Can't Stop: How Context and Individual Traits Influence Effectiveness of Different Gradual Interventions for Infinite Scrolling on Short-Form Video Platforms

LUCA-MAXIM MEINHARDT, Ulm University, Germany and Santa Clara University, United States
MANUELA DRAGIC, Ulm University, Germany
MARK COLLEY, UCL Interaction Centre, United Kingdom
KAI LUKOFF, Santa Clara University, United States
ENRICO RUKZIO, Ulm University, Germany

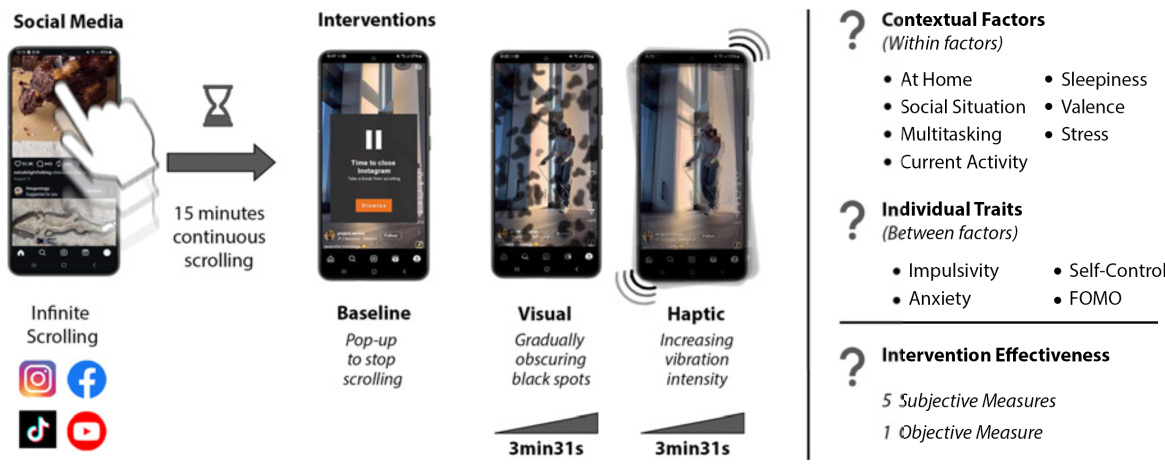


Fig. 1. Overview of the 7-day field study. Participants' continuous infinite scrolling on four major short-form video platforms (left) triggered one of three interventions (center). The baseline intervention used a pop-up, the two gradual interventions applied increasing levels of design friction via either the visual or haptic modality for up to 3 min 31 s. We investigated how the effectiveness of these interventions (subjective and objective) interacted with seven contextual factors and four individual traits (right).

Infinite scrolling on short-form video platforms like TikTok encourages prolonged engagement and post-usage regret. Interventions aim to mitigate such behavior, but their effectiveness may depend on the interplay between intervention type, contextual factors, and individual traits. In a 7-day within-subject randomized field study (N=104), we compared a baseline pop-up and two gradually intensifying design frictions (visual and haptic). We evaluated behavioral changes and user experience using objective and subjective measures. Results showed that the pop-up was initially effective but quickly

Authors' Contact Information: Luca-Maxim Meinhardt, luca.meinhardt@uni-ulm.de, Ulm University, Ulm, Germany and Santa Clara University, Santa Clara, United States; Manuela Dragic, manuela.dragic@uni-ulm.de, Ulm University, Ulm, Germany; Mark Colley, m.colley@ucl.ac.uk, UCL Interaction Centre, London, United Kingdom; Kai Lukoff, klukoff@scu.edu, Santa Clara University, Santa Clara, United States; Enrico Rukzio, enrico.rukzio@uni-ulm.de, Ulm University, Ulm, Germany.

ACM 2474-9567/2026/9-ART140
https://doi.org/10.1145/3831979

lost impact, whereas the visual gradual intervention sustained subjective ratings the longest. Bayesian modeling revealed that self-regulation traits moderate how participants responded to the three intervention types. For participants with low impulsivity, the type of intervention had little influence on its subjective effectiveness. For participants with high impulsivity, however, differences between intervention types were substantial, with the explicit baseline pop-up being most effective compared to the novel gradual interventions. Contextual factors, in contrast, showed little influence. These findings suggest that intervention modality and individual differences in self-regulation shape intervention effectiveness.

CCS Concepts: • **General and reference** → *Surveys and overviews*; • **Human-centered computing** → *HCI design and evaluation methods*; **Scenario-based design**; **Empirical studies in HCI**.

Additional Key Words and Phrases: infinite scrolling, context, individual traits, field study, digital well-being, Bayesian statistics

ACM Reference Format:
Luca-Maxim Meinhardt, Manuela Dragic, Mark Colley, Kai Lukoff, and Enrico Rukzio. 2026. Can't Stop: How Context and Individual Traits Influence Effectiveness of Different Gradual Interventions for Infinite Scrolling on Short-Form Video Platforms. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 10, 3, Article 140 (September 2026), 36 pages. https://doi.org/10.1145/3831979

1 Introduction

Social media (SoMe) platforms like TikTok and Instagram have reshaped user engagement with digital content through mechanisms like infinite scrolling. Infinite scrolling automatically loads new content as users scroll, substantially extending screen time [74]. This interaction design encourages unintentional and habitual engagement [77, 96], often leaving users with a sense of post-usage regret [17] and with the feeling of temporarily disconnecting from their surroundings [4]. Consequently, infinite scrolling has been identified as an attention-capturing dark pattern [99], designed to manipulate user behavior in ways that may conflict with their interests [36]. Infinite scrolling on short-form video platforms such as TikTok raises particular concerns, as these platforms are associated with significantly longer screen time and more negative emotions than text-based platforms [96]. Further, there is evidence that short-form video platforms decrease our ability to retain initial intentions [15] and to set boundaries [121]. Therefore, the European Commission scrutinized TikTok's design for potential violations of the Digital Services Act, noting that it "*[...] may stimulate behavioral addictions and/or create so-called 'rabbit hole effects'*" [32].

Even though users are generally aware of these effects and express a desire to limit their SoMe use, they frequently struggle to follow through [53]. To counteract this, interventions have been designed to help people manage their SoMe use. Most approaches directly restrict access to specific apps by enforcing time limits [40] or lockout tasks [50]. While these strategies effectively reduce screen time, they often come at the cost of users' agency [66]. Once a lockout is triggered, users are confronted with a hard boundary that they did not actively choose in that moment. According to Reactance Theory [9, 94], such externally imposed restrictions threaten users' sense of freedom, which can trigger negative emotions. Indeed, prior studies show that users frequently dismiss timers, disable reminders, or abandon lockout apps because they feel overly controlled [66, 81]. In the long term, this reactance not only undermines the intervention's effectiveness but also reduces its acceptability, as users perceive it as intrusive or paternalistic. In contrast, Lu et al. [64] showed that making common gestures like taps or swipes slightly more difficult to perform can more subtly nudge users to reduce smartphone use. This principle of design friction [20] introduces micro-boundaries "*[...] that provide a small obstacle prior to an interaction that prevents us rushing from one context to another*" [20, p. 1392]. These interruptions, though minor, create moments for reflection, prompting users to reconsider their behavior [6, 20, 73]. Hence, design friction has been successfully applied to mitigate SoMe overuse [39] and, more specifically, scrolling behavior [101]. One challenge with design friction is finding the right level of friction [49, 68]. Too little friction is easily ignored; too much can provoke reactance and abandonment of the intervention [66]. Gradually increasing intensity offers one potential, novel design strategy to find the best level of friction over time, as it allows interventions to become harder to ignore the longer scrolling continues. This motivates our first research question (RQ1):

RQ1: How do users experience interventions over time as friction intensity increases?

Whether such strategies succeed depends not only on how friction is designed, but also on where and when it is encountered [92], as their effectiveness is likely to vary across contexts. For instance, during work hours, subtle interventions may suffice as social norms already discourage phone use [34, 75], whereas during leisure time, stronger interventions may be necessary [96]. Meinhardt et al. [72] also argued to systematically explore "*various interventions to determine the most effective ones for specific contexts*" [72, p. 12]. Beyond context, prior work found that individual traits such as users' ability for self-regulation predict smartphone addiction [38] and problematic SoMe usage [63]. In particular, low self-control, high fear-of-missing-out (FOMO), and high impulsivity can lead to problematic SoMe use [37, 55, 106]. While the relationship between these traits and problematic use is well-known, the moderation between them and the effectiveness of specific interventions remains largely unexplored. Mark et al. [71] provide early evidence that this relationship matters in the workplace, where distraction interventions helped people with low self-control focus but caused stress for those with high self-control. Whether individual traits and contextual factors similarly moderate the effectiveness of different types of interventions for infinite scrolling during short-form video platforms is the central question of this work. This motivates our RQ2:

RQ2: How do *contextual factors* and *individual traits* moderate users' behavior change and experience of different interventions during infinite scrolling?

To answer both RQs, we operationalized intervention effectiveness along two dimensions: objective and subjective. Objective effectiveness was measured using *responsiveness* [72], defined as the time to stop infinite scrolling after an intervention occurred. However, prior research shows that reducing screen time alone can provoke negative reactions and relapse into old habits [81]. Therefore, we also considered subjective effectiveness as the weighted combination of *reactance* [31], *goal alignment* [67], *usefulness* [120], *agency* [67], and *satisfaction* [67]. In our 7-day field user study ($N = 104$), participants installed an Android application that monitored infinite scrolling behavior on TikTok, Instagram, Facebook, and YouTube Shorts. Once a scrolling session exceeded 15 min (in line with [72, 96, 114]), the app randomly triggered one of three interventions: The baseline intervention showed a pop-up that encouraged users to take a break from scrolling. The other two interventions gradually intensified either visual or haptic friction over 3 min and 31 s [72] (see Figure 1). In the visual condition, the content became progressively more challenging to view, while in the haptic condition, the phone increasingly vibrated. After users stopped scrolling due to the intervention, they rated their subjective intervention's effectiveness and reported their current context (e.g., location, valence, stress). Before the study, we collected the participants' traits for *impulsivity*, *FOMO*, *anxiety*, and *self-control*.

Our results show that the baseline pop-up intervention yielded the highest objective effectiveness, but its subjective experience diminished rapidly if users did not act immediately, suggesting that simple reminders work primarily for those already motivated to disengage. In contrast, the visual gradual intervention maintained stable subjective ratings throughout its increase in intensity, potentially supporting longer-term effectiveness by avoiding abrupt drops in the intervention's salience. We found that individual traits related to self-regulation, particularly self-control and impulsivity, moderate the relationship between our intervention types and both subjective and objective effectiveness. By contrast, contextual factors showed only limited interaction effects. For participants with high self-control or low impulsivity, the intervention type mattered less, likely because they could disengage from infinite scrolling without much external support. In contrast, participants with low self-control or high impulsivity responded most effectively to the baseline intervention. Its explicit prompt to *take a break* may provide the clear external signal these users need, whereas the gradual interventions might be too subtle for those with high impulsivity. While we tested only three different intervention types, our results suggest that future work should move beyond one-size-fits-all interventions. We suggest that interventions should be carefully designed to tailor them to individual differences. Further, intervention designers should be careful about what they optimize: objective behavior change vs. subjective experience. Our results reveal a trade-off between these two measures. Interventions that cut scrolling fastest can feel frustrating, whereas more acceptable designs may delay disengagement. Effective interventions should balance these goals, or adapt which goal they prioritize.

Contribution Statement [126]

Empirical study that tells us about people | Artifact. We contribute a 7-day field study (N = 104) comparing a baseline pop-up with two novel gradually intensifying interventions (visual vs. haptic) to mitigate infinite scrolling. We demonstrate that, within the scope of our study, intervention effectiveness depends not only on our tested designs but also on individual differences in self-regulation ability, such as impulsivity and self-control.

2 Related Work

This section reviews interventions aimed at reducing SoMe overuse and promoting digital well-being by limiting smartphone use. We also examine prior research on how contextual factors and individual traits influence problematic phone usage and behavior change. Chen et al. [14] classified moments of killing time and found that especially sessions of prolonged infinite scrolling were likely labeled as time-killing. In response, researchers have proposed interventions to reduce SoMe and smartphone use, which we briefly summarize below.

2.1 Interventions for Limiting Social Media Use

Digital interventions to limit SoMe use can be categorized into internal and external [67, 90]. Internal interventions are embedded in the app, such as removing the newsfeed to reduce exposure to endless content [90]. External interventions operate at the device level, modifying smartphone behavior without altering app functionalities. These external interventions vary in intrusiveness and can be grouped into four main types [97]: First, usage timers help users reflect on and potentially adjust their habits [40]. Second, persuasive interventions use prompts or notifications to encourage mindful engagement [89, 93]. Third, break reminders aim to interrupt usage and direct users toward more meaningful activities, such as breathing exercises [39, 114]. Lastly, phone blockers restrict access entirely or increase usage difficulty, providing support for individuals struggling with self-regulation [50]. As an alternative to such interventions, timeboxing structures phone use into bounded sessions with self-set time allowances rather than restricting access outright [86].

While internal and external interventions reduced SoMe use, internal interventions are more effective on passive SoMe features such as infinite scrolling [84]. Lee et al. [61], who tested an intervention that disabled attention-capturing dark patterns such as infinite scrolling, showed 21% less distraction when these dark patterns were disabled. Most interventions employ a rather intrusive approach, delivering them as a sudden interruption of the user's interaction. In contrast, the design friction approach advocates for small obstacles prior to the interaction to create moments of reflection to change behavior [6, 20, 73]. Recent work proposed applying design friction specifically to scrolling behavior: Ruiz et al. [101] required users to interact with each post before being allowed to scroll further. Participants reported that this increased their attention to individual posts. However, they also expressed frustration due to the repetitive interaction required. Despite their variety, many intervention mechanisms rely on uniform application across users, overlooking contextual and individual factors that could influence the interventions' effectiveness.

2.2 The Role of Context in Digital Media Use

Contextual factors, such as environmental setting, mobility, social interactions, multitasking, and distractions, can influence digital media use [1]. For instance, Böhmer et al. [12] found that users consume news content in the morning and shift to SoMe content in the evening. Further, Xu et al. [128] observed higher mobility among users of social networking and gaming apps. A large-scale behavioral study confirms these trends [41]. They further showed that smartphone use is nearly twice as high at home compared to the office [41]. Rixen et al. [96] also observed an 18% drop in infinite scrolling during work-related activities compared to leisure. They also found that infinite scrolling is often used as a coping mechanism for negative emotions or procrastination. Beyond temporal and locational influences, social context plays a crucial role. Do et al. [28] found that the presence of nearby Bluetooth devices was associated with increased smartphone use. Conversely, Weber et al. [122] showed that people were less likely to use their phones during group dining. These findings highlight the dynamic nature of digital media usage and introduce the importance of context for changing the behavior of digital media consumption [46, 115]. Acknowledging this, Ding et al. [26] emphasized the crucial role of time and location when it comes to setting reminders to change behavior. They argue that "*[...] context information plays a very important role in increasing the effectiveness and reducing the annoyingness of reminders [interventions]*" [26, p. 7].

2.3 Context-Aware and Tailored Interventions

Recent work argues for moving beyond one-size-fits-all strategies by tailoring SoMe interventions to users' situational context [83, 91, 96, 98, 100, 107, 119]. For example, Orzikulova et al. [85] designed adaptive just-in-time interventions based on phone usage, physical activity, time of day, location, and social setting. By using machine learning to predict optimal intervention moments, their system improved intervention accuracy by 33% compared to non-adaptive interventions. Similarly, Wu et al. [127] showed that mental states (boredom, stress, or inertia) shape how users respond to interventions. They generated personalized pop-ups using large language models and found that integrating mental state increased acceptance by up to 22.5%. Meinhardt et al. [72] investigated how contextual factors influence user responses during infinite scrolling. They found that low valence, combined with being at home, led to users disengaging from scrolling more quickly. In contrast, when sleepy, they were more likely to accept interventions but were less likely to act on them. However, their study focused on a single intervention type and did not examine whether individual differences also moderate intervention effectiveness. This limitation is important because individual traits are known predictors of problematic SoMe use [37, 55, 106]. Moreover, interventions themselves are not uniformly effective. Distraction blockers, for example, helped people with low self-control to focus but caused stress for those with high self-control [71]. If such traits shape both usage patterns and reactions to restrictions, they will also likely determine how individuals perceive and respond to interventions during infinite scrolling.

Our work addresses this gap by systematically comparing how three intervention types differ in objective behavior change and subjective experience, and by testing whether individual traits and contextual factors moderate these differences. Unlike prior work that established relationships between individual self-regulation traits and problematic phone use [38], our study directly examines whether these traits also determine which intervention type is most effective for a given user.

3 Contextual Factors and Individual Traits for the User Study

To account for the diverse range of influences on intervention effectiveness during infinite scrolling, we identified eleven factors grounded in prior research. These include both between-subject individual differences and within-subject contextual factors.

3.1 Contextual Factors *(Within Factors)*

We included seven contextual factors (within-factors) in our study: *current activity*, *social situation*, *location (at home)*, *multitasking*, *valence*, and *sleepiness*. These factors influence how users engage with their devices and respond to interventions during infinite scrolling [72, 91, 96]. For example, interventions may be more effective when users are engaged in activities like work rather than during leisure [96], or when social norms discourage phone use, such as in social gatherings [75, 76]. Similarly, usage *at home* tends to be longer [41]; therefore, during this context, the intervention effectiveness may be reduced [72]. Further, *multitasking* (e.g., eating or watching TV while being on the phone) creates cognitive distractions to redirect attention away from the phone, especially when being *at home* or during low *valence* [72]. Regarding the latter, Rixen et al. [96] found that extensive scrolling sessions were linked to lower *valence*, suggesting a negative emotional impact. Since users often cope with negative emotions via smartphones [24], emotional state may also influence the intervention's effectiveness.

We considered including time of day as a contextual factor but ultimately excluded it, in line with Meinhardt et al. [72]. Due to its periodic nature and strong individual variability in daily routines (e.g., between shift workers and students), time of day is a poor proxy for user state in our modeling. Instead, we included *sleepiness* as a more individual temporal indicator. Sleepiness varies throughout the day, can be modeled linearly, and meaningfully predicts user behavior Meinhardt et al. [72]. In addition to the above-mentioned six contextual factors, we included *stress*. As Vanden Abeele [119] mention *stress* as a key determinant of digital well-being, it is reasonable to assume that *stress* may also shape how users perceive and respond to interventions.

3.2 Individual Traits *(Between Factors)*

Digital well-being is also shaped by individual differences, particularly traits such as *impulsivity*, *self-control*, *anxiety*, and *FOMO* [119], all of which have established links to problematic SoMe use. *Impulsivity* is the tendency to act in favor of immediate rewards without deliberation [79] and is strongly linked to SoMe addiction [37]. *Self-control* describes the capacity to override such urges in favor of long-term goals [113], with low self-control relating to more problematic use [106]. Both constructs are typically strongly negatively correlated [70], yet they are not redundant. Impulsivity "*is not merely a lack of self-control, but rather a manifestation of relatively high levels of appetitive motivation for which self-control may serve as a buffer [...]*" [70, p. 74]. *Anxiety* reflects a disposition toward anticipatory apprehension, or worry, about perceived future threats [22] and relates positively to problematic SoMe usage [105]. Lastly, *FOMO* describes the apprehension of missing rewarding experiences others might be having [88] and it increases the risk of problematic phone use [55].

While the relationship between these four traits and problematic phone and SoMe usage is well established, less is known about whether they also moderate the effectiveness of interventions designed to mitigate such behaviors. Initial evidence from Mark et al. [71] suggests that individual differences in self-control shape how users respond to an intervention, but this question has not been examined across different intervention types for infinite scrolling on SoMe. Based on their influence on SoMe usage, we included *impulsivity*, *anxiety*, *self-control*, and *FOMO* as between-subject factors to examine their moderating role in intervention effectiveness during infinite scrolling.

4 User Study

To examine how contextual factors and individual differences moderate objective and subjective effectiveness of three intervention types during infinite scrolling, we conducted a 7-day within-subject field study with $N = 104$ participants. This duration was chosen to capture a sufficient range of intervention encounters across varied situational contexts rather than to demonstrate lasting behavior change, which would require longer observation periods [39]. A week also spans a full cycle of participants' weekday and weekend routines, so that recurring daily situations are represented at least once in the data. We focused on four of the most widely used short-form video platforms in the United States in 2023 (Instagram, TikTok, Facebook, and YouTube Shorts) [112], aligned with prior work [72, 96].

4.1 Apparatus

Our study was built on prior work, which open-sourced their application to track infinite scrolling [72]. We extended this work by adding two novel, gradually intensifying interventions (see subsubsection 4.1.1). The tracking mechanism remained the same: using Android's Accessibility Service [35], the app accessed the content variable of each app's hierarchy to identify infinite scrolling. Active engagement with SoMe, such as messaging or content creation, was excluded. For example, if a user switched from Instagram's "Reels" tab to the messaging interface, the content variable shifted, which was interpreted as a scrolling interruption. This focus on passive scrolling aligns with findings that interventions are more effective when targeted at specific in-app features rather than overall SoMe time [84].

A scrolling session was defined as uninterrupted feed consumption until the user either switched to a non-scrolling activity or closed the app. After 15 min of continuous scrolling, the app randomly triggered one of the three interventions with equal probability (1/3 each). Randomization occurred independently at each intervention moment, so participants could receive the same or different interventions across sessions (see subsubsection 4.1.1). We used the 15 min threshold as done in previous works [72] based on the findings that negative emotions from smartphone use often emerge after 10-20 min [114] and that sessions exceeding 10 min are typically dominated by infinite scrolling [96]. When users stopped scrolling after an intervention, a short questionnaire was presented to collect contextual information (subsubsection 4.3.2) and assess users' subjective intervention effectiveness (subsubsection 4.3.1). This event-based Experience Sampling Method (ESM) was chosen over fixed-interval ESM because it captures context immediately after disengagement from infinite scrolling, improving ecological validity and response rates [117]. However, this method only captures the participant's context at the moment they decide to stop scrolling. Hence, we missed data from those who continued scrolling. Despite this, we chose to collect subjective measures after infinite scrolling ended rather than during scrolling because interrupting the activity would have disrupted the scrolling behavior we aimed to observe. This approach is consistent with other ESM-based SoMe studies (e.g., [5, 13, 16, 96]).

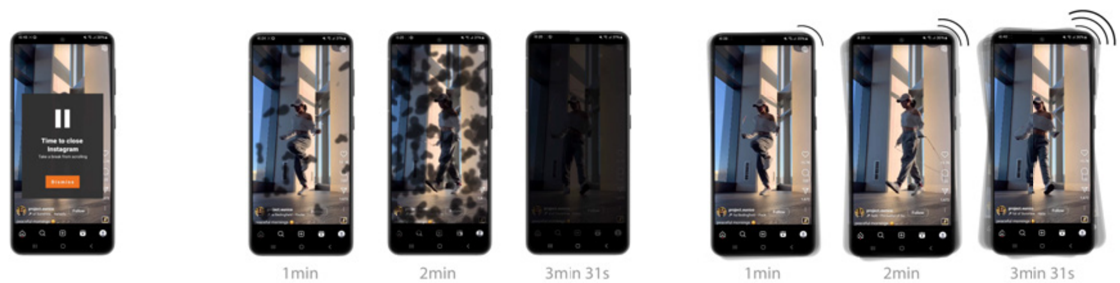
We also recorded the time that passed between the intervention occurring and the user stopping their scrolling behavior. We used his duration (*responsiveness*) to calculate the users' objective effectiveness of the intervention (see subsubsection 4.3.1).

4.1.1 Interventions. The baseline intervention replicated prior work [72] and existing screen-time reminders used on TikTok [116] and Instagram [42]. The intervention consisted of a single pop-up overlay (310×400 dp), presented once per scrolling session, that informed users it was time to close the SoMe app and take a break from scrolling. Users could dismiss this message by tapping the "Dismiss" button, allowing them to continue scrolling (see Figure 2a). We compared this baseline intervention with two novel interventions that gradually intensify based on haptic and visual modality. Drawing on the principle of gradual stimuli for behavior changes [52], these interventions subtly disrupt the user experience of infinite scrolling over time. By gradually increasing the cost of infinite scrolling through visual and haptic interference, we aimed to encourage disengagement without abrupt disruption. This approach aligns with design friction [20, 101], which advocates for thoughtful interaction interruptions to support digital self-regulation. Similar to the baseline, the gradual interventions were triggered after 15 min of uninterrupted infinite scrolling and reached maximum intensity after 3 min 31 s. This duration was grounded in related work [72], where users took an average of 3 min 31 s to stop infinite scrolling after an intervention occurred.

For the **visual intervention**, 12 different semi-transparent black spot images (37×41 px to 84×83 px) were placed at random positions across the phone's screen, inspired by the visual impairment of diabetic retinopathy [78]. A total of 3,000 spots were added over the time window, with 90% of all spots appearing in the final 30 s. Each spot ran an independent 15 s fade-in animation, scaling from 10% to 100% of its size while increasing from fully transparent to near-opaque. For the **haptic intervention**, we build on the approach from

Okeke et al. [82]. A recurring 500 ms vibration pulse started at amplitude 30 (of 255) [3], incrementing by 3 per pulse until reaching maximum. The initial pause between pulses was 5 s, decreasing by 70 ms after each pulse until reaching 0 ms. The 5 s initial pause follows Pielot et al. [87], who showed that this pause between vibrations prevents users from quickly adapting to repetitive haptic feedback.

Our design choices for the gradual interventions aimed to compare how contextual factors and individual traits moderate the effectiveness of different friction modalities suitable for in-field deployment. Semi-transparent visual occlusion was chosen to introduce perceptual friction while keeping content visible. Haptic friction was chosen as a non-visual modality that conveys friction with minimal reliance on sound, which may be perceived as intrusive in public or quiet environments and, therefore, less suitable for passive smartphone use [87].



(a) Baseline Intervention (b) Visual Intervention (c) Haptic Intervention

Fig. 2. Visualization of the three interventions that were used in the user study. The Visual and Haptic Intervention increases friction intensity over time, maximizing at 3 min 31 s.

4.2 Procedure

Recruitment and screening were conducted on Prolific with participants from the United States and the United Kingdom. Eligibility was restricted to individuals who owned a device with Android 10 or later to leverage platform-specific permission handling introduced with Android 10. At the time of the study, this version was installed on over 87% of Android devices in both countries [111]. After accepting the Prolific study, participants were redirected to an LimeSurvey instance where they completed the registration survey, including the trait questionnaires assessing *anxiety*, *self-control*, *FOMO*, and *impulsivity*. To ensure attentiveness, we included the attention check: "I breathe more than once a day." Upon completion, they received a tutorial video and a QR code (for desktop users) or a direct download link (for mobile users) to install the study application on their primary Android smartphone. After installation, the app displayed a consent form, which participants were asked to review and accept. Participants who installed the application and completed the full 7-day duration (regardless of whether they provided data) were compensated £1.22/$1.55. No dedicated fraud screening was implemented, but automated participation was unlikely as in-app questionnaires only appeared after 15 minutes of natural scrolling on participants' phones, making their timing unpredictable in daily use.

The study protocol, including the consent procedure, received formal approval from the university's Ethics Committee. Particularly, those ensuring participant anonymity were strictly adhered to. Following consent, participants were guided through enabling the necessary permissions. The app then prompted users to enter their age and select their gender. Once these demographic items were completed, the app activated its main tracking function. At the end of the 7-day period, the app notified participants that the study had concluded and that they could uninstall the application. We then compensated the participants according to the number of questionnaires filled out (£0.50/$0.60 per questionnaire).

4.3 Measurements

To minimize participant burden and reduce the risk of survey fatigue, we primarily relied on short, single-item measures. Although single-item assessments may introduce some measurement error [23], their use is widely accepted and validated in SoMe research, offering a practical trade-off between accuracy and user engagement [5, 7, 13, 72, 96]. To ensure data quality, we included random attention checks into the questionnaires. A full list of the items used in the study is provided in Appendix B.

4.3.1 Intervention Effectiveness. We assessed the effectiveness of the three interventions using both objective and subjective measures to capture behavioral change and user experience.

Objective effectiveness was measured using the time it took participants to disengage from infinite scrolling after the intervention occurred. This *responsiveness* was measured in seconds. To address the right-skewness in the *responsiveness*, we applied a logarithmic transformation [29]. Subsequently, to calculate the objective effectiveness score, we normalized the data to a 0–1 scale, using min–max normalization based on the observed range, and reversed the outcome so that higher scores indicate faster disengagement.

For the subjective effectiveness, participants rated the interventions based on five scales after they stopped infinite scrolling. These measures were derived from (1) Lukoff et al. [67], (2) Venkatesh and Davis [120], and (3) Meinhardt et al. [72]. From Lukoff et al. [67], we used the items to measure *goal alignment*, *satisfaction*, and *agency*; all three items were rated on a 7-point Likert scale. The wording of the items was slightly modified to ensure they referred directly to the intervention experience. For example, to assess perceived agency, participants were asked: "*For this intervention, how much did you feel out of or in control?*". In addition to these three measures, we included the perceived *usefulness* item drawn from the extended Technology Acceptance Model [120]. Perceived usefulness is a determinant for the intention to use [120]; therefore, we believe it is an ideal sub-indicator for the subjective effectiveness of an intervention. Usefulness was measured by the single-item "*I find the intervention to be useful in my current situation*" rated on a 7-point Likert scale. Lastly, to capture potential negative reactions, we measured *Reactance* using the threat subscale of the Reactance Scale for HCI (RSHCI) [31], following the approach of Meinhardt et al. [72]. Reactance refers to the resistance individuals feel when their freedom is restricted [31]. This resistance is rooted in psychological models [25], implying that "*messages [interventions] designed with the objective of behavior change must necessarily (implicitly or explicitly) limit an audience's freedom*" [94, p. 67]. Thus, it is particularly relevant here, as interventions may limit users' ability to continue scrolling. For the threat subscale in RSHCI, five items were used, such as "*I don't want the intervention to tell me what to do*", which were rated on a 5-point Likert scale (1= "strongly disagree", 5="strongly agree"). We reversed the rates for reactance and normalized each score to a 0-1 range, respectively, to each questionnaire scale's boundary. Subsequently, we combined the five subjective measures into a single weighted score for the subjective intervention effectiveness. Hence, we conducted an exploratory factor analysis [125] using the minimum residual method, which supported a unidimensional structure. All items loaded substantially on the latent factor (threshold > 0.4 [95]): Reactance (0.49), Goal Alignment (0.89), Satisfaction (0.88), Agency (0.51), and Usefulness (0.85). Model fit indices acceptable fit (RMSR = 0.07), with strong reliability statistics ($\omega = 0.90$). Following established practice [27], we used the normalized factor loadings as weights to calculate the combined subjective effectiveness score.

4.3.2 Contextual Factors. To determine the effects of context on intervention effectiveness, we included seven contextual factors. For *current activity*, we applied the interval scale developed by Samdahl [102], ranging from -3 ("definitely leisure") to +3 ("definitely not leisure"). The *social situation* was measured using a question adopted from Akpinar et al. [1], asking participants: "Which one of these best describes people around you?" with two response options: alone and with acquaintances (e.g., friends, family, colleagues). To determine whether participants were *at home*, we asked a simple yes/no question: "Are you currently at home?" adopted from Meinhardt et al. [72]. We determined *multitasking* by asking, "Did you do anything else besides being on [App Name]?" which participants could also answer with yes or no [72]. For evaluating participants' *valence*, we used the Self-Assessment Manikin (SAM) scale [8], as previously employed by Rixen et al. [96]. Sleepiness was measured using the Karolinska Sleepiness Scale (KSS) [104], which ranges from 1 ("extremely alert") to 9 ("extremely sleepy"). Lastly, stress was determined by the stress numerical rating scale [47] on an 11-point Likert-type scale, ranging from 0 ("no stress") to 10 ("worst stress possible"). Using the fixed boundaries of each questionnaire, we normalized all contextual factors to a range from 0 to 1.

4.3.3 Individual Traits. We measured *impulsivity* using the 15-item short version of the Barratt Impulsivity Scale [109]. This scale assesses the subcategories of non-planning, motor impulsivity, and attentional impulsivity on a 4-point Likert scale. For *self-control*, we used the 13-item Self-Control Scale developed by Tangney et al. [113]. It is rated on a 5-point Likert scale. For *anxiety*, we used the short version of the Spielberger Trait Anxiety Inventory, which consists of 5 items [129], using a 4-point Likert scale ranging from 1=almost never to 4=almost always. *FOMO* was assessed via the trait part of the Trait-State FOMO scale, which uses 5 items on a 5-point Likert scale [123]. Using the fixed boundaries of each questionnaire, we normalized all individual traits to a range from 0 to 1.

4.4 Participants and Power Simulation

Only participants who experienced all three intervention types were included in the final analysis to maintain the integrity of our within-subject design. To determine the required sample size for achieving 80% statistical power [19] for our subsequent analysis (see subsubsection 4.5.4), we followed the simulation-based approach proposed by Kumle et al. [60]. After collecting data from an initial batch of $n = 30$ participants ($M = 34.33$, $SD = 10.76$ years; 11 female, 18 male, one non-binary, 0 prefer not to say), (334 data points), we attempted to run the power simulation using the full linear mixed model (LMM) with all 11 factors (see section 3). However, this model proved computationally infeasible to simulate for power, even after one month of runtime.

To address this, we reduced complexity by identifying the most important factors. Hence, we fit the full LMM to the pilot data and inspected the t-values of all interaction terms involving *Intervention Type*. Following the threshold of $|t| \geq 2$ [60], we identified *Impulsivity ($t = -2.69$)*, *Anxiety ($t = 3.51$)*, and *FOMO ($t = 2.22$)* as the factors with sufficiently strong interaction effects. The corresponding t-values are reported in Table 6. Based on these results, we simplified the model for the power simulation. To reduce computational effort in simulating two models for subjective and objective intervention effectiveness, we combined both scales, leading to the following model:

$$\text{Inter. Effectiveness (combined)} \sim \text{Inter. Type} \times (\text{Impulsivity} + \text{Anxiety} + \text{FOMO}) + (1 \mid \text{Participant})$$

Using the `mixedpower` package [60], we ran simulations with the simplified model for increasing sample sizes ranging from 30 to 150 in steps of 30, with 1000 simulations per sample size and a critical value of 2 (approximating a two-tailed α level of .05 [60]). The results of the simulation (see Appendix C) showed that the interaction between *Intervention Type × Impulsivity* reached more than 80% power at $n = 30$. *Intervention Type × FOMO* reached 80% power by $n = 60$. *Intervention Type × Anxiety* reached 80% power by $n = 90$. Therefore, a sample size of at least $n = 90$ is sufficient to detect the interaction effects with acceptable power [19].

Hence, participant recruitment proceeded in batches of approximately 30 participants over a four-month period. 792 individuals completed the registration phase. Of these, 54 were excluded for failing attention checks in the individual traits questionnaires, and 524 were asked to return their submissions because they did not install the application. Participants who opted out often mentioned download difficulties or the burden of a 7-day study. In addition, we believe this dropout range can largely be attributed to the low initial effort required for registration. This left 214 participants who successfully completed the 7-day app-based study and received compensation for their participation. After filtering out participants who did not experience all three intervention types (see above), the final dataset included $N = 104$ participants ($M = 36.95$, $SD = 10.68$ years; 44 female, 58 male, 2 non-binary, 0 preferred not to say). Of the 110 excluded participants, 64 experienced two or fewer interventions in total, suggesting they either lost interest during the study period or did not exceed the 15-minute scrolling threshold frequently enough to trigger all conditions. The remaining 46 participants experienced three or more interventions overall, but due to the randomized assignment order, they were not exposed to all three types. To assess whether this filtering decision influenced the results, we conducted sensitivity analyses with two relaxed inclusion criteria ($N = 150$ and $N = 214$; see Appendix E).

Of the included participants, an additional 49 data points were excluded due to failed attention checks in the in-app surveys, and 23 incomplete entries were removed as they were likely due to technical errors. Subsequently, we applied the z-score method to detect outliers across all six intervention effectiveness measures. Data points exceeding ±3 SD were excluded, which resulted in removing 24 observations for *responsiveness*. After this step, 1,294 data points from 104 participants remained for the final analysis ($M = 12.40$, $SD = 9.36$, $Md = 10$ interventions per participant), distributed across platforms as follows: 41.27% TikTok, 22.91% Instagram, 17.91% Facebook, and 17.91% YouTube Shorts. Each intervention was distributed nearly equally across participants, with a median of three occurrences per participant for all three conditions (Baseline: $M = 4.26$, $SD = 3.86$; Visual: $M = 4.23$, $SD = 3.52$; Haptic: $M = 4.03$, $SD = 3.35$). Further descriptive statistics are provided in Appendix A.

4.5 Results

Initially, we conducted a survival analysis to examine disengagement behavior over time while accounting for censored observations. Second, we examined how participants' experience of the interventions changes as their friction intensifies and when the gradual intervention becomes too intense and thus less effective (RQ1). Finally, we present two Bayesian mixed-effects models, which test how contextual factors and individual traits moderate subjective and objective intervention effectiveness (RQ2).

4.5.1 Disengagement from Infinite Scrolling: Descriptive and Survival Analysis. Participants disengaged fastest with the baseline intervention ($Md = 7s$), followed by the haptic gradual intervention ($Md = 28s$), and slowest with the visual gradual intervention ($Md = 56s$). Further, the baseline exhibited lower variability ($IQR = 45s$) compared to the gradual interventions (visual: $IQR = 1min52s$; haptic: $IQR = 1min21s$). As our data captured disengagement as a time-to-event process, we analyzed responsiveness using survival analysis. Survival analysis is well-suited here because it accounts for censoring [45], that is, cases where participants continued scrolling beyond the observation window. We defined this window as the point of maximum intervention intensity (3 min 31 s). Responsiveness values below this cutoff were coded as events (scrolling stopped), values exceeding it were treated as censored. We then fitted a Cox proportional hazards model, a regression approach within survival analysis that estimates how predictors (here, intervention type) affect the hazard rate (HR), which is the instantaneous probability of stopping scrolling at a given time [21]. To account for repeated observations per participant, clustering was included at the participant level. The model indicated significant overall differences between interventions, $\chi^2(2) = 22.13$, $p < .001$. Pairwise comparisons using estimated marginal means (on the log-hazard scale) showed that the visual gradual intervention led to significantly slower disengagement than the baseline pop-up, $HR = 0.71$, 95% CI [0.59, 0.85], $p < .001$. The haptic gradual intervention also reduced disengagement relative to the pop-up ($HR = 0.80$, 95% CI [0.65, 0.99]), but this effect was weaker and only marginally reliable after Tukey correction ($p = .043$, p_{adj}=0.11). The visual and haptic interventions did not significantly differ from each other, $HR = 0.88$, 95% CI [0.72, 1.08], p_{adj}=0.21. Kaplan-Meier curves visualize these effects (see Figure 3).

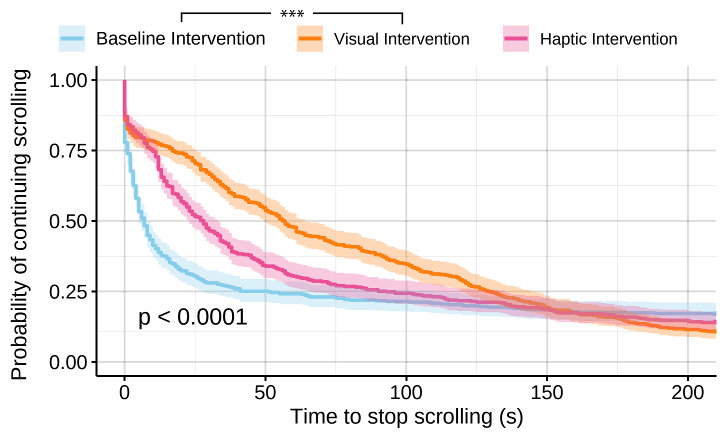


Fig. 3. Kaplan-Meier curves show the probability of the participants continuing to scroll after an intervention occurred.

Re-engagement Analysis. We analyzed re-engagement patterns between consecutive scrolling sessions to assess whether the interventions influenced users to resume scrolling directly after disengagement. Hence, we measured the break time from the disengagement of scrolling due to an intervention to when users began scrolling again in the next session. It is essential to note that only sessions in which participants exceeded the 15-minute threshold (triggering an intervention) were recorded; sessions in which users naturally disengaged before reaching this threshold generated no data points. Across 1,190 consecutive intervention-triggered sessions, 4.87% re-engaged within one minute, 11.5% within 3 minutes, 13.70% within five minutes, and 17.5% within ten minutes. When broken down by intervention type, rapid re-engagement (within 5 minutes) rates were similar: Baseline (12.5%), visual (14.2%), and haptic (14.4%). The median break time was 3.4 h overall, with slight variations across interventions: Baseline (Md = 4.3 h, IQR = 18.5 h), visual (Md = 3.3 h, IQR = 13.7 h), and haptic (Md = 3.2 h, IQR = 14.6 h). See Appendix A, Figure 10 for the histogram.

4.5.2 Subjective Effectiveness of the Gradual Interventions Over Time. RQ1 examines how participants' perceptions of intervention effectiveness change over time as the gradual interventions increase friction. To explore this, we analyzed how the subjective effectiveness developed in relation to the time it took participants to stop infinite scrolling (see Figure 4). The baseline intervention showed a peak in all ratings around 15 s after the pop-up appeared, but these ratings dropped and fluctuated afterwards (high standard error), suggesting a fading impact if no immediate action was taken. The haptic gradual intervention also peaked at around 15 s, but ratings then remained comparatively stable as the vibration intensity increased. The visual gradual intervention showed a different pattern: ratings stayed steady throughout without an initial peak, maintaining a level similar to the haptic intervention but lacking the early boost observed in the other two conditions.

To identify when interventions began to lose subjective effectiveness, we defined a *tipping point* as the moment when ratings dropped by at least one point on their respective Likert scales. Since measures were normalized, this corresponded to a drop of at least 1/7 for seven-point scales and 1/5 for the five-point scale of *reactance* (reversed). One Likert step represents the minimal perceptible change a participant could report. The weighted means (using the factor loadings from our exploratory factor analysis (see subsubsection 4.3.1) of the five subjective ratings were calculated. These tipping points are visualized in Figure 4 and summarized in Table 1.

The results indicate that ratings for the haptic intervention declined earlier than those for the visual intervention. For the haptic intervention, the subjective effectiveness dropped at 1 min 6 s, while the visual intervention's subjective effectiveness dropped at more than double the time (2 min 43 s). The baseline's subjective effectiveness

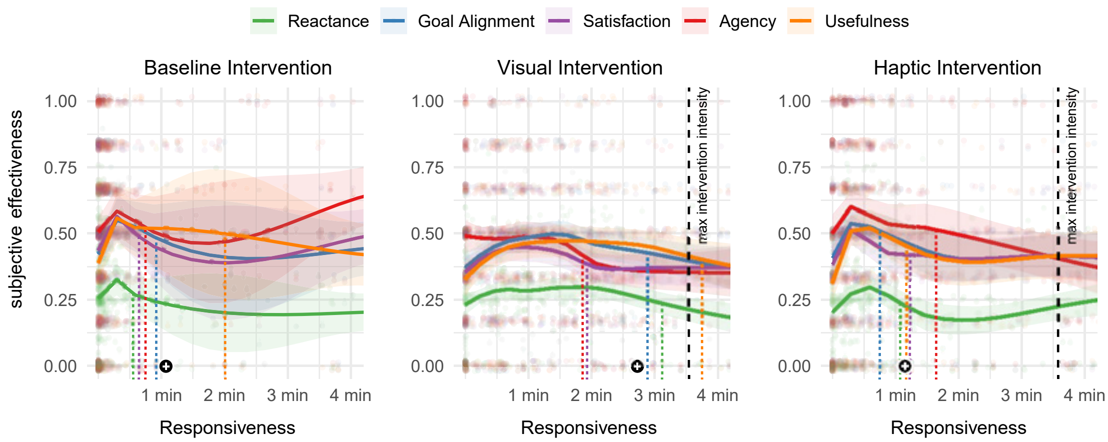


Fig. 4. Relationship between responsiveness and interventions' subjective ratings. Dots represent individual data points, and lines show LOESS-smoothed (*span* = 0.75, *default*). The x-axis shows the time participants needed to respond to the intervention by stopping infinite scrolling. After 3 min 31 s, the gradual interventions reached their maximum in intensity. The colored dashed lines mark the points at which each subjective rating dropped below at least one point on the respective Likert scale. The black dot, with the white cross, indicates the weighted mean where the subjective effectiveness drops.

Table 1. Tipping points where the subjective effectiveness drops by at least one step in the respective Likert scale

Subjective Measure	**Baseline Inter.**	**Visual Inter.**	**Haptic Inter.**
Agency	45 s	111 s	97 s
Goal Alignment	57 s	173 s	71 s
Satisfaction	39 s	115 s	42 s
Reactance (reversed)	31 s	185 s	64 s
Usefulness	120 s	221 s	69 s
Weighted Mean	*62.21 s*	*163.06 s*	*66.19 s*

dropped the earliest at 62.21 s. However, this timing should be interpreted cautiously due to the large standard error after 15 s.

4.5.3 Correlation Analysis Between Individual Traits. We computed pairwise Pearson correlations at the participant level to assess overlap among the four individual traits. Impulsivity and self-control showed the strongest negative correlation ($r = -.72$, $p < .001$). Strong correlations also emerged between self-control and FOMO ($r = -.62$, $p < .001$) and between anxiety and FOMO ($r = .60$, $p < .001$). The remaining pairs showed moderate correlations: self-control and anxiety ($r = -.54$, $p < .001$), impulsivity and FOMO ($r = .45$, $p < .001$), and impulsivity and anxiety ($r = .41$, $p < .001$).

4.5.4 Bayesian Mixed-Effect-Models. RQ2 addresses how contextual factors (within-subject) and individual traits (between-subject) moderate the effectiveness of interventions. We distinguish between subjective (user experience) and objective effectiveness (behavior change). Although our central focus is on moderation effects, we first report descriptive results to provide an overall orientation. Across all participants, the baseline intervention achieved the highest subjective and objective effectiveness (subjective: $M = 0.48$, $SD = 0.23$; objective: $M = 0.65$, $SD = 0.30$). The haptic gradual intervention followed (subjective: $M = 0.42$, $SD = 0.23$; objective: $M = 0.55$,

$SD = 0.25$), while the visual gradual intervention was rated lowest on both dimensions (subjective: $M = 0.39$, $SD = 0.21$; objective: $M = 0.51$, $SD = 0.26$).

Unlike classical models, Bayesian methods allow quantifying the significance and existence of effects, making them particularly suitable for small-sample and novel HCI contexts [48, 103]. Although we carefully simulated the sample size prior to our user study (see subsection 4.4), using Bayesian statistics enables us to perform a more robust analysis. While still uncommon in HCI research [48], Bayesian methods are increasingly applied in behavioral sciences [118].

Initially, we tested for normality of our data. A Shapiro–Wilk test indicated non-normality for subjective ($W = 0.98$, $p < .001$) and objective effectiveness ($W = 0.95$, $p < .001$). Therefore, we specified Student-t distributions. The t distribution is useful for Bayesian models "*when data appear to have outliers beyond what would be accommodated by a normal distribution*" [57, p. 458]. The model was fitted using Markov-chain Monte-Carlo (MCMC) sampling (4 chains, 11,000 iterations per chain, 1,000 warmup). As suggested by Lemoine [62], we used $B \sim \mathcal{N}(0, 1)$ as a weakly informative prior. Such priors act as a form of regularization, constraining estimates to plausible ranges and improving robustness, especially in complex models like ours, where even a reasonably large sample size (see subsection 4.4) may not fully counteract the risk of non-convergence, when not using priors [103]. The model structure was:

$$\text{Subj./Obj. Effectiveness} \sim \text{Inter. Type} \times (\text{Within-Factors} + \text{Between-Factors}) + (1 + \text{Inter. Type} + \text{Within-Factors} \mid \text{Participant})$$

The within-subject factors were valence, stress, sleepiness, multitasking, at home, current activity, and social situation; between-subject factors included impulsivity, anxiety, FOMO, stress, and self-control. The model for the subjective effectiveness showed substantial explanatory power (*Conditional* $R^2 = 0.80$ (95% CrI $[0.79, 0.81]$), *Marginal* $R^2 = 0.28$, (95% CrI $[0.21, 0.35]$) while the low model for the objective effectiveness was lower powered (*Conditional* $R^2 = 0.45$ (95% CrI $[0.40, 0.50]$), *Marginal* $R^2 = 0.12$, (95% CrI $[0.09, 0.17]$). All potential scale reduction factors for both models' parameters were below 1.01, indicating good convergence [11]. Trace plots of the MCMC permutations were inspected for divergent transitions. In addition, the effective sample sizes (ESS) exceeded 12,100 for all parameters, well above the $ESS > 10{,}000$ threshold suggested by Kruschke [57] for obtaining reasonably stable estimates.

Makowski et al. [69] describes an effect as possible existing when their $pd > 95\%$. Further, they mention that an effect is considered as probably significant when less than 2.5% of the 89% Highest Density Interval (HDI) lies inside the Region of Practical Equivalence (ROPE), indicating a meaningful deviation from the null region. According to Kruschke and Liddell [59], the ROPE range can be defined as a range corresponding to ± 0.1 of the standard deviation of the target variable [18]. We calculated the ROPE for the subjective effectiveness to be $[-0.02, 0.02]$ and for the objective effectiveness to be $[-0.03, 0.03]$. In the following, we report only the interaction terms that involve intervention type, as these directly address our RQ2. Several substantial main effects that replicate results from prior work [72, 96] were also identified, such as on valence and sleepiness. Additionally, impulsivity showed a significant main effect on subjective effectiveness (see Appendix D). However, they fall outside the scope of RQ2 and are therefore not detailed here.

We used the baseline intervention (Pop-up) as a reference condition, comparing it to the haptic and visual gradual intervention. Each effect is summarized by its posterior median, 95% CrI, probability of direction (pd), and percentage of the posterior distribution within ROPE. We followed the reporting guidelines of Kruschke [58] and Makowski et al. [69].

Subjective Effectiveness. Participants with higher impulsivity rated the visual intervention as less effective than the baseline (see Figure 5 left). This interaction between *Visual Intervention × Impulsivity*, observed on the subjective effectiveness as the weighted mean of the five subjective measures, probably exists ($pd = 98.69\%$; Md = -0.29, 95% CrI [-0.56, -0.03]) and is considered significant (0.00% in ROPE).

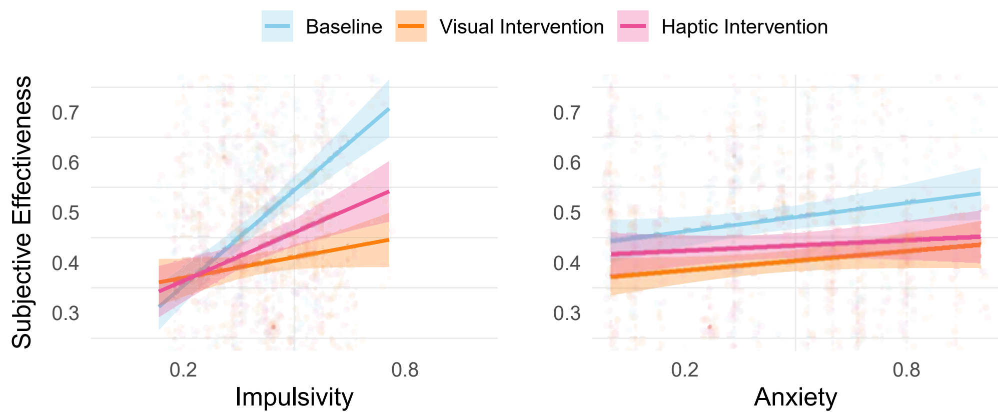


Fig. 5. Interaction effects for impulsivity, and anxiety with intervention type on **subjective effectiveness**

We further found a trend for *Visual Intervention × Anxiety* (Md = 0.12, 95% CrI [0.00, 0.23], pd = 97.81%, percentage in ROPE = 2.82%). While the pd is >95%, suggesting a possible existing effect, the percentage in ROPE shows undecided significance. Hence, the effect should be interpreted with caution (see Figure 5 right).

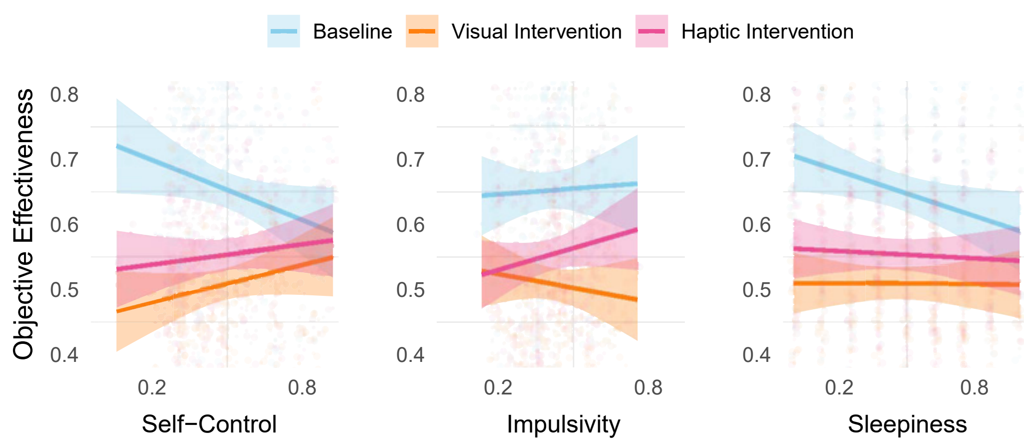


Fig. 6. Interaction effects for self-control, impulsivity, and sleepiness with intervention type on **objective effectiveness**

Objective Effectiveness. Participants with low self-control yielded more objective effectiveness with the baseline than with the haptic intervention, while participants with high self-control yielded similar effectiveness with both intervention types (see Figure 6 left). This interaction between *Haptic Intervention × Self-Control* (Md = 0.58, 95% CrI [0.18, 0.97], *pd* = 99.7%, 0% in ROPE) probably exists and is significant.

Further, for participants with low impulsivity, the haptic intervention yielded lower objective effectiveness than the baseline, a difference that diminished with increasing impulsivity (see Figure 6 center). This interaction between *Haptic Intervention × Impulsivity* (Md = 0.42, 95% CrI [-0.03, 0.86]) probably exists (*pd* = 96.6%) and is probably significant (1.91% in ROPE).

Last, for sleepy participants, the objective effectiveness of the visual intervention was comparable to the baseline, whereas for more alert participants, the baseline intervention was more objectively effective (see Figure 6 right). This interaction between *Visual Intervention × Sleepiness* (Md = 0.13, 95% CrI [0.01, 0.26], *pd* = 98.0%) likely exists, yet, as the percentage in ROPE is greater than 2.5% (2.8%), it has undecided significance and this trend should be handled with care.

Other Effects. Several other interactions, including those with FOMO, or contextual factors such as multitasking, being at home, current activity, and social situation, showed uncertain existence (pd < 95%) or undecided significance (ROPE > 2.5%). This indicates that these factors are unlikely to moderate the relationship between intervention type and intervention effectiveness meaningfully. Full results of the Bayesian model are provided in Appendix D. To assess the robustness of these findings, we conducted sensitivity analyses by re-calculating both models under two alternative participant filtering conditions ($N = 150$ and $N = 214$; see Appendix E). The *Visual Intervention* × *Anxiety* (subjective effectiveness) and *Haptic Intervention* × *Self-Control* (objective effectiveness) interactions remained significant across all filtering. The interaction between impulsivity and the intervention types attenuated when participants without complete within-person exposure were included.

5 Discussion

This study examined how the effectiveness of interventions for mitigating infinite scrolling on short video SoMe platforms is shaped by situational context and individual differences in the traits impulsivity, self-control, anxiety, and FOMO. We compared three designs: a baseline pop-up intervention, a gradually intensifying visual overlay, and a gradually intensifying haptic vibration. We assumed that no single intervention works best for all situations or all users. To answer our two RQs (see section 1), we conducted a 7-day field study ($N = 104$) using a custom Android app to track scrolling on TikTok, Instagram, Facebook, and YouTube Shorts, with interventions triggering after 15 min. During the study, upon stopping their scrolling activity, we assessed the intervention's effectiveness via objective and subjective measures and asked participants to report their current context (e.g., sleepiness, valence, etc.).

5.1 User Experience as Intervention Intensity Increases (RQ1)

To address user experience when the friction of interventions increases (see RQ1), we analyzed when subjective effectiveness declined, interpreting these declines as tipping points that may indicate the intervention was perceived as too intrusive. The baseline pop-up intervention worked well initially, but quickly lost subjective impact when participants did not disengage immediately and ignored it. The haptic gradual intervention extended subjective effectiveness somewhat beyond the baseline, but its ratings also declined after about 66 s, likely because vibration intensity quickly became intrusive. In contrast, our gradual visual intervention maintained stable ratings for much longer, with declines occurring only after about 163 s. While we can assume that it was better tolerated, it also yielded significantly slower behavioral responses than the baseline. These differences might be explained by the *orienting response* [108], where unexpected signals in a different sensory modality automatically redirect attention. Because infinite scrolling is mainly a visually immersive activity, haptic vibrations triggered a stronger orienting response, making them harder to ignore but also more disruptive. In contrast, the increasing dots of the visual intervention blended into the visual stream of infinite scrolling, which may make them less intrusive and more naturally integrated, but also slower to prompt disengagement.

5.2 Self-Regulation Moderates the Effectiveness of Interventions (RQ2)

Subjective and objective effectiveness of the baseline pop-up for impulsivity diverged. Objectively, impulsivity did not affect disengagement time, but subjectively, highly impulsive participants rated the intervention as significantly more effective than less impulsive ones. One interpretation is that impulsive users may recognize their difficulty in self-regulating and appreciate the explicit cue of the pop-up as a supportive nudge on their behavior. Even if their stopping times are not substantially shorter, the explicit intervention may resonate with their self-perception of struggling to disengage, which makes the intervention subjectively experienced more effective. In contrast, less impulsive users, who rely more successfully on internal regulation, may perceive the same prompt as unnecessary, resulting in lower subjective ratings. However, subjective ratings were collected only from sessions in which participants stopped scrolling, so individual differences in response may also contribute to how the intervention was experienced.

The divergence between subjective and objective effectiveness for impulsivity directly connects to the interaction effects with self-control on objective effectiveness: The baseline pop-up was objectively most effective for users with low self-control, but its impact declined for people with higher self-control. In contrast, the gradual interventions maintained relatively stable effectiveness across different levels of self-control. Taken together, these findings highlight that explicit interventions like our pop-up design are especially valuable for users who struggle with self-regulation (high impulsivity or low self-control). This interpretation aligns with Mark et al. [71] showing that people with lower self-control perceive intervention as more beneficial than those with high self-control. While prior work has well established that traits like low self-control predict higher vulnerability to problematic SoMe use [106], our results explicitly extend this relationship: individual traits do not merely predict problematic use, but *moderate* the effectiveness of behavioral interventions.

5.2.1 Limited Interaction Between Contextual Factors and Intervention Type. Among all contextual factors, only sleepiness showed a likely existing interaction effect with intervention type. For sleepy participants, all interventions were similarly effective, whereas alert participants disengaged faster with the baseline pop-up. These findings resonate with research on bedtime procrastination, where people continue late-night scrolling despite intending to stop, often due to reduced self-regulatory capacity [56]. In such situations, users' intent to continue scrolling may override any intervention regardless of its design. Yet while this effect likely existed, it did not reach significance (2.83% in ROPE). The limited moderation of the contextual factors aligns with Roffarello and de Russis [97], who argue that "*users consider their behaviors problematic independently of their contextual situation*" [97, p. 11]. Yet, this contrasts prior work showing that contextual factors such as valence and sleepiness do affect overall intervention effectiveness as main effects [72]. Our results suggest that these two findings are compatible, as contextual factors may influence the overall effectiveness of interventions (main effects), but they do not appear to determine which of our three intervention types is more effective than the others (interaction effects). Hence, stable individual differences appear to be more dominant than momentary contextual factors.

5.3 The Baseline Outperformed the Gradual Interventions

The baseline pop-up showed higher median values than both gradual interventions across agency and responsiveness (see Appendix A), diverging from prior design-friction insights [64, 101]. One explanation for the slower time-to-stop-scrolling might concern insufficient salience of the gradual interventions. As Cox et al. [20] emphasize, interventions must interrupt automatic behavior and trigger deliberate thought. When friction is too subtle, users may adapt without stopping their habituated scrolling. Our results suggest that despite the growing call for subtle design friction in digital well-being [6, 20, 73, 101], participants with low self-control or high impulsivity responded better to the explicit baseline pop-up than to the gradual interventions (see Figure 6). For these users, the baseline pop-up might have provided the clear interruption needed to break infinite scrolling, while the gradual interventions did not. Consequently, moving away from explicit intervention toward more subtle friction designs may actively undermine intervention efficacy for those who need it most.

5.4 Balancing Subjective and Objective Effectiveness in Intervention Designs

Our findings reveal a central design dilemma: should designers create interventions for objectively reducing screen time or for a subjective positive experience? Prior work has cautioned against evaluating digital well-being interventions solely by screen time [2, 40, 65]. Our results also show a divergence between objective and subjective effectiveness: interventions that quickly reduce scrolling are perceived as frustrating, while those rated more positively often delay disengagement. This tension can be interpreted through the distinction between the *experiencing self* and the *remembering self* [44]. The experiencing self evaluates interventions in the moment, where abrupt disruptions like pop-ups may feel intrusive or irritating. The remembering self reflects afterwards on whether the intervention ultimately supported meaningful goals, such as spending time with friends [114]. This suggests that effective intervention must not only nudge users to stop scrolling, but also be framed so that users can look back on the experience as supportive rather than coercive. Designing for the remembering self may justify interventions that momentarily frustrate, as long as they contribute to longer-term well-being.

5.5 Toward Tailored Interventions

Our findings indicate that individual differences in self-regulation, specifically impulsivity and self-control, moderate the effectiveness of the three interventions we tested. If this moderation pattern generalizes beyond our three designs, it would suggest that interventions could benefit from being adapted to users' self-regulation abilities rather than being deployed uniformly. Related work on adaptive interventions for smartphone overuse has already demonstrated that tailoring intervention timing to contextual factors such as location, activity, and app usage patterns can improve intervention accuracy [85]. Our findings add to this by indicating that stable individual traits may also inform the selection of intervention type. A central challenge for this tailoring is identifying relevant traits without requiring users to complete questionnaires. Recent work suggests that smartphone-usage data itself can serve as a proxy for stable individual differences. For instance, Stachl et al. [110] demonstrated that app usage and communication patterns can predict users' Big Five personality traits, particularly conscientiousness and extraversion. Similarly, Wen et al. [124] showed that impulsivity can be inferred from passively collected phone data such as call logs and charging behavior. Because conscientiousness is closely linked to self-control [80], passively detected personality traits could provide an indirect predictor for self-regulation ability. This opens up the possibility of adaptive systems that adjust intervention type or intensity to individual needs. For example, offering more explicit interventions to highly impulsive users while relying on subtler gradual friction for those with greater self-control. Such trait-based tailoring would also be efficient to operate. Because self-regulation traits are stable, the inference step needs to run only rarely. This contrasts with context-driven adaptation, where systems must sense the user's context at every moment to decide when or how to intervene [85]. Thus, continuous sensing incurs recurring costs in battery power, computation, and access to potentially sensitive signals such as location or activity. Our results suggest that, at least for selecting the intervention type, this cost may be avoidable, as the contextual factors we measured showed little moderating influence, whereas stable traits did.

5.6 Limitations

Several limitations should be considered. First, our participant pool was limited to Android users in the US and UK, excluding iOS users, which limits the generalizability of the results. Second, our re-engagement analysis (see section 4.5.1) showed that 13.7% of consecutive sessions resumed within 5 min. Such rapid returns may reflect natural scrolling behavior (e.g., briefly replying to a message before continuing to scroll), but could also indicate attempts to intentionally increase compensation by triggering additional data points. Because rapid re-engagement may also indicate limited intervention effectiveness, we retained all such data to avoid artificially removing behaviors that may occur in real use. However, this choice prevents us from distinguishing re-engagement from incentive-driven behavior, limiting the precision of our behavioral interpretation. A related measurement limitation concerns our use of time-to-stop as objective effectiveness. This measure inherently reflects both the time needed to perceive an intervention and the time needed to decide to stop scrolling. Since gradual interventions introduce friction gradually, their effects may unfold over a longer perceptual timescale than sudden pop-ups. As a result, direct comparisons across these intervention types are imperfect, and responsiveness may systematically advantage abrupt cues. We nevertheless used this metric because it is one of the standard metrics in intervention research [40, 51, 64, 72, 85]. An additional limitation concerns our ESM design. Because the subjective effectiveness questionnaire was triggered only after participants stopped scrolling, all subjective ratings and contextual reports are conditioned on the decision to disengage. This introduces a potential self-selection bias, as differences in subjective ratings across interventions or context may partly reflect who stopped rather than how effective the intervention was perceived to be. Nonetheless, we chose this post-session timing because collecting subjective measures during scrolling would have interrupted the activity under study.

Our study examined only three interventions grounded in prior work [52, 82, 101], representing a narrow subset of possible friction designs. Their specific parameters (e.g., the 3 min 31 s linear increase, visual/haptic designs, or the 15 min trigger threshold) represent only one design in a larger space. Alternative friction designs, such as faster/slower increases, different slopes, or hybrid designs combining pop-ups with gradual friction, might yield different results. Thus, our findings highlight the importance of specific design choices rather than implying general implications of friction-based approaches. Still, the main effects of individual differences that emerged independently of the intervention type (see Appendix D) suggest that, besides our limited design space, individual differences shape how users respond to interventions. Further, because intervention types were assigned by full randomization at each decision point, 46 participants received three or more interventions in total yet were never exposed to all three types. A counterbalanced randomization system that cycles through all conditions before repeating any type would have prevented this attrition while still preserving the benefits of random assignment.

Finally, the study lasted only 7 days. While this ensured ecologically valid data collection, longer-term effects such as habituation remain unexplored [54], and meaningful behavioral changes typically emerge over several weeks (e.g., 13 weeks [39]). The short study duration may also blur the comparison of intervention types, because the designs differed in how familiar they were to participants. The baseline is based on the screen-time reminders that platforms already show their users [42, 116], whereas the gradual interventions were new to participants. Part of the baseline's advantage in responsiveness and ratings may therefore stem from users recognizing a familiar intervention rather than from the design itself, an advantage that could fade as users grow equally accustomed to the gradual interventions. Over longer durations, these differences could shift. The trait moderation, in contrast, rests on stable individual differences and should be less sensitive to the study duration, yet whether it persists over months remains open. Further, habituation poses a particular challenge for non-adaptive interventions, as users may initially respond to them but gradually ignore them as novelty fades [54]. This limitation strengthens the case for interventions that can dynamically adjust as users become habituated, potentially maintaining effectiveness over longer-term use.

5.7 Future Work

Future work should explore a broader design space of intervention designs. An interesting approach to optimize these design parameters could be Multi-Objective Bayesian Optimization, an algorithmic method that efficiently explores design parameter spaces to optimize for objective and subjective intervention effectiveness. This method showed success in related UI optimization challenges [10, 30, 43]. Whether our moderation findings translate into practical benefits when used to tailor interventions to individual traits remains an open question. Since such traits can be inferred from phone usage patterns [110, 124], future work should investigate whether adaptive systems that dynamically adjust intervention parameters based on detected traits outperform fixed intervention designs. However, the same technology that enables personalization also enables surveillance, requiring careful attention to user privacy. Therefore, adaptive interventions should strike a balance between tailoring and privacy, prioritizing people's well-being over optimizing for (dis-)engagement alone.

5.8 Policy Implications

Our findings have implications for ongoing regulatory debates, especially those related to Article 28 of the Digital Services Act [33], which requires online platforms to ensure a high level of privacy, safety, and security for minors. The accompanying guidelines specify what this means for intervention design. They recommend "*information or friction that slows down content display [...] giving users an opportunity to think before they decide if they want to see more content*" [33, §57.b.ix], and at the same time define interventions as effective if they"*deter minors from spending more time on the platform*" [33, §61.c]. Therefore, the guidelines define intervention effectiveness as reduced usage time. Our results suggest this criterion is too narrow. Participants with low self-control and high impulsivity benefited from the explicit baseline pop-up but not from either gradual intervention (see subsubsection 4.5.4), even though gradual designs are the kind of low-friction implementation a usage-reduction criterion would most readily accept. Additionally, since users dismiss or abandon unacceptable interventions [66, 81], platforms have their own incentive to favor seamless designs. A usage-reduction criterion does not counter this, leaving limited pressure to develop stronger designs for the users most at risk.

We therefore suggest two refinements. First, audits should report outcomes separately for user groups defined by self-regulation traits such as self-control and impulsivity, and should combine objective behavior change with subjective measures, rather than relying on population averages or on usage reduction alone. Second, regulation should specify measurable outcomes, such as low rates of intervention dismissal, rather than prescribing particular designs, since no single design in our study served every user equally.

6 Conclusion

This paper investigated whether the effectiveness of different intervention types for mitigating infinite scrolling depends on users' context and individual traits. In a 7-day field study with $N = 104$ participants, we compared a baseline pop-up with gradually intensifying visual and haptic interventions. By investigating both objective and subjective intervention effectiveness, we found that users responded differently to the three interventions: the baseline pop-up yielded the fastest disengagement but quickly lost influence if ignored, the haptic intervention disrupted scrolling earlier but was subjectively perceived as less effective as intensity increased, and the visual intervention was slower to trigger disengagement but maintained higher subjective effectiveness over time. These results highlight that interventions that effectively reduce screen time may not always be what users find acceptable. Our Bayesian analysis further showed that individual differences in self-regulation, specifically self-control and impulsivity, shaped how participants responded to the interventions, while contextual factors showed only limited moderation of which intervention worked most effectively. Participants with low self-control benefited most from the baseline pop-up, which offered an explicit prompt to stop scrolling, while those with higher self-control were less dependent on our tested intervention types. Overall, our findings show that intervention design should balance objective impact with subjective experience and account for differences in individuals' self-regulation traits. However, because both gradual interventions were outperformed by a simple pop-up in our study, the moderation effects we observed should be understood as motivation for future research on tailored interventions rather than as direct evidence that tailoring improves outcomes.

Open Science

The Android application used for this study, the study data, and the R scripts used for analysis are openly available at: https://github.com/luca-maxim/Cant_Stop. To protect participant privacy, all study data is shared in anonymized form, with Prolific IDs replaced by unique sequential participant identifiers.

Acknowledgments

This research was supported by the German Research Foundation (DFG) through the project "Beyond Screen Time: Context- and Content-tailored Interventions to Social Media Usage to Enhance Digital Well-being " (project number: 561828495). We are further grateful to Lukas Gruler for building an initial web-based feasibility study of the gradual interventions, to Maryam Elhaidary for drafting early Android implementations of the gradual interventions, and to Albin Zeqiri for exploring first approaches to analyzing the user data. Thanks to RHCP for the main theme of this work.

References

[1] Elgin Akpinar, Yeliz Yeşilada, and Pinar Karagöz. 2023. Effect of Context on Smartphone Users' Typing Performance in the Wild. *ACM Transactions on Computer-Human Interaction* 30, 3 (2023), 1–44. doi:10.1145/3577013

[2] Sultan Almoallim and Corina Sas. 2022. Toward Research-Informed Design Implications for Interventions Limiting Smartphone Use: Functionalities Review of Digital Well-being Apps. *JMIR Form Res* 6, 4 (19 Apr 2022), e31730. doi:10.2196/31730

[3] Android. 2025. *VibrationEffect.createOneShot.* Android Developers. Accessed: 2025-11-19. https://developer.android.com/reference/android/os/VibrationEffect#createOneShot(long,%20int)

[4] Amanda Baughan, Mingrui Ray Zhang, Raveena Rao, Kai Lukoff, Anastasia Schaadhardt, Lisa D. Butler, and Alexis Hiniker. 2022. "I Don't Even Remember What I Read": How Design Influences Dissociation on Social Media. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–13. doi:10.1145/3491102.3501899

[5] Joseph Bayer, Nicole Ellison, Sarita Schoenebeck, Erin Brady, and Emily B. Falk. 2018. Facebook in context(s): Measuring emotional responses across time and space. *New Media & Society* 20, 3 (2018), 1047–1067. doi:10.1177/1461444816681522

[6] Steve Benford, Chris Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, and Tom Rodden. 2012. Uncomfortable interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Austin Texas USA, 2005–2014. doi:10.1145/2207676.2208347

[7] Ine Beyens, J Loes Pouwels, Irene I van Driel, Loes Keijsers, and Patti M Valkenburg. 2020. The effect of social media on well-being differs from adolescent to adolescent. *Scientific Reports* 10, 1 (2020), 1–11. doi:10.1038/s41598-020-67727-7

[8] M. M. Bradley and P. J. Lang. 1994. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59. doi:10.1016/0005-7916(94)90063-9

[9] Jack W Brehm. 1966. A theory of psychological reactance. (1966).

[10] Eric Brochu, Tyson Brochu, and Nando De Freitas. 2010. A Bayesian interactive optimization approach to procedural animation design. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. ACM, New York, NY, USA, 103–112.

[11] Stephen P. Brooks and Andrew Gelman. 1998. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics* 7, 4 (1998), 434–455. arXiv:https://www.tandfonline.com/doi/pdf/10.1080/10618600.1998.10474787 doi:10.1080/10618600.1998.10474787

[12] Matthias Böhmer, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Bauer. 2011. Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, Stockholm Sweden, 47–56. doi:10.1145/2037373.2037383

[13] Yung-Ju Chang, Gaurav Paruthi, and Mark W. Newman. 2015. A Field Study Comparing Approaches to Collecting Annotated Activity Data in Real-World Settings. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 671–682. doi:10.1145/2750858.2807524

[14] Yu-Chun Chen, Yu-Jen Lee, Kuei-Chun Kao, Jie Tsai, En-Chi Liang, Wei-Chen Chiu, Faye Shih, and Yung-Ju Chang. 2023. Are You Killing Time? Predicting Smartphone Users' Time-killing Moments via Fusion of Smartphone Sensor Data and Screenshots. In *CHI*, Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, Anicia Peters, Stefanie Mueller, Julie R. Williamson, and Max L. Wilson (Eds.). ACM, New York, NY, USA, 1–19. doi:10.1145/3544548.3580689

[15] Francesco Chiossi, Luke Haliburton, Changkun Ou, Andreas Martin Butz, and Albrecht Schmidt. 2023. Short-Form Videos Degrade Our Capacity to Retain Intentions: Effect of Context Switching On Prospective Memory. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, Anicia Peters, Stefanie

Mueller, Julie R. Williamson, and Max L. Wilson (Eds.). ACM, New York, NY, USA, 1–15. doi:10.1145/3544548.3580778

[16] Hyunsung Cho, DaEun Choi, Donghwi Kim, Wan Ju Kang, Eun Kyoung Choe, and Sung-Ju Lee. 2021. Reflect, not Regret: Understanding Regretful Smartphone Use with App Feature-Level Analysis. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–36. doi:10.1145/3479600

[17] Minjoo Cho and Daniel Saakes. 2017. Calm Automaton. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, Gloria Mark, Susan Fussell, Cliff Lampe, m.c. schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor (Eds.). ACM, New York, NY, USA, 393–396. doi:10.1145/3027063.3052968

[18] Jacob Cohen. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed ed.). L. Erlbaum Associates, Hillsdale, N.J.

[19] Jacob Cohen. 1992. A power primer. *Psychological Bulletin* 112, 1 (1992), 155–159. doi:10.1037/0033-2909.112.1.155

[20] Anna L. Cox, Sandy J.J. Gould, Marta E. Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, San Jose California USA, 1389–1397. doi:10.1145/2851581.2892410

[21] D. R. Cox. 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 34, 2 (1972), 187–220. http://www.jstor.org/stable/2985181

[22] Michelle G Craske and Murray B Stein. 2016. Anxiety. *The Lancet* 388, 10063 (Dec. 2016), 3048–3059. doi:10.1016/S0140-6736(16)30381-6

[23] Egon Dejonckheere, Febe Demeyer, Birte Geusens, Maarten Piot, Francis Tuerlinckx, Stijn Verdonck, and Merijn Mestdagh. 2022. Assessing the reliability of single-item momentary affective measurements in experience sampling. *Psychological assessment* 34, 12 (2022), 1138–1154. doi:10.1037/pas0001178

[24] Sarah Diefenbach and Kim Borrmann. 2019. The Smartphone as a Pacifier and Its Consequences: Young Adults' Smartphone Usage in Moments of Solitude and Correlations to Self-Reflection. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300536

[25] James Price Dillard and Lijiang Shen. 2005. On the Nature of Reactance and its Role in Persuasive Health Communication. *Communication Monographs* 72, 2 (2005), 144–168. doi:10.1080/03637750500111815

[26] Xiang Ding, Jing Xu, Honghao Wang, Guanling Chen, Herpreet Thind, and Yuan Zhang. 2016. WalkMore: promoting walking with just-in-time context-aware prompts. In *2016 IEEE Wireless Health (WH)*. 1–8. doi:10.1109/WH.2016.7764558

[27] Christine DiStefano, Min Zhu, and Diana Mîndrilã. [n. d.]. Understanding and Using Factor Scores: Considerations for the Applied Researcher. ([n. d.]). Publisher: University of Massachusetts Amherst. doi:10.7275/DA8T-4G52

[28] Trinh Minh Tri Do, Jan Blom, and Daniel Gatica-Perez. 2011. Smartphone Usage in the Wild: A Large-Scale Analysis of Applications and Context. In *Proceedings of the 13th International Conference on Multimodal Interfaces* (Alicante, Spain) *(ICMI '11)*. Association for Computing Machinery, New York, NY, USA, 353–360. doi:10.1145/2070481.2070550

[29] Norman R Draper and Harry Smith. 1998. *Applied regression analysis*. Vol. 326. John Wiley & Sons. doi:10.1002/9781118625590

[30] John J. Dudley, Jason T. Jacques, and Per Ola Kristensson. 2019. Crowdsourcing Interface Feature Design with Bayesian Optimization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300482

[31] Patrick Ehrenbrink. 2020. Reactance Scale for Human–Computer Interaction. In *The Role of Psychological Reactance in Human–Computer Interaction*, Patrick Ehrenbrink (Ed.). Springer International Publishing, Cham, 71–81. doi:10.1007/978-3-030-30310-5_8

[32] European Commission. 2024. *Commission opens formal proceedings against TikTok under the Digital Services Act.* , 19 February 2024. https://ec.europa.eu/commission/presscorner/detail/en/IP_24_926

[33] European Commission. 2025. *Communication from the Commission – Guidelines on measures to ensure a high level of privacy, safety and security for minors online, pursuant to Article 28(4) of Regulation (EU) 2022/2065.* Official Journal of the European Union, 10 October 2025. http://data.europa.eu/eli/C/2025/5519/oj

[34] Deborah Kirby Forgays, Ira Hyman, and Jessie Schreiber. 2014. Texting everywhere for everything: Gender and age differences in cell phone etiquette and use. *Computers in Human Behavior* 31 (2014), 314–321. doi:10.1016/j.chb.2013.10.053

[35] Google Play Store. [n. d.]. *Use of the AccessibilityService API.* Retrieved Aug 10, 2024 from https://support.google.com/googleplay/android-developer/answer/10964491?hl=en

[36] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Regan Mandryk, Mark Hancock, Mark Perry, and Anna Cox (Eds.). ACM, New York, NY, USA, 1–14. doi:10.1145/3173574.3174108

[37] Zhihua Guo, Shuyi Liang, Lei Ren, Tianqi Yang, Rui Qiu, Yang He, and Xia Zhu. 2022. Applying network analysis to understand the relationships between impulsivity and social media addiction and between impulsivity and problematic smartphone use. *Frontiers in Psychiatry* 13 (2022). doi:10.3389/fpsyt.2022.993328

[38] Şahin Gökçearslan, Filiz Kuşkaya Mumcu, Tülin Haşlaman, and Yasemin Demiraslan Çevik. 2016. Modelling smartphone addiction: The role of smartphone usage, self-regulation, general self-efficacy and cyberloafing in university students. *Computers in Human Behavior* 63 (Oct. 2016), 639–649. doi:10.1016/j.chb.2016.05.091

[39] Luke Haliburton, David Joachim Grüning, Frederik Riedel, Albrecht Schmidt, and Nađa Terzimehić. 2024. A Longitudinal In-the-Wild Investigation of Design Frictions to Prevent Smartphone Overuse. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–16. doi:10.1145/3613904.3642370

[40] Alexis Hiniker, Sungsoo Hong, Tadayoshi Kohno, and Julie A. Kientz. 2016. MyTime: Designing and Evaluating an Intervention for Smartphone Non-Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4746–4757. doi:10.1145/2858036.2858403

[41] Daniel Hintze, Philipp Hintze, Rainhard D. Findling, and René Mayrhofer. 2017. A Large-Scale, Long-Term Analysis of Mobile Device Usage Characteristics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 13 (jun 2017), 21 pages. doi:10.1145/3090078

[42] Instagram. 2025. Set a daily time limit on Instagram. https://help.instagram.com/2049425491975359. Accessed: 2025-03-14.

[43] Florian Kadner, Yannik Keller, and Constantin Rothkopf. 2021. AdaptiFont: Increasing Individuals' Reading Speed with a Generative Font Model and Bayesian Optimization. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 585, 11 pages. doi:10.1145/3411764.3445140

[44] Daniel Kahneman. 2024. *Thinking, fast and slow* (reissued ed.). Penguin Books, London.

[45] E. L. Kaplan and Paul Meier. 1958. Nonparametric Estimation from Incomplete Observations. *J. Amer. Statist. Assoc.* 53, 282 (1958), 457–481. arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1958.10501452 doi:10.1080/01621459.1958.10501452

[46] Pasi Karppinen, Harri Oinas-Kukkonen, Tuomas Alahäivälä, Terhi Jokelainen, Anna-Maria Teeriniemi, Tuire Salonurmi, and Markku J Savolainen. 2018. Opportunities and challenges of behavior change support systems for enhancing habit formation: A qualitative study. *Journal of biomedical informatics* 84 (2018), 82–92. doi:10.1016/j.jbi.2018.06.012

[47] Despoina Karvounides, Patricia M. Simpson, William H. Davies, Khushnuma A. Khan, Susan J. Weisman, and Kathryn R. Hainsworth. 2016. Three studies supporting the initial validation of the stress numerical rating scale-11 (Stress NRS-11): A single item measure of momentary stress for adolescents and adults. *Pediatric Dimensions* 1, 4 (2016), 105–109. doi:10.15761/PD.1000124

[48] Matthew Kay, Gregory L. Nelson, and Eric B. Hekler. 2016. Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 4521–4532. doi:10.1145/2858036.2858465

[49] Jaejeung Kim, Hayoung Jung, Minsam Ko, and Uichin Lee. 2019. GoalKeeper. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1 (2019), 1–29. doi:10.1145/3314403

[50] Jaejeung Kim, Joonyoung Park, Hyunsoo Lee, Minsam Ko, and Uichin Lee. 2019. LocknType. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Stephen Brewster, Geraldine Fitzpatrick, Anna Cox, and Vassilis Kostakos (Eds.). ACM, New York, NY, USA, 1–12. doi:10.1145/3290605.3300927

[51] Jaejeung Kim, Joonyoung Park, Hyunsoo Lee, Minsam Ko, and Uichin Lee. 2019. LocknType. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Stephen Brewster, Geraldine Fitzpatrick, Anna Cox, and Vassilis Kostakos (Eds.). ACM, New York, NY, USA, 1–12. doi:10.1145/3290605.3300927

[52] Stephanie L. Kincaid. 2023. Gradual Change Procedures in Behavior Analysis. *Behavior Analysis in Practice* 16, 1 (March 2023), 117–126. doi:10.1007/s40617-022-00689-6

[53] Minsam Ko, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Daehee Shin, Koji Yatani, Junehwa Song, and Kyong-Mee Chung. 2015. NUGU: A Group-based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. (2015), 1235–1245. doi:10.1145/2675133.2675244

[54] Geza Kovacs, Zhengxuan Wu, and Michael S. Bernstein. 2018. Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 1–25. doi:10.1145/3274364

[55] Hayri Koç, Zeynep Şimşir Gökalp, and Tolga Seki. 2023. The Relationships Between Self-Control and Distress Among Emerging Adults: A Serial Mediating Roles of Fear of Missing Out and Social Media Addiction. *Emerging Adulthood* 11, 3 (June 2023), 626–638. doi:10.1177/21676968231151776

[56] Floor M. Kroese, Denise T. D. de Ridder, Catharine Evers, and Marieke A. Adriaanse. 2014. Bedtime procrastination: introducing a new area of procrastination. *Frontiers in psychology* 5 (2014), 611. doi:10.3389/fpsyg.2014.00611

[57] John K. Kruschke. 2015. *Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan* (edition 2 ed.). Academic Press, Boston.

[58] John K. Kruschke. 2021. Bayesian Analysis Reporting Guidelines. *Nature Human Behaviour* 5, 10 (Aug. 2021), 1282–1291. doi:10.1038/s41562-021-01177-7

[59] John K. Kruschke and Torrin M. Liddell. 2018. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review* 25, 1 (Feb. 2018), 178–206. doi:10.3758/s13423-016-1221-4

[60] Levi Kumle, Melissa L-H Võ, and Dejan Draschkow. 2021. Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior research methods* 53, 6 (2021), 2528–2543. doi:10.3758/s13428-021-01546-0

[61] Hao-Ping Lee, Yi-Shyuan Chiang, Lan Gao, Stephanie Yang, Philipp Winter, and Sauvik Das. 2025. Purpose Mode: Reducing Distraction Through Toggling Attention Capture Damaging Patterns on Social Media Websites. *ACM Transactions on Computer-Human Interaction* (Jan. 2025), 3711841. doi:10.1145/3711841

[62] Nathan P. Lemoine. 2019. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* 128, 7 (July 2019), 912–928. doi:10.1111/oik.05985

[63] Qianqian Li, Tianlong Chen, Shujing Zhang, Chuanhua Gu, and Zongkui Zhou. 2025. The mediating role of intentional self-regulation in the constructive and pathological compensation processes of problematic social networking use. *Addictive Behaviors* 160 (Jan. 2025), 108188. doi:10.1016/j.addbeh.2024.108188

[64] Tao Lu, Hongxiao Zheng, Tianying Zhang, Xuhai “Orson” Xu, and Anhong Guo. 2024. InteractOut: Leveraging Interaction Proxies as Input Manipulation Strategies for Reducing Smartphone Overuse. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–19. doi:10.1145/3613904.3642317

[65] Kai Lukoff. 2019. *Digital wellbeing is way more than just reducing screen time.* Retrieved Feb 16, 2023 from https://uxdesign.cc/digital-wellbeing-more-than-just-reducing-screen-time-46223db9f057

[66] Kai Lukoff, Ulrik Lyngs, and Lize Alberts. 2022. Designing to support autonomy and reduce psychological reactance in Digital Self-Control Tools.

[67] Kai Lukoff, Ulrik Lyngs, Karina Shirokova, Raveena Rao, Larry Tian, Himanshu Zade, Sean A. Munson, and Alexis Hiniker. 2023. SwitchTube: A Proof-of-Concept System Introducing “Adaptable Commitment Interfaces” as a Tool for Digital Wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI ’23)*. Association for Computing Machinery, New York, NY, USA, Article 197, 22 pages. doi:10.1145/3544548.3580703

[68] Ulrik Lyngs, Kai Lukoff, Laura Csuka, Petr Slovák, Max Van Kleek, and Nigel Shadbolt. 2022. The Goldilocks level of support: Using user reviews, ratings, and installation numbers to investigate digital self-control tools. *International Journal of Human-Computer Studies* 166 (2022), 102869. doi:10.1016/j.ijhcs.2022.102869

[69] Dominique Makowski, Mattan S. Ben-Shachar, S. H. Annabel Chen, and Daniel Lüdecke. 2019. Indices of Effect Existence and Significance in the Bayesian Framework. *Frontiers in Psychology* 10 (Dec. 2019), 2767. doi:10.3389/fpsyg.2019.02767

[70] Tianxin Mao, Weigang Pan, Yingying Zhu, Jian Yang, Qiaoling Dong, and Guofu Zhou. 2018. Self-control mediates the relationship between personality trait and impulsivity. *Personality and Individual Differences* 129 (July 2018), 70–75. doi:10.1016/j.paid.2018.03.013

[71] Gloria Mark, Mary Czerwinski, and Shamsi T. Iqbal. 2018. Effects of Individual Differences in Blocking Workplace Distractions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–12. doi:10.1145/3173574.3173666

[72] Luca-Maxim Meinhardt, Maryam Elhaidary, Mark Colley, Michael Rietzler, Jan Ole Rixen, Aditya Kumar Purohit, and Enrico Rukzio. 2025. Scrolling in the Deep: Analysing Contextual Influences on Intervention Effectiveness during Infinite Scrolling on Social Media. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–17. doi:10.1145/3706598.3713187

[73] Thomas Mejtoft, Sarah Hale, and Ulrik Söderström. 2019. Design Friction. In *Proceedings of the 31st European Conference on Cognitive Ergonomics*. ACM, BELFAST United Kingdom, 41–44. doi:10.1145/3335082.3335106

[74] Thomas Mildner and Gian-Luca Savino. 2021. Ethical User Interfaces: Exploring the Effects of Dark Patterns on Facebook. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, and Takeo Igarashi (Eds.). ACM, New York, NY, USA, 1–7. doi:10.1145/3411763.3451659

[75] Aimee E. Miller-Ott and Lynne Kelly. 2017. A Politeness Theory Analysis of Cell-Phone Usage in the Presence of Friends. *Communication Studies* 68, 2 (2017), 190–207. arXiv:https://doi.org/10.1080/10510974.2017.1299024 doi:10.1080/10510974.2017.1299024

[76] Shalini Misra, Lulu Cheng, Jamie Genevie, and Miao Yuan. 2016. The iPhone Effect: The Quality of In-Person Social Interactions in the Presence of Mobile Devices. *Environment and Behavior* 48, 2 (2016), 275–298. arXiv:https://doi.org/10.1177/0013916514539755 doi:10.1177/0013916514539755

[77] Elizabeth L. Murnane, Saeed Abdullah, Mark Matthews, Tanzeem Choudhury, and Geri Gay. 2015. Social (media) jet lag: how usage of social technology can modulate and reflect circadian rhythms. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Osaka Japan, 843–854. doi:10.1145/2750858.2807522

[78] National Eye Insititute. [n. d.]. Diabetic Retinopathy | National Eye Institute. https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy

[79] Joel T. Nigg. 2017. Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry* 58, 4 (April 2017), 361–383. doi:10.1111/jcpp.12675

[80] Fredrik A. Nilsen, Henning Bang, and Espen Røysamb. 2024. Personality traits and self-control: The moderating role of neuroticism. *PLOS ONE* 19, 8 (Aug. 2024), e0307871. doi:10.1371/journal.pone.0307871

[81] Fabian Okeke, Michael Sobolev, Nicola Dell, and Deborah Estrin. 2018. Good vibrations. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, Lynne Bailie and Nuria Oliver (Eds.). ACM, New York, NY, USA, 1–12. doi:10.1145/3229434.3229463

[82] Fabian Okeke, Michael Sobolev, Nicola Dell, and Deborah Estrin. 2018. Good Vibrations: Can a Digital Nudge Reduce Digital Overload?. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI ’18)*.

Association for Computing Machinery, New York, NY, USA. doi:10.1145/3229434.3229463

[83] Fabian Okeke, Michael Sobolev, and Deborah Estrin. 2018. Towards A Framework for Mobile Behavior Change Research. In *Proceedings of the Technology, Mind, and Society*. ACM, New York, NY, USA, 1–6. doi:10.1145/3183654.3183706

[84] Adiba Orzikulova, Hyunsung Cho, Hye-Young Chung, Hwajung Hong, Uichin Lee, and Sung-Ju Lee. 2023. FinerMe: Examining App-level and Feature-level Interventions to Regulate Mobile Social Media Use. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Sept. 2023), 1–30. doi:10.1145/3610065

[85] Adiba Orzikulova, Han Xiao, Zhipeng Li, Yukang Yan, Yuntao Wang, Yuanchun Shi, Marzyeh Ghassemi, Sung-Ju Lee, Anind K Dey, and Xuhai Xu. 2024. Time2Stop: Adaptive and Explainable Human-AI Loop for Smartphone Overuse Intervention. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–20. doi:10.1145/3613904.3642747

[86] Joonyoung Park, Hyunsoo Lee, Sangkeun Park, Kyong-Mee Chung, and Uichin Lee. 2021. GoldenTime: Exploring System-Driven Timeboxing and Micro-Financial Incentives for Self-Regulated Phone Use. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, Takeo Igarashi, Pernille Bjørn, and Steven Drucker (Eds.). ACM, New York, NY, USA, 1–17. doi:10.1145/3411764.3445489

[87] Martin Pielot, Karen Church, and Rodrigo de Oliveira. 2014. An in-situ study of mobile phone notifications. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services (MobileHCI '14)*. 233–242. doi:10.1145/2628363.2628364

[88] Andrew K. Przybylski, Kou Murayama, Cody R. DeHaan, and Valerie Gladwell. 2013. Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior* 29, 4 (July 2013), 1841–1848. doi:10.1016/j.chb.2013.02.014

[89] Aditya Kumar Purohit, Torben Jan Barev, Sofia Schöbel, Andreas Janson, and Adrian Holzer. 2023. Designing for Digital Wellbeing on a Smartphone: Co-creation of Digital Nudges to Mitigate Instagram Overuse. In *Proceedings of the 56th Hawaii International Conference on System Sciences*. HICSS, Hawaii, 4087–4096. doi:10.24251/HICSS.2023.499

[90] Aditya Kumar Purohit, Kristoffer Bergram, Louis Barclay, Valéry Bezençon, and Adrian Holzer. 2023. Starving the Newsfeed for Social Media Detox: Effects of Strict and Self-Regulated Facebook Newsfeed Diets. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 196, 16 pages. doi:10.1145/3544548.3581187

[91] Aditya Kumar Purohit and Adrian Holzer. 2019. Functional Digital Nudges. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, Stephen Brewster, Geraldine Fitzpatrick, Anna Cox, and Vassilis Kostakos (Eds.). ACM, New York, NY, USA, 1–6. doi:10.1145/3290607.3312876

[92] Aditya Kumar Purohit and Adrian Holzer. 2019. Functional Digital Nudges: Identifying Optimal Timing for Effective Behavior Change. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3290607.3312876

[93] Aditya K. Purohit and Adrian Holzer. 2021. Unhooked by Design: Scrolling Mindfully on Social Media by Automating Digital Nudges. In *Proceedings of the 27th Americas Conference on Information Systems, AMCIS 2021*. AIS, Virtual Conference, 1–10. https://aisel.aisnet.org/amcis2021/sig_hci/sig_hci/7

[94] Stephen A. Rains. 2013. The Nature of Psychological Reactance Revisited: A Meta-Analytic Review. *Human Communication Research* 39, 1 (2013), 47–73. doi:10.1111/j.1468-2958.2012.01443.x

[95] Tenko Raykov and George A. Marcoulides. 2009. *Introduction to Psychometric Theory* (0 ed.). Routledge. doi:10.4324/9780203841624

[96] Jan Ole Rixen, Luca-Maxim Meinhardt, Michael Glöckler, Marius-Lukas Ziegenbein, Anna Schlothauer, Mark Colley, Enrico Rukzio, and Jan Guggenheimer. 2023. The Loop and Reasons to Break It: Investigating Infinite Scrolling Behaviour in Social Media Applications and Reasons to Stop. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 1228 (sep 2023), 22 pages. doi:10.1145/3604275

[97] Alberto Monge Roffarello and Luigi de Russis. 2019. The Race Towards Digital Wellbeing: Issues and Opportunities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300616

[98] Alberto Monge Roffarello and Luigi De Russis. 2021. Understanding, Discovering, and Mitigating Habitual Smartphone Use in Young Adults. *ACM Trans. Interact. Intell. Syst.* 11, 2, Article 13 (jul 2021), 34 pages. doi:10.1145/3447991

[99] Alberto Monge Roffarello and Luigi de Russis. 2022. Towards Understanding the Dark Patterns That Steal Our Attention. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3491101.3519829

[100] Alberto Monge Roffarello and Luigi De Russis. 2023. Achieving Digital Wellbeing Through Digital Self-Control Tools: A Systematic Review and Meta-Analysis. *ACM Trans. Comput.-Hum. Interact.* 30, 4, Article 53 (sep 2023), 66 pages. doi:10.1145/3571810

[101] Nicolas Ruiz, Gabriela Molina León, and Hendrik Heuer. 2024. Design Frictions on Social Media: Balancing Reduced Mindless Scrolling and User Satisfaction. In *Proceedings of Mensch und Computer 2024*. ACM, Karlsruhe Germany, 442–447. doi:10.1145/3670653.3677495

[102] Diane M. Samdahl. 1991. Measuring Leisure: Categorical or Interval? *Journal of Leisure Research* 23, 1 (1991), 87–93. doi:10.1080/00222216.1991.11969845

[103] Daniel J. Schad, Michael Betancourt, and Shravan Vasishth. 2021. Toward a principled Bayesian workflow in cognitive science. *Psychological Methods* 26, 1 (Feb. 2021), 103–126. doi:10.1037/met0000275

[104] Azmeh Shahid, Kate Wilkinson, Shai Marcu, and Colin M. Shapiro (Eds.). 2012. *STOP, THAT and One Hundred Other Sleep Scales*. Springer New York, New York, NY. doi:10.1007/978-1-4419-9893-4

[105] Holly Shannon, Katie Bush, Paul J Villeneuve, Kim Gc Hellemans, and Synthia Guimond. 2022. Problematic Social Media Use in Adolescents and Young Adults: Systematic Review and Meta-analysis. *JMIR Mental Health* 9, 4 (April 2022), e33450. doi:10.2196/33450

[106] Zeynep Simsir-Gokalp and Muhammet Ibrahim Akyurek. 2024. Self-control and Problematic Social Media Use: A Meta-Analysis. *Journal of Education in Science, Environment and Health* (July 2024), 199–215. doi:10.55549/jeseh.722

[107] Michael Sobolev, Rachel Vitale, Hongyi Wen, James Kizer, Robert Leeman, J. P. Pollak, Amit Baumel, Nehal P. Vadhan, Deborah Estrin, and Frederick Muench. 2021. The Digital Marshmallow Test (DMT) Diagnostic and Monitoring Mobile Health App for Impulsive Behavior: Development and Validation Study. *JMIR mHealth and uHealth* 9, 1 (2021), e25018. doi:10.2196/25018

[108] Eugene Nikolaievich Sokolov. 1960. Neuronal models and the orienting reflex. *The central nervous system and behaviour* (1960).

[109] Marcello Spinella. 2007. Normative data and a short form of the Barratt Impulsiveness Scale. *International Journal of Neuroscience* 117, 3 (January 2007), 359–368. doi:10.1080/00207450600588881

[110] Clemens Stachl, Quay Au, Ramona Schoedel, Samuel D. Gosling, Gabriella M. Harari, Daniel Buschek, Sarah Theres Völkel, Tobias Schuwerk, Michelle Oldemeier, Theresa Ullmann, Heinrich Hussmann, Bernd Bischl, and Markus Bühner. 2020. Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences* 117, 30 (July 2020), 17680–17687. doi:10.1073/pnas.1920484117

[111] statcounter. 2023. *Mobile Android Version Market Share United States Of America*. Retrieved May 10, 2025 from https://gs.statcounter.com/android-version-market-share/mobile/united-states-of-america/#monthly-202208-202310-bar

[112] Statista. 2023. *Social network usage by brand in the U.S. as of December 2023*. https://www.statista.com/forecasts/997135/social-network-usage-by-brand-in-the-us

[113] June P. Tangney, Roy F. Baumeister, and Angie Luzio Boone. 2004. High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality* 72, 2 (April 2004), 271–324. doi:10.1111/j.0022-3506.2004.00263.x

[114] Nađa Terzimehić and Sarah Aragon-Hahner. 2022. I Wish I Had: Desired Real-World Activities Instead of Regretful Smartphone Use. In *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia*, Tanja Döring, Susanne Boll, Ashley Colley, Augusto Esteves, and João Guerreiro (Eds.). ACM, New York, NY, USA, 47–52. doi:10.1145/3568444.3568465

[115] Kelly J Thomas Craig, Laura C Morgan, Ching-Hua Chen, Susan Michie, Nicole Fusco, Jane L Snowdon, Elisabeth Scheufele, Thomas Gagliardi, and Stewart Sill. 2021. Systematic review of context-aware digital behavior change interventions to improve health. *Translational behavioral medicine* 11, 5 (2021), 1037–1048. doi:10.1093/tbm/ibaa099

[116] TikTok. 2025. Screen time. https://support.tiktok.com/en/account-and-privacy/account-information/screen-time. Accessed: 2025-03-14.

[117] Niels van Berkel, Jorge Goncalves, Lauri Lovén, Denzil Ferreira, Simo Hosio, and Vassilis Kostakos. 2019. Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies* 125 (2019), 118–128. doi:10.1016/j.ijhcs.2018.12.002

[118] Rens Van De Schoot, Sonja D. Winter, Oisín Ryan, Mariëlle Zondervan-Zwijnenburg, and Sarah Depaoli. 2017. A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods* 22, 2 (June 2017), 217–239. doi:10.1037/met0000100

[119] Mariek M. P. Vanden Abeele. 2021. Digital Wellbeing as a Dynamic Construct. *Communication Theory* 31, 4 (2021), 932–955. doi:10.1093/ct/qtaa024

[120] Viswanath Venkatesh and Fred D. Davis. 2000. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science* 46, 2 (2000), 186–204. doi:10.1287/mnsc.46.2.186.11926

[121] Clara Virós-Martín, Mireia Montaña-Blasco, and Mònika Jiménez-Morales. 2024. Can't stop scrolling! Adolescents' patterns of TikTok use and digital well-being self-perception. *Humanities and Social Sciences Communications* 11, 1 (Oct. 2024), 1444. doi:10.1057/s41599-024-03984-5

[122] Philip Weber, Philip Engelbutzeder, and Thomas Ludwig. 2020. "Always on the Table": Revealing Smartphone Usages in Everyday Eating Out Situations. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (Tallinn, Estonia) *(NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 2, 13 pages. doi:10.1145/3419249.3420150

[123] Elisa Wegmann, Ursula Oberst, Benjamin Stodt, and Matthias Brand. 2017. Online-specific fear of missing out and Internet-use expectancies contribute to symptoms of Internet-communication disorder. *Addictive Behaviors Reports* 5 (June 2017), 33–42. doi:10.1016/j.abrep.2017.04.001

[124] Hongyi Wen, Michael Sobolev, Rachel Vitale, James Kizer, J P Pollak, Frederick Muench, and Deborah Estrin. 2021. mPulse Mobile Sensing Model for Passive Detection of Impulsive Behavior: Exploratory Prediction Study. *JMIR Mental Health* 8, 1 (Jan. 2021), e25019. doi:10.2196/25019

[125] William Revelle. 2025. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.5.6. https://CRAN.R-project.org/package=psych

[126] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-Computer Interaction. *Interactions* 23, 3 (apr 2016), 38–44. doi:10.1145/2907069

[127] Ruolan Wu, Chun Yu, Xiaole Pan, Yujia Liu, Ningning Zhang, Yue Fu, Yuhan Wang, Zhi Zheng, Li Chen, Qiaolei Jiang, Xuhai Xu, and Yuanchun Shi. 2024. MindShift: Leveraging Large Language Models for Mental-States-Based Problematic Smartphone Use Intervention. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–24. doi:10.1145/3613904.3642790

[128] Qiang Xu, Jeffrey Erman, Alexandre Gerber, Zhuoqing Mao, Jeffrey Pang, and Shobha Venkataraman. 2011. Identifying diverse usage behaviors of smartphone apps. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, Berlin Germany, 329–344. doi:10.1145/2068816.2068847

[129] Andras N. Zsido, Szidalisz A. Teleki, Krisztina Csokasi, Sandor Rozsa, and Szabolcs A. Bandi. 2020. Development of the short version of the Spielberger State–Trait Anxiety Inventory. *Psychiatry Research* 291 (September 2020), 113223. doi:10.1016/j.psychres.2020.113223

A Descriptive Data of the User Study

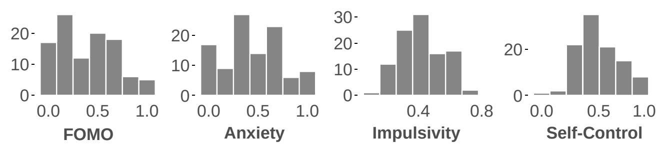


Fig. 7. Distribution of the between-factors variables of the participants (normalized)

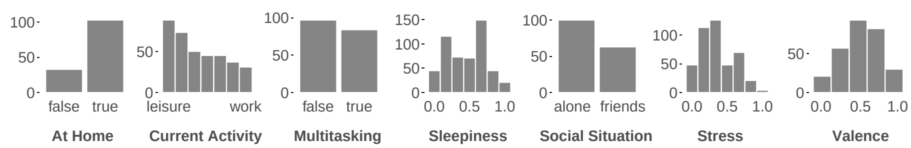


Fig. 8. Distribution of the within-factors variables of the participants (normalized)

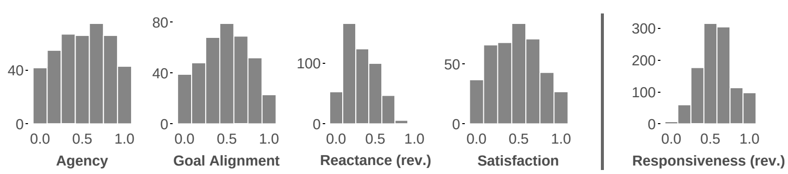


Fig. 9. Distribution of the six measures (normalized) that were used to assess objective and subjective intervention effectiveness. reactance and responsiveness are reversed (see Appendix B)

Table 2. Descriptive statistics for contextual factors, individual traits, and intervention effectiveness. All values are normalized, except for responsiveness.

Contextual Factor	**Min**	**Max**	**Mean**	**Median**	**SD**	**Distribution**
Sleepiness	0.00	1.00	0.46	0.50	0.28	
Current Activity	0.00	1.00	0.27	0.17	0.31	
Valence	0.00	1.00	0.55	0.50	0.23	
At Home						true (89.72%), false (10.28%)
Multitasking						true (40.42%), false (59.58%)
Social Situation						alone (72.80%), friends (27.20%)
Individual Traits						
Self-Control	0.06	0.92	0.50	0.48	0.19	
Anxiety	0.00	1.00	0.43	0.47	0.29	
FOMO	0.00	1.00	0.41	0.35	0.28	
Impulsivity	0.13	0.76	0.41	0.40	0.14	
Intervention Effectiveness						
Subjective Effectiveness						
Baseline	0.00	0.98	0.48	0.51	0.24	
Visual Intervention	0.00	0.92	0.40	0.42	0.22	
Haptic Intervention	0.00	0.97	0.43	0.43	0.25	
Objective Effectiveness						
Baseline	0.00	1.00	0.65	0.71	0.30	
Visual Intervention	0.03	1.00	0.51	0.44	0.26	
Haptic Intervention	0.00	1.00	0.55	0.54	0.25	
Individual measures						
Responsiveness – Baseline	0s	23min 10s	2min 2s	7s	4min 40s	
Responsiveness – Visual	0s	18min 10s	1min 42s	56s	2min 28s	
Responsiveness – Haptic	0s	23min 13s	1min 31s	28s	2min 52s	
Reactance (rev.) – Baseline	0.10	1.00	0.72	0.75	0.20	
Reactance (rev.) – Visual	0.00	1.00	0.74	0.75	0.19	
Reactance (rev.) – Haptic	0.20	1.00	0.76	0.75	0.20	
Goal Alignment – Baseline	0.00	1.00	0.51	0.50	0.29	
Goal Alignment – Visual	0.00	1.00	0.43	0.50	0.27	
Goal Alignment – Haptic	0.00	1.00	0.47	0.50	0.30	
Satisfaction – Baseline	0.00	1.00	0.51	0.50	0.29	
Satisfaction – Visual	0.00	1.00	0.40	0.42	0.28	
Satisfaction – Haptic	0.00	1.00	0.44	0.50	0.31	
Agency – Baseline	0.00	1.00	0.59	0.67	0.31	
Agency – Visual	0.00	1.00	0.45	0.50	0.31	
Agency – Haptic	0.00	1.00	0.53	0.50	0.33	
Usefulness – Baseline	0.00	1.00	0.49	0.50	0.33	
Usefulness – Visual	0.00	1.00	0.41	0.33	0.30	
Usefulness – Haptic	0.00	1.00	0.44	0.50	0.33	
Intervention Distribution						
Baseline (34.23%), Visual Intervention (33.69%), Haptic Intervention (32.07%)						
App Distribution						
TikTok (41.27%), Instagram (22.91%), Facebook (17.91%), YouTube Shorts (17.91%)						

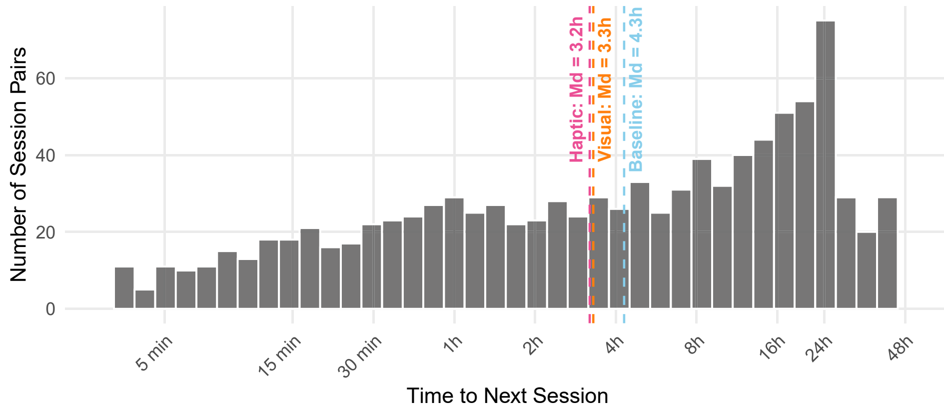


Fig. 10. Distribution of break times between consecutive (intervention-triggering) scrolling sessions. Each dotted line represents the median break time per intervention type before the next session. Only sessions exceeding the 15-minute threshold (triggering an intervention) are included.

B Question Items Used in the User Study

Table 3. Question items used in the user study to determine **context**

Measurement	Question Item	Answer Items	Ref.
Current Activity	What is your current activity?	7-point Likert-type scale from -3 ("definitely leisure") to 3 ("definitely not leisure")	[102]
Valence	How do you feel?	five images of manikin showing different valence levels	[8]
Sleepiness	What is your level of sleepiness?	9-point Likert-type scale from 1 ("extremely alert") to 9 ("extremely sleepy")	[104]
Social Situation	Which one of these best describes people around you?	"alone", "with friends/colleagues/family members"	[1]
Stress	What number best describes your level of stress right now?	11-point Likert-type scale, ranging from 0 ("no stress") to 10 ("worst stress possible")	[47]
Multitasking	Did you do anything else besides being on [app name]?	"yes", "no"	[72]
At Home	Are you currently at home?	"yes", "no"	[72]

Table 4. Measures and questionnaire items to determine subjective and objective **intervention effectiveness**

Measurement	Question Item	Answer Items	Ref.	Calculation
Objective Effectiveness				
Responsiveness			[72]	$log(1 + Responsiveness)$, normalized (0-1), and reversed
Subjective Effectiveness (weighted measures)				
Reactance	I want to be in control, not the intervention. I like to act independently from the intervention. I don't want the intervention to tell me what to do. I don't let the intervention impose its will on me. I alone determine what to do, not the intervention.	5-point Likert scale from "strongly disagree", to "strongly agree"	[31]	normalized (0-1) and reversed
Agency	For this intervention, how much did you feel out of or in control?	7-point Likert-type scale, ranging from 1 ("very out of control") to 7 ("very in control")	[67]	normalized (0-1)
Satisfaction	For this intervention, how much did you feel dissatisfied or satisfied?	7-point Likert-type scale, ranging from 1 ("very dissatisfied") to 7 ("very satisfied")	[67]	normalized (0-1)
Goal Alignment	For this intervention, how much did it conflict with or support your personal goals?	7-point Likert-type scale, ranging from 1 ("very in conflict") to 7 ("very supported")	[67]	normalized (0-1)
Usefulness	I find the intervention to be useful in my current situation.	7-point Likert-type scale, ranging from 1 ("strongly disagree") to 7 ("strongly agree")	[120]	normalized (0-1)

Table 5. Question items used in the user study to determine **individual traits**

Measurement	Question Item	Answer Items	Ref.
Impulsivity	I act on impulse. [inverted] I act on the spur of the moment. I do things without thinking. I say things without thinking. I buy things on impulse. I plan for job security. [inverted] I plan for the future. [inverted] I save regularly. [inverted] I plan tasks carefully. [inverted] I am a careful thinker. [inverted] I am restless at lectures or talks. I squirm at plays or lectures. I concentrate easily. [inverted] I don't pay attention. Easily bored solving thought problems.	4-point Likert scale, ranging from 1 ("rarely/never") to 4 ("almost always")	[109]
Anxiety	I feel that difficulties are piling up so that I can't overcome them. I worry too much over something that really doesn't matter. Some unimportant thought runs through my mind and bothers me. I take disappointments so keenly that I can't put them out of my mind. I get in a state of tension or turmoil as I think over my recent concerns and interests.	4-point Likert scale, ranging from 0 ("almost never") to 3 ("almost always")	[129]
FOMO	I fear others have more rewarding experiences than me. I fear my friends have more rewarding experiences than me. I get worried when I find out my friends are having fun without me. I get anxious when I don't know what my friends are up to. When I miss out on a planned get-together it bothers me.	5-point Likert scale, ranging from 1 ("totally disagree") to 5 ("totally agree")	[123]
Self-Control	I am good at resisting temptation. I have a hard time breaking bad habits. [inverted] I am lazy. [inverted] I say inappropriate things. [inverted] I do certain things that are bad for me, if they are fun. [inverted] I refuse things that are bad for me. I wish I had more self-discipline. [inverted] People would say that I have iron self-discipline. Pleasure and fun sometimes keep me from getting work done. [inverted] I have trouble concentrating. [inverted] I am able to work effectively toward long-term goals. Sometimes I can't stop myself from doing something, even if I know it is wrong. [inverted] I often act without thinking through all the alternatives. [inverted]	5-point Likert scale, ranging from 1 ("not at all") to 5 ("very much")	[113]

C Mixed Power Simulation

To determine the required sample size for sufficient statistical power, we conducted a simulation-based power analysis using pilot data from the first batch of 30 participants, following the approach by Kumle et al. [60]. However, simulating the full model with all contextual variables was computationally intensive. To address this, we first fitted the full model on the data of the 30 participants and identified factors with $|t| > 2$ (see Table 6a). We then limited the power simulation to those factors where $|t| > 2$ [60]: anxiety, impulsivity, and FOMO. The simulation indicated that a sample size of 90 participants would be sufficient to achieve 80% power [19] (see Table 6b).

Table 6. Feature importance (a) and corresponding power estimates (b) for included predictors and their interactions

(a) t-values of the full model's interactions on the pilot data (N = 30)

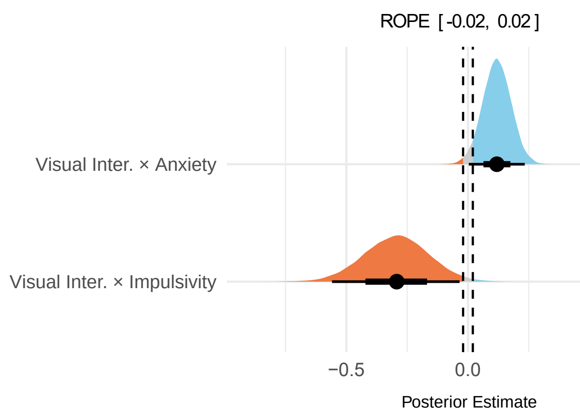
Interaction Effect	t-value
Visual Inter. × Valence	-0.32
Haptic Inter. × Valence	0.11
Visual Inter. × Impulsivity	**-2.69**
Haptic Inter. × Impulsivity	0.22
Visual Inter. × Sleepiness	-0.70
Haptic Inter. × Sleepiness	-0.59
Visual Inter. × FOMO	1.20
Haptic Inter. × FOMO	**2.22**
Visual Inter. × Stress	0.75
Haptic Inter. × Stress	-1.21
Visual Inter. × Multitasking	-0.37
Haptic Inter. × Multitasking	0.65
Visual Inter. × At Home	0.05
Haptic Inter. × At Home	-1.19
Visual Inter. × Self-Control	0.47
Haptic Inter. × Self-Control	-0.02
Visual Inter. × Anxiety	**3.51**
Haptic Inter. × Anxiety	-1.15
Visual Inter. × Current Activity	-0.88
Haptic Inter. × Current Activity	-0.63
Visual Inter. × Social Situation	-1.63
Haptic Inter. × Social Situation	-0.08

(b) Power Simulation Estimates for Interactions (based on 1000 simulation runs)

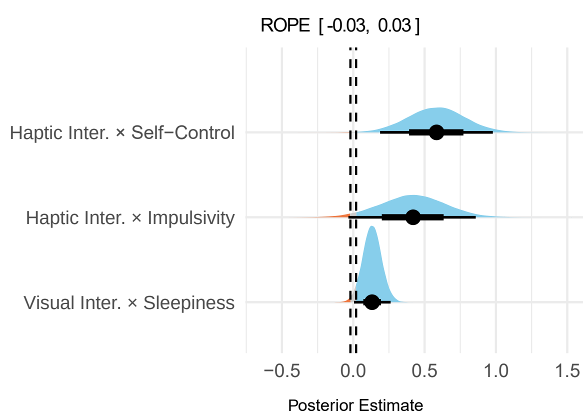
Interaction Effect	Power per Sample Size				
	30	60	90	120	150
Visual Interv. × Anxiety	0.43	**0.92**	**0.99**	**0.99**	**1.00**
Haptic Interv. × Anxiety	0.12	0.31	0.45	0.55	0.67
Visual Interv. × Impulsivity	**0.92**	**1.00**	**1.00**	**1.00**	**1.00**
Haptic Interv. × Impulsivity	0.04	0.06	0.05	0.04	0.06
Visual Interv. × FOMO	0.31	0.74	**0.89**	**0.96**	**0.99**
Haptic Interv. × FOMO	0.28	0.73	**0.88**	**0.95**	**0.98**

D Results of the Bayesian Model

We estimated two Bayesian mixed-effects models to examine how intervention type and individual/contextual factors affect objective and subjective intervention effectiveness. This section presents the complete model results in Table 8 for the objective effectiveness and Table 7 for the subjective effectiveness of the interventions. Figure 11 shows the existing interaction effects (pd > 95%) between intervention type and individual traits on both subjective and objective intervention effectiveness. The vertical dashed lines indicate the ROPE where effects are considered practically negligible. When the percentage in ROPE is below 2.5% an effect can be considered as probably significant [69].



(a) Interaction effects on **subjective effectiveness**



(b) Interaction effects on **objective effectiveness**

Fig. 11. Interaction Effects between the intervention type and the individual traits where the $pd > 95\%$

Table 7. Results of the Bayesian mixed-effects model for **subjective effectiveness**. Effects with a *pd* > 95% are highlighted in bold.

Subjective Effectiveness					
Parameter	Median	95% CrI	pd	% in ROPE	ESS
Intercept	*0.12*	*[-0.16, 0.40]*	*81.18%*	*9.56%*	*18,506*
Visual Inter.	-0.06	[-0.29, 0.16]	71.12%	14.91%	18,619
Haptic Inter.	-0.03	[-0.24, 0.18]	61.38%	18.08%	19,4831
Valence	**0.38**	**[0.29, 0.46]**	**100%**	**0%**	**21,333**
Sleepiness	**0.08**	**[0.03, 0.13]**	**99.95%**	**0%**	**28,091**
Stress	0.02	[-0.06, 0.10]	67.97%	42.43%	26,071
Multitasking [Yes]	0.00	[-0.03, 0.03]	52.98%	95.78%	26420
At Home [True]	0.00	[-0.06, 0.04]	62.44%	63.86%	28,445
Current Activity	0.01	[-0.04, 0.06]	69.70%	62.03%	26,789
Social Situation [Friends]	0.00	[-0.04, 0.03]	52.52%	88.33%	23,941
Impulsivity	**0.33**	**[0.01, 0.67]**	**97.77%**	**0.65%**	**15,797**
FOMO	-0.11	[-0.27, 0.05]	92.03%	9.41%	24,501
Self-Control	0.00	[-0.28, 0.28]	50.25%	14.51%	23,342
Anxiety	0.06	[-0.08, 0.21]	80.57%	19.19%	27,244
Visual Inter. × Valence	0.04	[-0.05, 0.13]	82.52%	28.08%	25,621
Haptic Inter. × Valence	0.02	[-0.07, 0.11]	68.30%	37.55%	25,146
Visual Inter. × Sleepiness	-0.03	[-0.10, 0.04]	81.62%	38.95%	27,279
Haptic Inter. × Sleepiness	0.00	[-0.05, 0.07]	61.76%	54.44%	31,846
Visual Inter. × Stress	-0.02	[-0.12, 0.08]	67.34%	35.44%	23,862
Haptic Inter. × Stress	-0.04	[-0.13, 0.05]	82.66%	28.05%	29,071
Visual Inter. × Multitasking [Yes]	0.02	[-0.02, 0.05]	80.69%	66.75%	29,839
Haptic Inter. × Multitasking [Yes]	0.00	[-0.03, 0.04]	63.69%	79.53%	33,202
Visual Inter. × At Home [True]	0.01	[-0.06, 0.07]	62.09%	54.61%	30,071
Haptic Inter. × At Home [True]	0.02	[-0.04, 0.08]	74.75%	49.97%	33,719
Visual Inter. × Current Activity	0.02	[-0.04, 0.08]	72.12%	49.22%	27,196
Haptic Inter. × Current Activity	0.00	[-0.07, 0.06]	58.94%	55.90%	27,197
Visual Inter. × Social Situation [Friends]	-0.01	[-0.05, 0.03]	73.39%	68.89%	26,516
Haptic Inter. × Social Situation [Friends]	0.00	[-0.04, 0.05]	58.23%	73.80%	25,636
Visual Inter. × Impulsivity	**-0.29**	**[-0.56, -0.03]**	**98.69%**	**0%**	**19,785**
Haptic Inter. × Impulsivity	-0.17	[-0.41, 0.06]	93.03%	5.79%	23,975
Visual Inter. × FOMO	0.00	[-0.12, 0.13]	53.83%	31.42%	22,350
Haptic Inter. × FOMO	0.06	[-0.05, 0.17]	84.44%	21.43%	27,067
Visual Inter. × Self-Control	0.08	[-0.15, 0.30]	75.25%	14.23%	18,754
Haptic Inter. × Self-Control	0.05	[-0.15, 0.26]	69.58%	17.31%	21,192
Visual Inter. × Anxiety	**0.12**	**[0.00, 0.23]**	**97.81%**	**2.82%**	**24,627**
Haptic Inter. × Anxiety	0.00	[-0.11, 0.11]	50.89%	35.80%	30,901
Conditional R^2 = 0.80 (95% CrI [0.79, 0.81])					
Marginal R^2 = 0.28, (95% CrI [0.21, 0.35]					

Table 8. Results of the Bayesian mixed-effects model for **objective effectiveness**. Effects with a *pd* > 95% are highlighted in bold.

Objective Effectiveness					
Parameter	Median	95% CrI	pd	% in ROPE	ESS
Intercept	*0.83*	*[0.48, 1.19]*	*100%*	*0.00%*	*11,143*
Visual Inter.	**0.36**	**[0.01, 0.72]**	**97.67%**	**0.83%**	**17,219**
Haptic Inter.	0.03	[-0.31, 0.37]	55.77%	13.04%	18,762
Valence	-0.07	[-0.21, 0.06]	86.42%	18.99%	15,431
Sleepiness	-0.08	[-0.18, 0.02]	94.89%	12.61%	19,957
Stress	-0.02	[-0.16, 0.12]	61.79%	30.37%	17,187
Multitasking [Yes]	0.03	[-0.03, 0.09]	80.52%	50.68%	21,658
At Home [True]	0.01	[-0.08, 0.09]	55.09%	49.31%	12,706
Current Activity	0.04	[-0.06, 0.13]	79.05%	34.84%	17,694
Social Situation [Friends]	-0.04	[-0.11, 0.02]	90.70%	31.62%	21,886
Impulsivity	0.06	[-0.34, 0.46]	61.86%	11.21%	12,324
FOMO	0.00	[-0.21, 0.20]	51.79%	22.57%	13,987
Self-Control	-0.19	[-0.52, 0.16]	85.20%	7.49%	12,573
Anxiety	-0.06	[-0.25, 0.13]	73.47%	20.13%	17,986
Visual Inter. × Valence	0.07	[-0.11, 0.24]	78.06%	18.93%	17,218
Haptic Inter. × Valence	0.12	[-0.05, 0.29]	90.88%	11.15%	14,751
Visual Inter. × Sleepiness	**0.13**	**[0.01, 0.26]**	**98.02%**	**2.83%**	**19,762**
Haptic Inter. × Sleepiness	0.08	[-0.04, 0.21]	89.93%	16.19%	21,316
Visual Inter. × Stress	0.07	[-0.12, 0.25]	75.94%	19.47%	17,595
Haptic Inter. × Stress	0.00	[-0.19, 0.17]	53.60%	25.19%	16,668
Visual Inter. × Multitasking [Yes]	0.02	[-0.06, 0.10]	70.20%	48.98%	21,934
Haptic Inter. × Multitasking [Yes]	0.03	[-0.05, 0.11]	77.85%	41.59%	21,536
Visual Inter. × At Home [True]	0.03	[-0.09, 0.14]	66.31%	34.76%	18,697
Haptic Inter. × At Home [True]	0.05	[-0.07, 0.17]	81.37%	25.68%	14,291
Visual Inter. × FOMO	0.02	[-0.22, 0.25]	54.97%	19.02%	14,414
Haptic Inter. × FOMO	0.12	[-0.10, 0.35]	85.88%	11.28%	15,072
Visual Inter. × Current Activity	-0.04	[-0.16, 0.09]	72.55%	30.20%	18,940
Haptic Inter. × Current Activity	0.00	[-0.12, 0.13]	52.12%	35.02%	20,771
Visual Inter. × Social Situation [Friends]	0.02	[-0.06, 0.10]	69.94%	46.03%	23399
Haptic Inter. × Social Situation [Friends]	0.02	[-0.06, 0.10]	67.55%	46.07%	23216
Visual inter. × Impulsivity	-0.04	[-0.52, 0.43]	56.00%	9.82%	13,301
Haptic Inter. × Impulsivity	**0.42**	**[-0.03, 0.86]**	**96.61%**	**1.91%**	**13,337**
Visual Inter. × Self-Control	0.30	[-0.11, 0.71]	92.58%	3.92%	13,539
Haptic Inter. × Self-Control	**0.58**	**[0.18, 0.97]**	**99.74%**	**0.00%**	**11,703**
Visual Inter. × Anxiety	0.03	[-0.19, 0.25]	60.27%	20.16%	17,304
Haptic inter. × Anxiety	0.02	[-0.19, 0.24]	58.59%	20.26%	18,816
Conditional $R^2 = 0.45$ (95% CrI $[0.40, 0.50]$)					
Marginal $R^2 = 0.12$, (95% CrI $[0.09, 0.17]$)					

E Sensitivity Analysis

To assess the robustness of the moderation effects reported in the main analysis, we re-calculated both Bayesian mixed-effects models under two alternative participant filtering conditions. The primary analysis retained only participants who experienced all three intervention types ($N = 104$). In the first sensitivity condition, we relaxed this criterion to include all participants who completed at least three interventions in total, regardless of whether all three types were represented ($N = 150$). In the second condition, we included all 214 participants who completed the 7-day study, with no filtering based on intervention exposure. In both sensitivity conditions, the same preprocessing pipeline (outlier removal, normalization, model specification, and priors) was applied. Table 9 shows the results for all interaction effects that were possibly existing ($pd > 95\%$) in at least one of the three filtering conditions. The results show that the anxiety and self-control interaction effects were robust across all three conditions, with significance strengthening as more participants were included. The impulsivity interactions attenuated but remained directionally consistent, which is expected given the reduced within-person balance in the larger samples. Several additional effects emerged only under relaxed filtering (e.g., *Visual Inter. × Self-Control*, *Haptic Inter. × FoMo*) and should be interpreted cautiously. Overall, these results support the primary filtering decision while demonstrating that the core findings are not artifacts of this particular filtering threshold.

Table 9. Sensitivity analysis across three participant filtering conditions. Effects with $pd > 95\%$ are bold.

Interaction Effect	Filtering	Median	95% CrI	pd	% in ROPE
Subjective Effectiveness					
Visual Inter. × Impulsivity	$N = 104$ (primary)	**−0.29**	**[−0.56, −0.03]**	**98.69%**	**0.00%**
	$N = 150$ (intermediate)	−0.11	[−0.33, 0.12]	82.91%	10.66%
	$N = 214$ (relaxed)	−0.07	[−0.28, 0.14]	74.52%	14.67%
Visual Inter. × Anxiety	$N = 104$ (primary)	**0.12**	**[0.00, 0.23]**	**97.81%**	**2.82%**
	$N = 150$ (intermediate)	**0.15**	**[0.04, 0.26]**	**99.61%**	**0.00%**
	$N = 214$ (relaxed)	**0.15**	**[0.05, 0.25]**	**99.74%**	**0.00%**
Visual Inter. × Self-Control	$N = 104$ (primary)	0.08	[−0.15, 0.30]	75.25%	14.23%
	$N = 150$ (intermediate)	0.16	[−0.04, 0.36]	93.73%	6.08%
	$N = 214$ (relaxed)	**0.17**	**[−0.02, 0.36]**	**96.33%**	**3.69%**
Objective Effectiveness					
Haptic Inter. × Self-Control	$N = 104$ (primary)	**0.58**	**[0.18, 0.97]**	**99.74%**	**0.00%**
	$N = 150$ (intermediate)	**0.61**	**[0.26, 0.96]**	**99.98%**	**0.00%**
	$N = 214$ (relaxed)	**0.60**	**[0.26, 0.93]**	**99.97%**	**0.00%**
Haptic Inter. × Impulsivity	$N = 104$ (primary)	**0.42**	**[−0.03, 0.86]**	**96.61%**	**1.91%**
	$N = 150$ (intermediate)	0.30	[−0.09, 0.69]	93.77%	3.49%
	$N = 214$ (relaxed)	0.28	[−0.10, 0.65]	92.88%	3.98%
Visual Inter. × Sleepiness	$N = 104$ (primary)	**0.13**	**[0.01, 0.26]**	**98.02%**	**2.83%**
	$N = 150$ (intermediate)	**0.11**	**[−0.01, 0.23]**	**96.72%**	**5.96%**
	$N = 214$ (relaxed)	0.09	[−0.04, 0.21]	91.65%	14.68%
Visual Inter. × Self-Control	$N = 104$ (primary)	0.30	[−0.11, 0.71]	92.58%	3.92%
	$N = 150$ (intermediate)	**0.33**	**[0.00, 0.67]**	**96.94%**	**1.92%**
	$N = 214$ (relaxed)	**0.34**	**[0.00, 0.67]**	**97.49%**	**1.23%**
Haptic Inter. × FoMo	$N = 104$ (primary)	0.12	[−0.10, 0.35]	85.88%	11.28%
	$N = 150$ (intermediate)	**0.20**	**[0.00, 0.40]**	**97.32%**	**2.41%**
	$N = 214$ (relaxed)	**0.23**	**[0.04, 0.43]**	**99.04%**	**0.00%**