'. In the Default arm, where 'All content types' were pre-selected, only 15% chose 'Reduced sensitive content'. The difference could be driven by inertia and the ease of staying with the default option, the perception that this is the recommended choice or because it represents the status quo (Jachimowicz *et al.*, 2019).

In contrast, more participants chose 'Reduced sensitive content' when examples were shown on the decision page (29%) compared to the Control group (24%). This resonates with information security research showing that threatening wording and visually salient messages increase users' likelihood of declining cookies (Ebert, Ackermann, & Bearth, 2022). In our case, the perceived threat level may have been higher in the Information salience group, as all participants saw the sensitive content examples, unlike the Control group, where only five clicked 'Learn more' and saw the examples.

After viewing the feed, participants were asked if the choice was still working for them (Figure 3). Surprisingly, the proportion of participants maintaining their initial

[7]Statistical significance tests were not conducted for exploratory descriptives.

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

choice remained similar across all interventions and the Control group, despite differences in initial selections. This lack of significant impact from the interventions may be affected by the high baseline, with 87% staying with the initial choice in the Control arm. The most common reasons for keeping or changing the initial choice aligned with our assumptions when designing the experiment. Participants kept their initial choice if it suited them, switched to a 'safer' option if the content upset them and switched to a less 'safe' option out of curiosity about potential changes.

Notably, the highest proportion of participants keeping their initial choice was in the Non-skippable microtutorial arm (91%), although it was not statistically significant after multiple comparisons correction. We expected that the microtutorial would make users pause and reflect on the information, and make a more informed initial choice, increasing the likelihood of sticking to it after the feed. This finding would have aligned with research, which demonstrated that microtutorials were an effective intervention for online safety by significantly boosting user reporting (Ofcom, 2023).

In addition to the high baseline, differences in context, nature (capability-building vs information provision) and psychological mechanisms (prompting an action vs shaping choice) could also explain the discrepancy between our findings and previous research. The lack of impact from the Skippable microtutorial is likely due to the fact that 73% of participants skipped it, effectively placing them in the Control condition. Overall, our findings indicate that the initial choice was 'sticky' even when participants were given a low-effort opportunity to revise it.[8]

The lack of impact of the interventions on comprehension may be driven by participants having their own pre-existing understanding of what constitutes sensitive content.[9] Another explanation may be that our interventions targeted attention rather than comprehension. Users know what sensitive content is but do not consider it when making online choices. Information salience was effective as it increased user attention, and not knowledge, leading to different initial choices.

### *Limitations*

The main limitations relate to the simulated environment which may not fully replicate real social media users' incentives and motivations, and the absence of more personalised or harmful content. Moreover, the short experiment timescale limits conclusions on long-term effects. Thus, our confidence lies more in the relative impact of interventions than in precise measures of their magnitude.

## Conclusion

These findings offer two key lessons for policymakers and regulators to enhance user experience and safety across digital platforms. Firstly, small changes in how content controls are presented may significantly influence user choice. This underscores

[8]Changing or keeping the initial choice during the review stage required the same level of effort – one click.

[9]We recognise that the phrasing of the question, 'Which of the following would you describe as sensitive content?' may have led participants to interpret it based on their personal understanding of sensitive content, rather than the platform's definition.

https://doi.org/10.1017/bpp.2025.10016 Published online by Cambridge University Press

the importance of platforms designing initial choices to support informed decision-making. Secondly, users tend to stick with their existing setting, even when offered easy opportunities to change. We hypothesise that some users may not have strong preferences regarding the two choice options offered. They may prefer more tailored methods, such as hiding an individual post. Future research could explore such control mechanisms.

**Supplementary material.** To view supplementary material for this article, please visit https://doi.org/10.1017/bpp.2025.10016

**Acknowledgements.** We would like to thank our colleagues at Ofcom and BIT for their support and contributions, including Amor Perez Pavon, Bobby Stuijfzand, Deborah Mc Crudden, Eva Kolker, John Ivory, Johnny Sutton, Jonathan Porter, Rhian Armstrong, Riccardo D'Adamo and Zak Soithongsuk.

**Funding statement.** Ofcom provided funding for this work.

**Competing interests.** The authors declare no competing interests.

**Data availability statement.** Data and code can be provided upon request. Please email the corresponding author.

## References

Atkins, L., J. Francis, R. Islam, D. O'Connor, A. Patey, N. Ivers, R. Foy, E. M. Duncan, H. Colquhoun, J. M. Grimshaw, R. A. Lawton and S. Michie (2017), 'A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems', *Implementation Science*, **12**: 1–18.

Bauer, J. M., R. Bergstrøm and R. Foss-Madsen (2021), 'Are you sure, you want a cookie? – The effects of choice architecture on users' decisions about sharing private online data', *Computers in Human Behavior*, **120**: 106729.

Busch, C. and A. Fletcher (2024). *Harmful Online Choice Architecture*. Centre on Regulation in Europe.

Centre for Data Ethics and Innovation. (2020). *Online targeting: final report and recommendations*. Centre for Data Ethics and Innovation.

Competition & Markets Authority (2022), *Online Choice Architecture: How Digital Design Can Harm Competition and Consumers*, Competition & Markets Authority. https://assets.publishing.service.gov.uk/media/624c27c68fa8f527710aaf58/Online_choice_architecture_discussion_paper.pdf, Accessed 10 October, 2025.

Ebert, N., K. A. Ackermann and A. Bearth (2022), 'When information security depends on font size: how the saliency of warnings affects protection behavior', *Journal of Risk Research*, **26**(3): 233–255.

Godefroid, M.-E., R. Plattfaut and B. Niehaves (2023), 'How to measure the status quo bias? A review of current literature', *Management Review Quarterly*, **73**(4): 1667–1711.

Hertwig, R. and T. Grüne-Yanoff (2017), 'Nudging and boosting: Steering or empowering good decisions', *Perspectives on Psychological Science*, **12**(6): 973–986.

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

Mertens, S., M. Herberz, U. J. Hahnel, and T. Brosch (2022), 'The effectiveness of nudging: a meta-analysis of choice architecture interventions across behavioral domains', *Proceedings of the National Academy of Sciences*, **119**(1): e2107346118.

Michie, S., M. Van Stralen and R. West (2011), 'The behaviour change wheel: a new method for characterising and designing behaviour change interventions', *Implementation Science*, **6**: 1–12.

Ofcom (2023), *Boosting Users' Safety Online: Microtutorials*, Ofcom. https://www.ofcom.org.uk/online-safety/safety-technology/boosting-users-safety-online-microtutorials, Accessed 10 October, 2025.

Ofcom (2024), *Terms and Conditions and Content Controls*, Ofcom. https://www.ofcom.org.uk/media-use-and-attitudes/online-habits/terms-and-conditions-and-content-controls, Accessed 10 October, 2025

Rosenkranz, S., K. Vringer, T. Dirkmaat, E. van den Broek, C. Abeelen and A. Travaille (2017), 'Using behavioral insights to make firms more energy efficient: a field experiment on the effects of improved communication', *Energy Policy*, **108**: 184–193.

Samuelson, W. and R. Zeckhauser (1988), 'Status quo bias in decision making', *Journal of Risk and Uncertainty*, **1**: 7–59.

Thaler, R. H., C. R. Sunstein and J. P. Balz (2014). Choice Architecture. *The Behavioral Foundations of Public Policy*.

**Cite this article:** Shakhina, N., P. Skotida, S. Krishnan-Barman, M. Wessel, E. Meyer Zu Brickwedde, T. House and R. Gill (2025), 'How does the design of social media content controls shape users' choice? Evidence from an online experiment', *Behavioural Public Policy*, 1–18. https://doi.org/10.1017/bpp.2025.10016