Protective defaults for social media platforms

A randomised controlled trial testing retention of defaults for children and adults

Behavioural Insights Hub Insights Paper

13 July 2026

The Discussion / Insights paper series

Ofcom is committed to encouraging debate on all aspects of communications regulation and to create rigorous evidence to inform that debate. One of the ways we do this is through publishing a series of research 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.

Disclaimer

Discussion and Insights papers contribute to the work of Ofcom by providing rigorous research and encouraging debate in areas of Ofcom's remit. These 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.

If you would like to comment on this Insights paper, please email to:
behavioural.insights@ofcom.org.uk

Contents

Section

Overview

N.B This work was produced before HM Government's announcement on June 15th about forthcoming measures on social media access for under 16s.

Background

- A default is a choice that has already been selected for the user. In other words, it is the outcome a user will receive unless they actively choose an alternative. Extensive research has shown that defaults can be powerful drivers of behaviour, often leading people to retain the pre-selected option.
- Several previous trials conducted by Ofcom's Behavioural Insights Hub have found that both adults and children are likely to retain default settings. Depending on how those defaults are configured, this can either improve online safety, for example, by reducing exposure to harmful content, or have the opposite effect.
- However, these earlier trials examined adults and children separately, making it difficult to directly compare how each group responds to defaults. Given evidence of developmental differences between children and adults, particularly in relation to risk-taking and decision making, this study aimed to test whether default safety settings influence the behaviour of these groups in different ways.

Method

- We conducted an online randomised controlled trial with 8,918 UK participants (3,201 adults and 5,717 children). Participants interacted with a mock social media feed. During this, they received a prompt that their settings had been updated, along with an invitation to review these settings.
- Participants who chose to review their settings were presented with three safety-related settings configured to their most protective level by default: location sharing, network expansion, direct messaging. Participants were randomly assigned to different versions of the settings screen that varied in the level of supportive information provided and amount of friction involved in changing settings.
- The study aimed to assess the extent to which participants retained these default settings and whether responses differed between adults and children.

Key Findings

- In each group, the majority of participants viewed and retained all of their safety defaults, with no differences between adults and children. None of our supportive messaging, or friction interventions further increased default retention.
- Retention rates were the same across all types of setting, which could indicate a broad tendency to retain defaults or maintain the status quo, rather than setting-specific preferences.
- The post-trial survey found broad support across all age groups for both default and fixed safety settings for under-18s, suggesting that participants generally recognised the value of these measures for online safety.
- While retention of defaults was high, findings should be interpreted with caution. Although mock online trials can provide useful insights into likely behaviour, they cannot fully replicate the real-world influences on decision-making, such as social norms and peer influence.

Policy and Research Context[1]

As an evidence-based regulator, Ofcom continues to incorporate behavioural research into its evidence base to inform policy development. Ofcom's Behavioural Insights Hub leads a programme of research that includes running randomised controlled trials (RCTs) to rigorously test and evaluate the effectiveness of different online safety interventions. These trials help uncover what works, what doesn't, and how the design of content controls affects user engagement with them. To date we have undertaken research on a range of topics including reporting mechanisms and alert messages, micro-tutorials, encouraging engagement with terms of service, high-level content control tools, content control tools which apply at the level of individual posts, and the use of verification badges like the classic "blue tick".

Defaults and public policy

In behaviourally informed public policy, defaults are a form of intervention that have been widely applied across a range of policy domains to good effect. A "default" refers to a choice pre-selected for the user; in other words, it is the outcome the user will get unless they actively make a different choice. Historically, behavioural science research has shown that people are often pre-disposed to retain default options. This has been shown not only in lab-based settings,[2] but also in real-world decision contexts such as organ donation[3], and retirement saving[4]. In short: the default option acts as a sort of "path of least resistance" for users, a quick option that reduces the effort needed to make a choice. Defaults are even perceived as "recommendations" in contexts where users don't feel they have sufficient knowledge to make an informed active choice.[5]

To that end, defaults are an incredibly important intervention to study in the context of online user empowerment and safety, which cuts across many of Ofcom's policy areas under the Online Safety Act.[6]

Previous Ofcom research on defaults

Ofcom's Behavioural Insights Hub has carried out several RCTs on the topic of defaults. These trials have tested the impact of defaults in representative social media contexts involving children (<18-

[1] To note: the work detailed in this report was conducted and written up before the UK's June 15th 2026 announcements on banning social media for under 16 year olds.
[2] Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, *1*(1), 7-59.
[3] Johnson, E. J., & Goldstein, D. G. (2004). Defaults and donation decisions. *Transplantation*, *78*(12), 1713-1716.
[4] Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, *112*(S1), S164–S187.
[5] McKenzie, C. R., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, *17*(5), 414-420.
[6] The Online Safety Act ('the Act') requires Ofcom to produce Codes of Practice outlining the measures that providers can take to comply with certain duties specified in the Act. These include the 'user empowerment' duties, which require providers of Category 1 services to offer adult users in the UK features to e.g. reduce the likelihood of encountering, or alert them to the presence of, certain types of content or to exert control over other aspects of their experience on the platform.

year-olds), and adult users. Below we outline some key examples of these previous trials, and findings.

Trial	Method	Sample	Findings
Autoplay/skip defaults on video-sharing platforms	Users interacted with a mock video-sharing platform where the default varied between auto-playing videos vs. auto-skipping. Users viewed a variety of videos, some of which were flagged to the user as containing harmful content.	2,800 adults	Significantly more people avoided seeing harmful content when those videos auto-skipped. When videos auto-played, similar numbers of people skipped overall, however, substantially more people *viewed* some of the harmful content before skipping.
Defaults for harmful content settings (adults)	Users interacted with a mock social media platform including the sign-up phase, and subsequently viewing a mock feed of posts. As part of sign-up, users choose a level of harmful content exposure with "All Content Types" either defaulted on or not.	Trial 1 – 3,500 adults	Trial 1 found that when "All Content Types" was defaulted, only 15% actively switched to the less harmful "Reduced Sensitive Content" option. When the "Reduced Sensitive Content" was defaulted, 42% of users stuck with it. A majority retained defaults: 88% of users kept their initial default (regardless of the option) after a later prompt gave them the chance to switch
Defaults for harmful content settings (children)	In a similar trial involving children, participants completed a mock sign-up process where the safer option ("Don't recommend harmful content") was either pre-selected by default or not. They were then shown a feed of video posts, although no harmful content was included in the feed.	2,403 children aged 13 – 17	Children were 1.4x more likely to keep the safe option when it was defaulted. However, almost a third of children later opted out of this initial default.

Ultimately, Ofcom's prior trials on defaults show that pre-selected options are strong drivers of choice which, depending on the direction of the default, could have serious implications for users' exposure to harmful content online. However, across these findings, there is some evidence to suggest that children may be more likely than adults to opt out of these types of safety defaults.

Why might further research on safety defaults be necessary?

While Ofcom's previous behavioural research on safety defaults focused primarily on sensitive content filtering, the protective value of defaults is now being considered more broadly. Ofcom's 2024 **Illegal Harms Statement** identified default settings as a frontline defensive mechanism for ensuring children are better protected from child sexual exploitation and abuse, grooming, and unsolicited online contact, with the subsequent **Illegal Content Codes of Practice for User-to-User Services** setting specific instructions on defaults that map directly onto the variables tested in the current trial:

- **Network Connection Settings:** Firms must implement default settings ensuring that child users are neither presented with nor included in network expansion prompts (e.g. "suggested friends") or visible connection lists.
- **Direct Message (DMs) Settings:** For services with user connections, default settings must prevent children from receiving DMs from non-connected users. For services without established connections, child users must be provided with a mechanism to actively confirm a DM before it becomes visible.
- **Location Sharing Settings:** Automated location information displays must be switched off by default for child users.

The Illegal Content Codes also specify a need for behavioural friction before a child user can disable a safety default: platforms must present supportive messaging outlining the potential risks and implications of turning the above defaults off. The experiment reported here aimed to directly test retention rates for the above defaults, and if supporting information played any moderating role. Additionally, Ofcom's 2025 **Women and Girls' Online Safety** guidance highlights that certain safety defaults should be applied universally across *all* user groups as a way of raising the overall floor of digital safety for groups who are at disproportionate risk.

Children vs. Adult Users: Why it matters

A limitation of Ofcom's previous default trials is that none of these trials' samples comprised both adults and children. As such, we are limited in being able to draw strong conclusions on the relative impacts that defaults may have between ages. While similar, the trials noted above were nonetheless different in terms of presentation, and the type of content experienced by the participants. A further limitation is that these trials exclusively tested safety defaults concerning exposure to harmful content.

There is good reason to hypothesise that children and adults' behavioural responses to defaults, particularly in the online space, could strongly differ. Young people, particularly teenagers, may behave differently from adults due to differences in brain development, such as heightened

sensitivity to rewards[7], ongoing development of impulse control[8], and increased susceptibility to peer influence[9].

These developmental factors could shape how children respond to online environments and safety features. For example, younger users' heightened sensitivity to social rewards could mean that their desire for peer validation – meaning a need to be seen, liked, viewed, shared, and contacted – might "override" any notion of retaining safety defaults, potentially increasing the user's risk of being exposed to harm.

Research Objectives

Our aim was to conduct a new trial that addressed the limitations arising out of the previous work reported above. We ran a trial that involved both children, and adult users being exposed to the same procedure, and the same defaults. We expanded the types of defaults included in the trial over what was tested in past Ofcom trials.

Our aims with this trial were to:

- Develop an understanding of **whether children and adults differ in their response** to default safety settings.
- Gain evidence on **which types of default settings** (e.g. network expansion, messaging, location sharing) adults and children are more likely to retain or override.
- Assess the effectiveness of different interventions to encourage adults and children to retain default **safety settings**.

[7] Casey, B. J., Jones, R. M., & Hare, T. A. (2010)**.** Adolescent development of the reward system. *Frontiers in Human Neuroscience*, *2*(6).
[8] Blakemore, S.-J., & Robbins, T. W. (2012). Decision-making in the adolescent brain. *Nature Neuroscience*, *15*(9), 1184–1191.
[9] Gardner, M., & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: An experimental study. *Developmental Psychology*, *41*(4), 625–635.

Trial Design and Setup

From December 2025 – January 2026, 3,201 adults and 5,717 children (aged 13-17)[10] took part in an online randomised controlled trial (RCT). We designed a bespoke mock social media platform to test three different interventions aimed at reducing the likelihood that default safety settings are overridden. The study was designed and delivered in partnership with Verian. Further details on the sample, procedures, and analytical approach are provided in the accompanying Technical Report published alongside this Insights Paper.

Experimental Design

The trial simulated a social media experience, asking participants to use a mock feed in a way that reflected typical platform use. Children and adults took part in identical experiments. Participants encountered a scrollable feed of up to 30 posts varying between purely text-based, image and video-based, designed to mirror a typical social media mix of neutral, age-appropriate content. Participants could interact with the feed and posts by scrolling and reacting (e.g. liking, commenting). All interactions were recorded automatically before participants were routed to the post-trial survey.

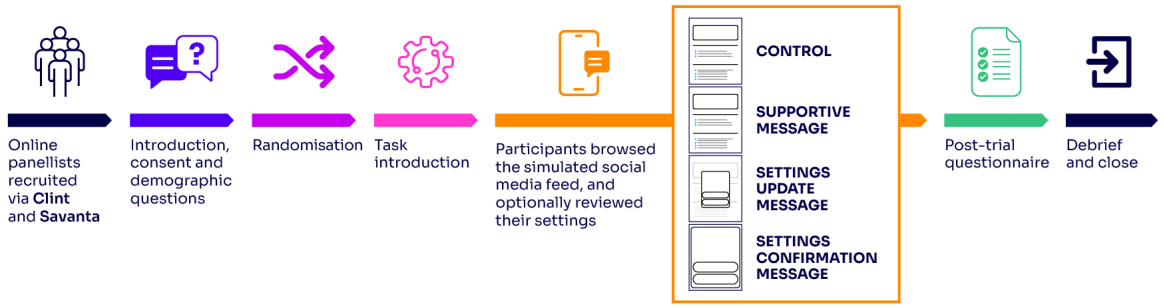
The general flow of the trial for participants was:

- Participants were first introduced to the social media feed and instructed to complete a browsing task in which they scrolled through the feed. After participants viewed the sixth post, they received a pop-up notification on screen that invited them to review their settings. Users could choose to ignore this by clicking an X in the top right corner of the pop-up which would appear after a few seconds' delay.
- If participants chose to review their settings by clicking "Review" on the pop-up notification, they were taken to a settings page which contained three types of settings: location sharing settings, user connection settings, and direct message settings. Each category contained multiple settings designed to mimic those typically seen on common social media platforms. All settings were defaulted on (i.e. protective) for all participants. Participants then interacted with the settings page.
- Participants were randomly allocated to one of four trial arms: a control condition and three intervention conditions. These conditions changed how the settings were displayed.
- After reviewing the settings, participants had to click "Confirm my settings" to return to the main feed.
- After completing the task, all participants completed a post-trial survey capturing their understanding of, and responses to, the settings and notifications shown.

Figure 1 summarises the experimental flow and user journey.

[10] Samples were representative of the UK population.

Figure 1: Experiment flow and user journey.



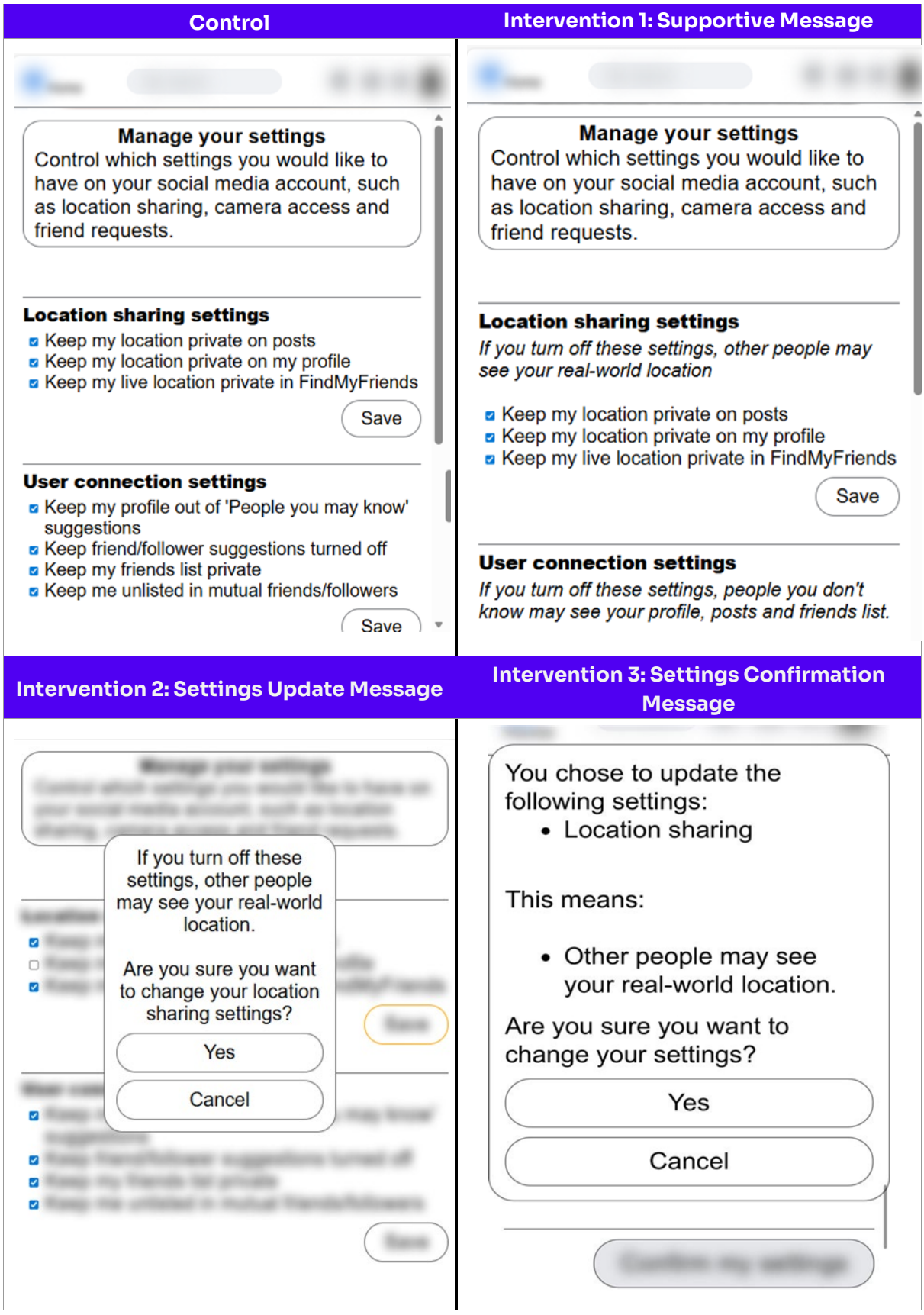
Interventions and Trial Arms

Participants were randomly and evenly distributed to either the control condition, or one of three intervention conditions. The control experience reflected the baseline settings page with no supportive messaging or prompts designed to inform or facilitate user choices. The intervention arms introduced specific changes to how platform settings were presented. These included presenting timely information on the function of the different settings, or in some cases altering how users confirmed their choices. We hypothesised that these interventions may introduce friction, making risks of turning off defaults more salient to users and decreasing the rates of changing defaulted settings.

Specifically, the four trial arms were:

- **Control:** No supportive messaging
    - Participants in the control group experienced the baseline settings page, with no additional supportive messaging or prompts regarding disabling safety settings.
- **Intervention 1**: Supportive message
    - Participants in the “supportive message” arm saw a message displayed underneath each settings category, which informed the user about the implications of disabling the default safety settings.
- **Intervention 2**: Settings update message
    - Participants in the “settings update message” arm were shown a pop-up with the message used in the “supportive message” arm for any of the three categories in which they de-selected any default setting.
    - This means participants were shown up to three of these, depending on the number settings categories for which they opted out.
- **Intervention 3**: Settings confirmation message
    - For participants in the “settings confirmation message” the process of deactivating a default was:
        i. changing the setting,
        ii. clicking “Save” for that category, and
        iii. making final confirmation of any and all changes by accepting a “Are you sure you want to change your settings?” message that appeared before users could exit the settings screen.

Figure 2: Experimental trial arms and associated interventions



The direct message portion of the settings screen read: "Direct Message settings: *If you turn off these settings, you may receive messages from people you don't know.*

- Only show messages from people I am connected with
- Warn me when a group includes people I’m not connected with”

Key Insights

1. Both children and adult participants viewed and retained their defaulted settings on at high rates

During the experiment, around 75% of adults viewed their settings, consistent across all experimental arms. A similar percentage of participants in the under-18 sample viewed their settings at some point during the experiment (all under-18s: 77%; young people aged 16–17: 79%; children aged 13–15: 75%).

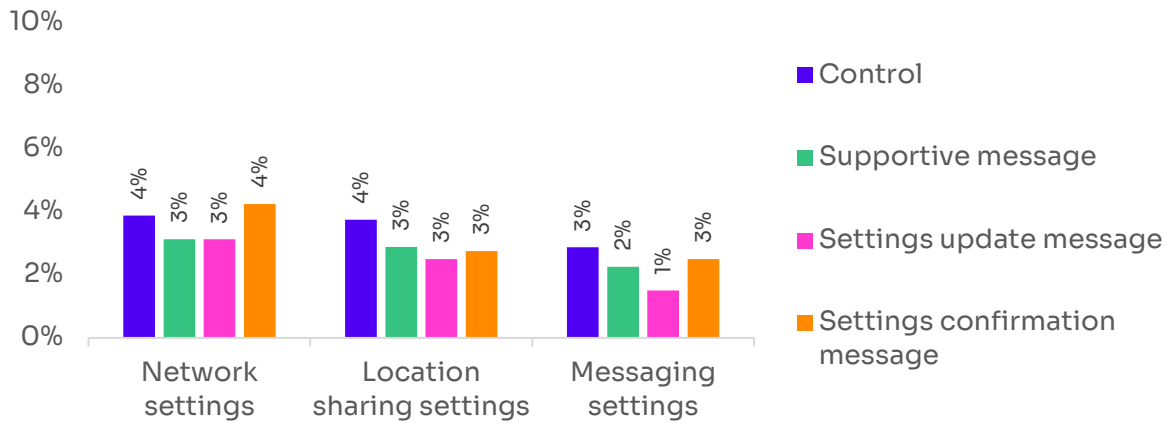
Across these participants, a vast majority had retained all their default settings by the end of the experiment. This was true for adults (92% who checked their settings retained all defaults), children aged 16–17 (also 92%), and children aged 13–15 (and also 92%). Taken together, these findings reinforce previous Ofcom findings, and others, that default settings are largely "sticky" for both children and adults, even when users actively view or engage with their settings.

However, aspects of the experimental design may have influenced participants' behaviour. First, participants were required to wait a few seconds before they could dismiss the initial prompt to review their settings, which may have led to greater attention than in real-world environments where such prompts can typically be skipped immediately. This otherwise increased salience of this active prompt could have raised the likelihood that participants checked and engaged with their default settings. Second, the use of a mock social media platform may have reduced participants' motivation to engage with their settings and change their default settings compared to real-world contexts, where choices have more meaningful personal consequences. For example, the platform did not include features such as messaging or the ability to share posts, so we were not able to test whether exposure to such features would affect participants' decisions about their settings in this trial.

2. Participants retained their defaults at similar rates across the three setting categories

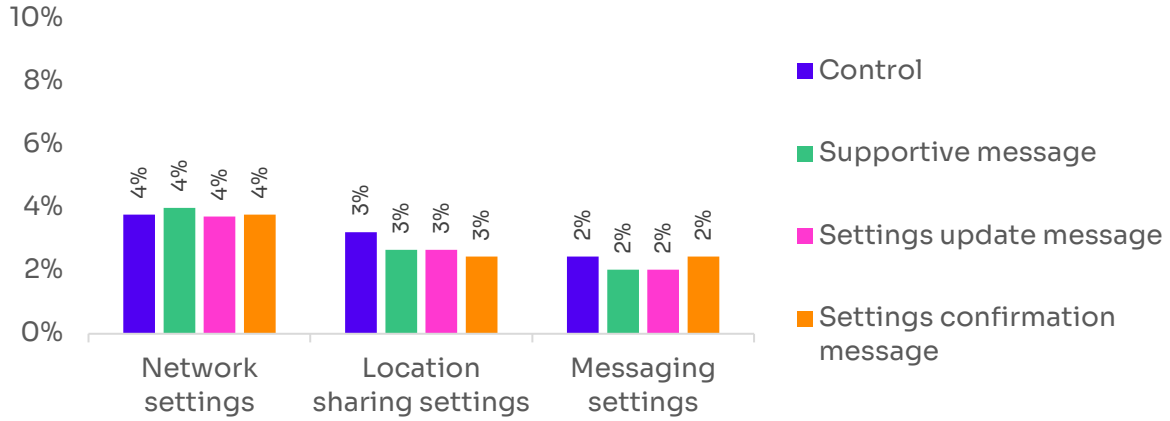
Participants who viewed their settings were free to opt-out of defaults in any or all of the three categories of setting (network settings, location sharing, and messaging). Retention rates were similarly high across all three categories (e.g. no one category had opt-out rates greater than 4% in the adult sample.)

Figure 3: Proportion of all participants who changed at least one safety setting, for each of the three settings categories



We found very similar retention rates for children where no category again had opt-out rates greater than 4%. There was no significant difference in retention rates between the different settings.

Figure 4: Proportion of all children who changed at least one safety setting, for each of the three settings categories



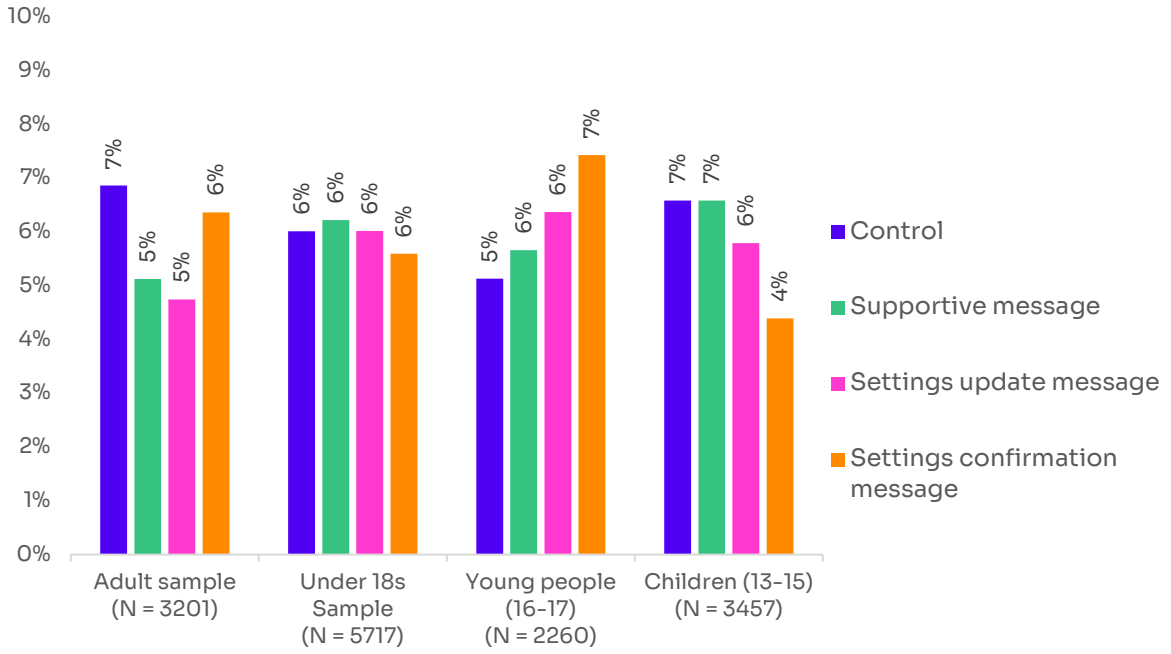
3. There was no differences in default retention rates due to the trial interventions

The experiment tested three different interventions which changed how the settings were presented, including the amount of supporting information and how choices were "saved". There were no statistically significant differences in the number of participants retaining default settings between any experimental arms. This was true for both adults and children (see Figure 5). As such, we have no clear evidence that the interventions we tested meaningfully affected the retention of default settings for either adults or children.

This is likely, in large part, owing to the finding that only a small proportion of participants (around 5%) changed any settings, which limits the ability to detect statistically significant differences between groups. As noted above, the experimental design also does not fully replicate real-world

platform use, making it difficult to draw firm conclusions about the effectiveness of specific interventions.

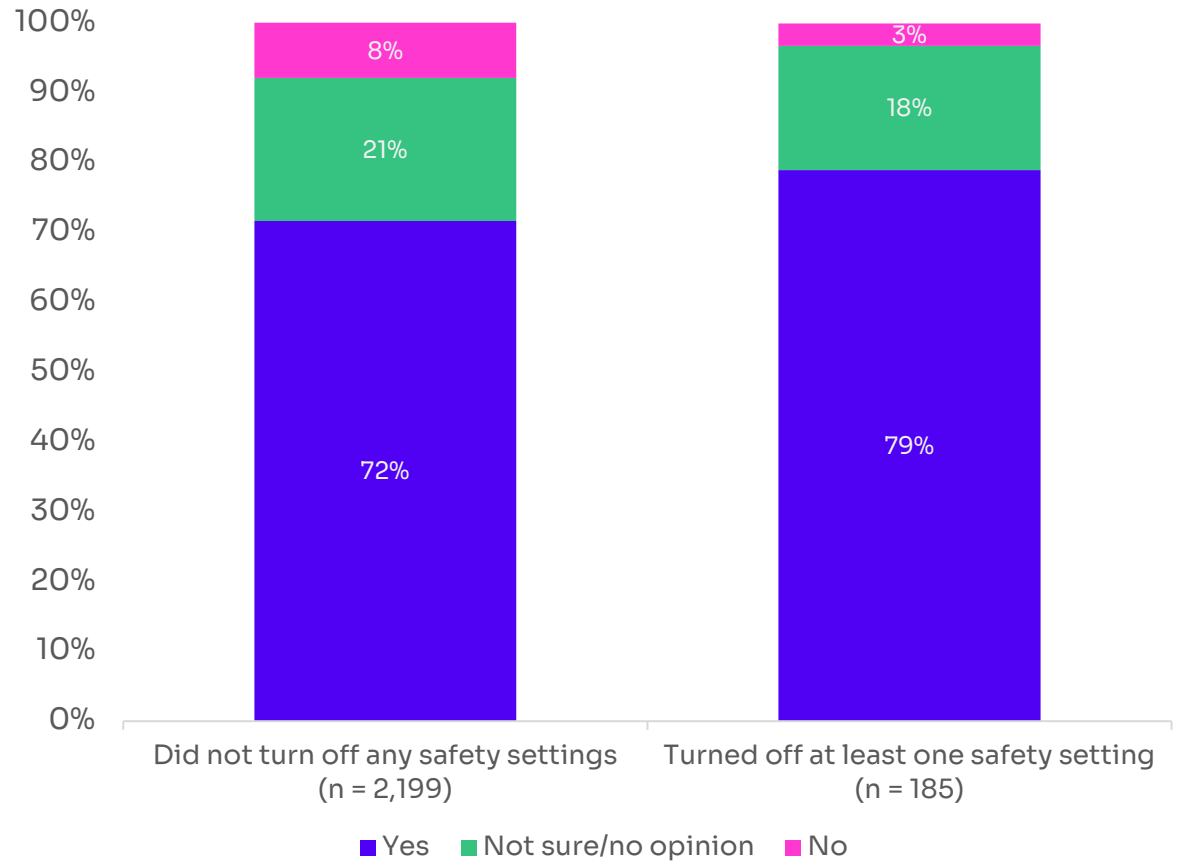
Figure 5: Proportion of all participants who turned off at least one safety setting by the end of the experiment, across all three settings categories combined



4. Satisfaction with, and comprehension of the options was higher among those who engaged with them

Participants who had viewed the settings during the experiment were asked to indicate whether they felt satisfied with the available options. Among participants who turned off at least one setting, satisfaction with the available options was high for all age groups (79%+).

Figure 6: Proportion of adults opting out of at least one default that indicated feeling satisfied with the available options



Our survey also tested participants' understanding of each category of settings.

Across several items, participants were shown an image depicting someone's settings screen with specific settings turned on or off, then asked questions such as "Do you think that other people who they are not connected with (not friends or followers) could send them direct messages?" (Yes/No/Don't Know).

In each of the three age groups, around two thirds of participants correctly identified the implications of settings. Across all age groups: comprehension was higher among those who had changed at least one setting compared with those who had not.

Figure 7: Proportion of adults and children who correctly understood settings and either did, or did not change a setting

Group	Participants correctly understanding settings	
	Changed a setting	*Did not change a setting*
Adults	76%	64%
16 -17-year-olds	55%	45%
13-15-year-olds	61%	47%

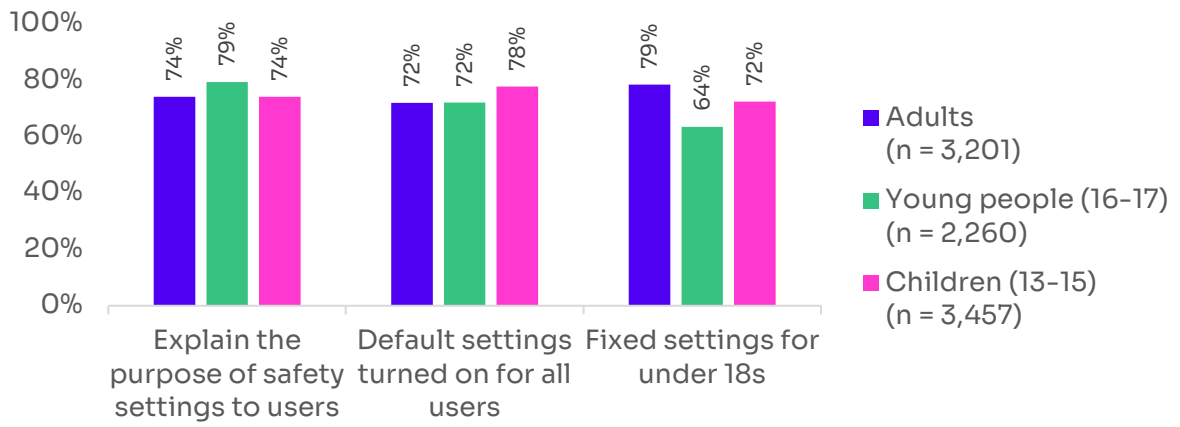
Comprehension also varied by setting type: for adults it was highest for messaging settings (69%) and lowest for network settings (60%). This was similarly true for each of the under-18s groups.

5. The majority of participants support the introduction of safety defaults, and feel they offer protective value

In our post-trial survey, we asked participants to consider whether, in the future, social media platforms should:

i) provide clearer explanations of the types of safety settings covered in the trial,
ii) turn these settings on by default for users (but allow users to then turn them off),
iii) fix these settings in place for under 18s (i.e. prevent children turning these settings off).

Figure 8, below, shows that the majority of participants across all age groups supported the introduction of all three of these measures.



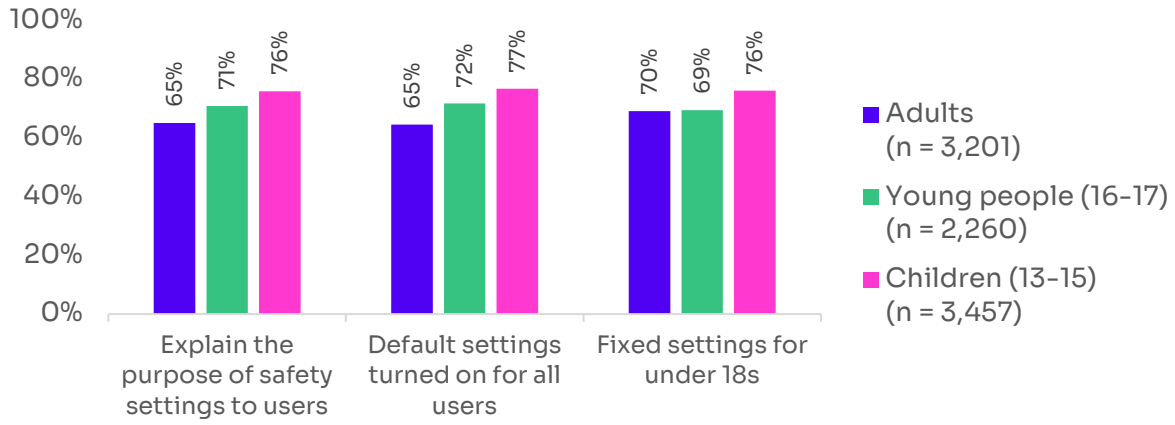
Of particular note, support was high across all age groups for fixed settings for under 18s (i.e. settings that are switched on and cannot be changed), suggesting that participants appreciate the safety benefits of these measures. Support for fixed settings was consistently strong, although slightly lower than support for default settings. The largest difference was among 16–17-year-olds, where 72% supported default settings compared with 64% who supported fixed settings.

We also asked each group the extent to which they agreed that these propositions would keep users safe. Figure 9, below, shows that the majority of participants thought that these propositions would keep users safe.

Support for all settings was broadly consistent across age groups. A notable difference was in relation to fixed settings for under 18s where children aged 16-17s were less likely to support this option that adults (18+) or younger children (13-15s).

This age group (16-17) were more likely to indicate support for defaulted settings for all age groups than fixed settings for under 18s.

Figure 9: Proportion of participants who agreed or strongly agreed with the value of safety settings propositions



Conclusions

While valuable for generating insights, mock online trials cannot fully replicate real-world behaviour. Studies of this kind, in which participants interact with simulated versions of platforms they use in everyday life, can provide useful evidence on attitudes, broad behavioural patterns, and the relative impact of different interventions. However, they are inherently limited in their ability to capture the wider social, contextual and emotional factors that often influence behaviour in real-world settings, such as peer influence, social norms, rewards, and personal stakes. Findings should therefore be interpreted with these limitations in mind and viewed as indicative of how users may respond, rather than as a direct measure of real-world behaviour.

1. Retention of defaults was high

Overall, the findings indicate that both adults and children largely retain safety-by-default options, at least when first exposed to these. The findings from the experiment align with behavioural literature and previous trial results showing that defaults act as a powerful "path of least resistance", reducing cognitive effort and simplifying decision-making in complex environments including online safety. The consistency of behaviour in retaining the default settings across setting types reinforces the generality of this effect. It could suggest that high retention reflects a broad behavioural tendency rather than preferences tied to specific settings or risks. Importantly, rates of viewing these settings were also high: around three-quarters of participants clicked through when prompted, yet this did not translate into widespread changing of settings.

2. Retention of defaults was similar between children and adults under the same conditions

The results also indicate a strikingly consistent pattern of behaviour across adults and children. Rates of viewing settings and retaining defaults were nearly identical between age groups, with only a small minority (around 5–7%) making any changes. This suggests that default effects operate similarly across age groups in this context.

This is particularly notable given our expectation that adolescents may behave differently due to heightened sensitivity to rewards or peer influence. Instead, the findings may suggest that the behavioural mechanisms underpinning default adherence—such as status quo bias and decision simplification—may be sufficiently robust to produce similar outcomes across age groups when choices are presented in comparable ways. However, as noted above these sorts of trials are generally unable to replicate these potential real-world drivers of behaviour (i.e. limited reward and no peer influence) and so results should be viewed as indicative of potential behaviour.

3. Participants were largely in support of defaulting users into protective settings, but support is nuanced

Our post-trial survey indicates that a majority support a platform-led, safety-by-default approach to these kinds of settings. Adults, young people, and children support platforms explaining the purpose of these kinds of settings, safe-by-default configurations, and tighter restriction for under 18s.

However, there are differences to note between younger and older children. While many 16–17 year olds are in favour of safe defaults (72%), their support drops significantly to 63% when considering *fixing* these settings for under 18-year-olds. This suggests that this age group, particularly, value having autonomy over these choices. This contrasts with the high levels of agreement among 13–15-year-olds regarding both the implementation (72%) and perceived safety value (76%) of fixed settings.

While it's perhaps not surprising that older children nearing adulthood particularly crave stronger autonomy, this group is at still at risk of online harm. For instance: older children are disproportionately affected by sextortion.[11] The higher vulnerability means even greater and more careful consideration may need to be given to the level of friction people in this age group are met with if they wish to opt-out of safety defaults.

In summary

The work reported here shows further supporting evidence that defaults – pre-selected options – can be substantial drivers of online behaviour. Using a mock social media platform setup, the current trial built on previous Ofcom behavioural trials by showing that those defaulted into protective settings for location sharing, network expansion, and direct messaging largely retained those settings despite being able to opt-out. This observation held across three age groups: 13-15-year-olds, 16-17-year-olds, and adults. We caveat that our mock platform was not deterministic and therefore may not have fully recreated factors that affect real-world decision making.

[11] Internet Watch Foundation (2025). Sexually Coerced Extortion.