

InteractOut: Leveraging Interaction Proxies as Input Manipulation Strategies for Reducing Smartphone Overuse

Tao Lu

University of Michigan
Ann Arbor, MI, USA
luttul@umich.edu

Hongxiao Zheng

University of Michigan
Ann Arbor, MI, USA
waleyz@umich.edu

Tianying Zhang

University of Michigan
Ann Arbor, MI, USA
danniez@umich.edu

Xuhai Xu

Massachusetts Institute of Technology
Cambridge, MA, USA
xoxu@mit.edu

Anhong Guo

University of Michigan
Ann Arbor, MI, USA
anhong@umich.edu

ABSTRACT

Smartphone overuse poses risks to people's physical and mental health. However, current intervention techniques mainly focus on explicitly changing screen content (i.e., output) and often fail to persistently reduce smartphone overuse due to being over-restrictive or over-flexible. We present the design and implementation of InteractOut, a suite of implicit input manipulation techniques that leverage interaction proxies to weakly inhibit the natural execution of common user gestures on mobile devices. We present a design space for input manipulations and demonstrate 8 Android implementations of input interventions. We first conducted a pilot lab study ($N=30$) to evaluate the usability of these interventions. Based on the results, we then performed a 5-week within-subject field experiment ($N=42$) to evaluate InteractOut in real-world scenarios. Compared to the traditional and common timed lockout technique, InteractOut significantly reduced the usage time by an additional 15.6% and opening frequency by 16.5% on participant-selected target apps. InteractOut also achieved a 25.3% higher user acceptance rate, and resulted in less frustration and better user experience according to participants' subjective feedback. InteractOut demonstrates a new direction for smartphone overuse intervention and serves as a strong complementary set of techniques with existing methods.

CCS CONCEPTS

- Human-centered computing → Interaction techniques; Gestural input; Empirical studies in HCI

KEYWORDS

Smartphone overuse, intervention design, interaction proxy, input manipulation, gestures

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0330-0/24...\$15.00
<https://doi.org/10.1145/3613904.3642317>

ACM Reference Format:

Tao Lu, Hongxiao Zheng, Tianying Zhang, Xuhai 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 (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 19 pages. <https://doi.org/10.1145/3613904.3642317>

1 INTRODUCTION

The rapid growth of digital technologies in the past two decades has led to a boom in convenient access for users to digital products and services. Mobile devices are one of the most highly used digital hardware technologies [68], which have become even more prevalent during the COVID-19 [68, 70]. Recent research has revealed concerns about the overuse of digital devices, such as poor sleep, impaired eye health, digital device addiction, and academic distraction [12, 16, 31, 41, 56, 62, 72]. Many users are aware of these concerns [43], and there is a growing need to reduce mobile device usage [44].

In response, there has been a wide range of research on smartphone overuse intervention techniques. Prior studies mainly focus on two directions: (i) restrictive blocking that introduces interaction barriers to encourage users to stop smartphone entertainment (e.g., [34, 35, 55, 78]), and (ii) alerts or reminders that direct users away from smartphones and penalize overuse (e.g., [24, 37, 50, 54]). However, blocking often becomes over-restrictive and causes a bad user experience, and some can even backfire and cause users to have more smartphone usage [34, 78]. On the other hand, reminders tend to be over-flexible and can be easily ignored. Therefore, there is a need for new intervention techniques that strike the right balance between restrictiveness and flexibility. Our work aims to achieve this goal.

Most of the prior intervention techniques adopted an explicit pattern to change the smartphone's *output* (e.g., locking, showing a notification). Many of these techniques are based on the Dual Process Theory [26], where changing the output can potentially prevent System 1 (an intuition and unconscious process, in this case, the desire for immediate gratification of using smartphones) and trigger System 2 (a logical and analytical process, in this case, the reasoning and reflection on whether they should continue smartphone usage) [44, 78]. However, there is very little work exploring the implicit intervention pattern that alters users' *input* to the

smartphone. Prior work in cognitive science and psychology [7, 19] suggests that the mismatch between the expected results (based on the issued command, i.e., user input) and the actual outcome (after input manipulation [49]) can provide an opportunity to trigger analytic reasoning, i.e., System 2. This suggests a promising direction for smartphone overuse intervention.

We present *InteractOut*, a suite of implicit input manipulation techniques that slightly inhibit the natural execution of common user gestures on smartphones, such as taps and swipes. These input manipulation techniques introduce interaction costs and decrease the smoothness of smartphone interaction to nudge users towards reducing usage. By varying the intervention intensity, we can have precise control over the balance between restrictiveness and flexibility. We present a design space of continuous and discrete manipulations on time, location, direction, and the number of fingers, which are the four main elements of gesture commands. Our manipulation strategies include multiple remapping actions, such as shift, complicate, extend, and disable (Figure 1). Inspired by Interaction Proxies [80], we then demonstrate 8 Android implementations of input interventions throughout our design space, including four tap interventions (Tap Delay, Tap Prolong, Tap Shift, and Tap Double), as well as four swipe interventions (Swipe Delay, Swipe Deceleration, Swipe Multiple Fingers, and Swipe Reverse).

We first conducted a pilot study in the lab ($N=30$) to evaluate how users respond to the interventions, which helped us fine-tune the specific parameters, and understand what might contribute to their effectiveness. Based on the lab study, we identified four most promising interventions (Tap Delay, Tap Shift, Swipe Delay, and Swipe Deceleration) and their intensities. We then performed a five-week field experiment to measure their effectiveness and user receptivity in a real-world deployment. We deployed *InteractOut* with 42 participants in an uncontrolled everyday setting, where users can make their own decisions of smartphone usage in situ. We compared *InteractOut* against the timed lockout baseline, one of the most common intervention techniques supported by modern Android and iOS operating systems [2, 5]. Our results show that *InteractOut* significantly outperformed the lockout baseline by reducing the usage time by an additional 15.0% and reducing the opening frequency by an additional 17.0% on target apps. Meanwhile, *InteractOut* achieved a 25.4% higher user acceptance rate than lockout, which indicated that participants had a significantly higher receptivity towards our techniques. Moreover, participants' subjective feedback indicated that they preferred the *InteractOut* interventions. Users could still continue using their phones with some friction on input. But unlike the timed lockout, where users may forget the time limit they requested, *InteractOut* presents continuous reminders that are more acceptable, has both lower user-perceived friction and higher effectiveness, and strikes a good balance between restrictiveness and flexibility.

InteractOut introduces a new perspective of smartphone overuse interventions that can be compatible with and complementary to existing intervention techniques. The main contributions of our work include:

- We introduce *InteractOut*, a novel set of implicit input manipulation techniques for smartphone overuse intervention. We present a design space of input manipulations that weakly inhibits the natural execution of common user gestures.

- We leverage interaction proxies and develop eight Android implementations of input intervention techniques throughout our design space to illustrate their expressiveness.
- We conduct a pilot lab study ($N=30$) to evaluate the usability of different intervention techniques. Our results identify four most promising intervention techniques: Tap Delay, Tap Shift, Swipe Delay, and Swipe Deceleration.
- We further conduct a five-week deployment study ($N=42$) in the wild. Our results demonstrate significant advantages of *InteractOut* over the common timed lockout baseline technique on smartphone usage time, opening frequency, user acceptance rate, and subjective user experience.

2 RELATED WORK

We first summarize existing intervention techniques for smartphone overuse reduction. We then introduce the theoretical foundation of our input manipulation design. Finally, we introduce related work in interaction proxies, which inspired our intervention design.

2.1 Designs for Smartphone Overuse Intervention

In an age where smartphones have become an integral part of our lives, their overuse has become an increasingly important problem, posing risks for physical and mental health [8, 16, 20, 23, 71, 72]. There exist many research studies, and commercial technologies that offer resources to aid in the reduction of smartphone overuse [42, 51]. According to the strength of obstacles to limit smartphone usage [6], current intervention techniques can be categorized into two groups: (i) strong interventions with restrictive blocking, and (ii) soft interventions with alerts or reminders.

Strong intervention techniques often block smartphone usage, forcing users to stop using the device. Such restrictive and coercive lockout was often used for productivity because it is the simplest way to control smartphone usage, although it often causes user frustration. For example, GoalKeeper [34] studied users' receptivity to different lockout intensity levels and found that users were frustrated by restrictive and coercive interventions. Interaction Restraint [55] and Lock'n>Type [35] explored how cognitive tasks like typing a series of digits could help reduce users' violation of their predefined smartphone usage limit. TypeOut [78] is a more recent study combining self-affirmation [61] and just-in-time intervention [45] to further reduce the annoyance of strong lockout interventions. In commercial applications, the lockout mechanism is also commonly used. For instance, Forest [9], PromodoLock [33], and LessPhone [11] are popular focusing tools using strong lockout to reduce smartphone distractions.

Soft intervention techniques, on the other hand, do not directly limit usage, and instead try to persuade users to reduce usage. Common examples include monitoring and notifications. For instance, MyTime [24] used persistent notifications that displayed usage to reduce users' desire to use smartphones. Social impact and competition have also been utilized to make smartphones less entertaining. NUGU [37] visualized smartphone usage on competitive ranking scoreboards to facilitate self-motivation. Lockn'LoL [36] also used social effects but focused on smaller groups of friends. Let's FOCUS [32] specifically targeted the distraction problem in the classroom

using both reminders and bounded space, limiting most smartphone functionalities. In commercial products, using social competition is also common, such as AntiSocial [79], ActionDash [60], and RealizD [29]. Besides introducing contents of screen output, Good Vibrations [50] studied the effect of subtle, repeating phone vibrations on reducing smartphone overuse.

However, existing techniques are either overly restrictive (i.e., strong intervention) that cause user frustration, or are overly flexible (i.e., soft intervention) that lead to low engagement. For both types of interventions, the effectiveness of existing techniques often comes with an expense of disrupting users. This is often unacceptable when users are in the middle of an important task (e.g., sending an important email). There is a need for new intervention techniques that strike the right balance. Meanwhile, existing intervention techniques mostly adopted an explicit pattern to change the smartphone's *output* (content on the screen, e.g., locking, showing a notification). However, there is no prior work exploring the implicit intervention pattern that alters users' *input* for smartphone intervention. Our work aims to strike the right balance between restrictiveness and flexibility by altering users' input.

2.2 Cognitive Theory Foundation of Input Manipulations

To design an effective intervention technique with minimal user frustration or annoyance, we need to investigate the underlying mental factors that lead to users' decisions to continue or stop smartphone usage. This decision-making process can be largely explained by Dual Process Theory [26, 73, 78]. In this theory, human behaviors are mentally controlled by two "systems" of thinking: an intuitive, unconscious, and spontaneous process (System 1) and an analytical, self-aware, and deliberative process (System 2). To engage users, app designers often try to cater to one's desire for immediate gratification and override their logical thinking [40, 44, 67]. In smartphone applications such as social media platforms, designers use techniques (e.g., following and recommendation features) that constantly refresh bite-size content to attract user attention. To obtain this content, users often just need to use simple tap-and-swipe interactions [17, 25, 65]. Such low cognitive and dexterity requirements can lead to longer usage and eventually addiction to the application.

Reversely, researchers found that harder or more sophisticated interactions can evoke users' System 2 thinking and increase their awareness of smartphone usage [65]. Moreover, existing research in behavior science showed that the mismatch between the expected results and the actual outcome can provide an opportunity to trigger System 2 thinking [7, 64]. This sheds light on our idea of manipulating *input* to introduce a mismatch between the issued command and the output on the screen [49]. We implement this idea, explore the design space of input manipulations, and build InteractOut. More specifically, we modify users' tap and swipe gestures such that they do not receive their expected output change on the screen in time. This expectation mismatch weakens System 1, so that System 2 can take more control of users' decisions.

2.3 Input Manipulation and Interaction Remapping

There has been a wealth of research to improve user experience and accessibility of application interfaces through user interaction remapping [13, 59, 66]. For example, Sugilite [39] is a programming-by-demonstration system that enables users to map a set of UI interaction workflows into a single command. Commercial technologies such as iOS Shortcuts [10] and IFTTT [28] also provide features that allow users to create personalized interaction remappings to simplify phone use. These techniques save operations for repetitive tasks by remapping a number of taps and swipes into zero or a few button clicks. Closer to our work, Interaction Proxies [80] presented a framework to modify user interactions. It serves as a layer between an application's original interface and the manifest interface that a person uses to perceive and manipulate the application. This strategy allows third-party developers and researchers to modify the app interactions without modifying the original application. The initial purpose of interaction proxies is to minimize the gulf of command execution [27], making applications easier to use and improving their accessibility. Their method of using a proxy layer to add or reinterpret the accessibility metadata demonstrates a simple way of runtime modification on smartphone applications. We are inspired by this method and investigate how to use it in the "opposite" way: we can leverage interaction proxies to manipulate input and make phones "harder to use." InteractOut utilizes interaction proxies to increase the difficulty for users to accomplish desired actions on mobile app interfaces. We investigate the effects of such intervention design in smartphone overuse reduction.

3 INTERACTOUT

We first present the design space of implicit input manipulation techniques for overuse intervention. We then demonstrate how our intervention works on the Android operating system, along with 8 Android implementations to instantiate our design space.

3.1 InteractOut Design Space

Most modern smartphones use a touch screen as the primary way for interaction. Touch gestures have become increasingly sophisticated, allowing users to manipulate their smartphones in more convenient and advanced ways [14, 57]. Still, simple tap and swipe gestures remain the most fundamental and common gestures due to their low cognitive cost [47]. Thus, we categorize touch-based interactions into two major groups, tap-based and swipe-based, to facilitate a holistic and systematic intervention design.

Tap-Based Smartphone Interactions. Most interactions users perform on the smartphone are tap-based. This group of interactions does not have continuous moves on the screen but generally requires a precise location. The location is associated with the user's intention of the tap, such as selection, confirmation, and cancellation. The common types of taps include single tap, double tap, long press (fixed duration), and hold (indefinite duration). Our intervention design centers on single and double taps. We focus on manipulating their *time* (i.e., duration) and *location*.

Swipe-Based Smartphone Interactions. Swipe-based interactions are another group of common interactions. They have continuous and directed movement on the screen, and the trajectories are the

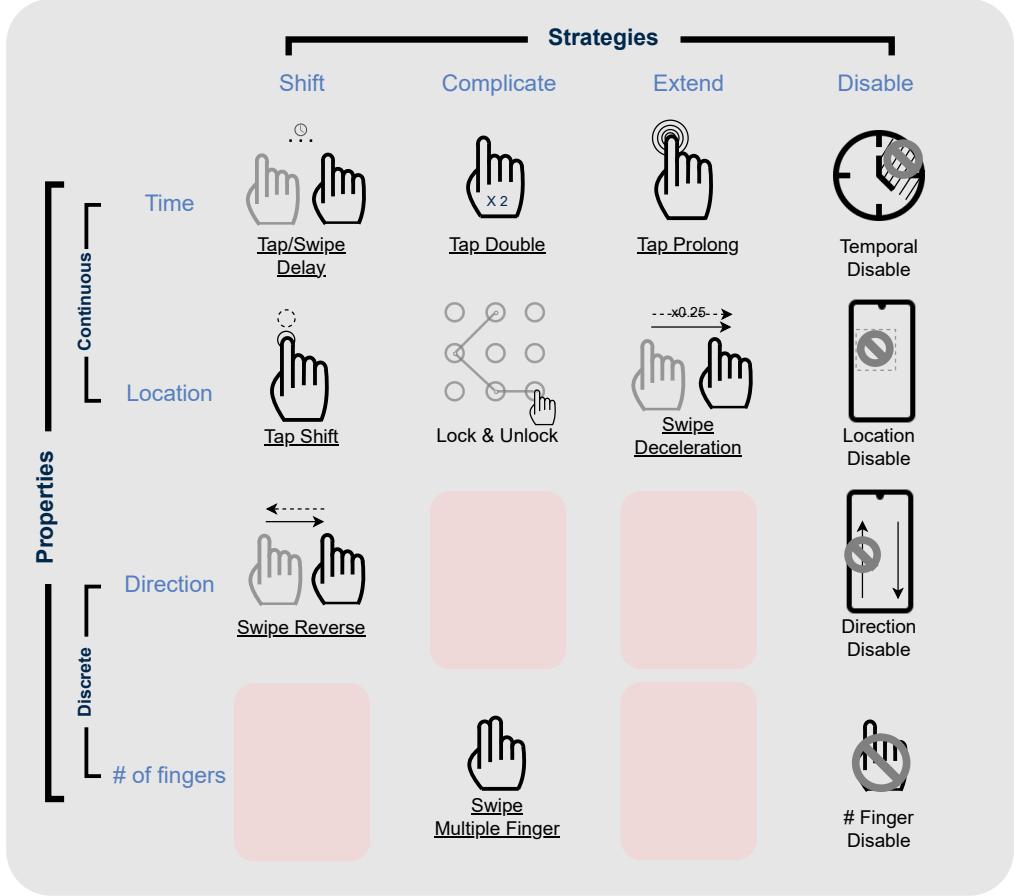


Figure 1: Design Space of InteractOut. The rows include the four essential interaction properties, and the columns show the four manipulation strategies. The underlined techniques are our demonstrated implementations, while others are techniques associated with interactions. Note that pink blocks indicate invalid types.

most important information of a swipe. Swipe-based interactions are often used for navigation and adjustment. For example, users swipe a list to view hidden items due to limited screen sizes or pinch the picture to examine the details. For swipe-based interactions, we focus on the manipulation of *time*, *number of fingers*, and *direction*.

We present a design space for our smartphone interventions. Based on the categorization above, we focus on four essential properties of the interaction: (1) *Time*, (2) *Location*, (3) *Direction*, and (4) *Number of fingers*. To manipulate these properties, we propose four types of strategies that change input from different perspectives:

- (a) *Shift*: Changing the input outcome based on the interaction property. This can be applied to *Time*, *Location*, and *Direction*.
- (b) *Complicate*: Asking for more operations from users to achieve the expected effect. This can be applied to *Time*, *Location*, and *Number of fingers*.
- (c) *Extend*: Requiring extended and longer interaction from users to achieve the expected effect. This can be applied to *Time* and *Location*.
- (d) *Disable*: Nullifying the input effect of the property. This can be applied to all four properties, including *Time*, *Location*, *Direction*, and *Number of fingers*.

This leads to a 4×4 matrix for potential intervention types, as shown in Figure 1. After removing invalid ones (shown as pink blocks in Figure 1), our design space includes 12 intervention types, including a few popular existing techniques, such as Lock & Unlock [2, 5] and the Temporal Disable strategy [9, 34]. Here we propose 8 input manipulation intervention techniques.

- Tap Delay: Users' tap is postponed for a certain period before it takes effect;
- Tap Prolong: Users need to press the screen longer than a threshold to trigger a tap;
- Tap Shift: Users' tap is shifted in a fixed direction away from the actual tap position;
- Tap Double: Users need to tap twice to trigger a single tap;
- Swipe Delay: Users' swipe is postponed for a certain period before it takes effect;
- Swipe Deceleration: The effect of a swipe is slower than the users' actual gesture trajectory;
- Swipe Reverse: Users' swipe direction is reversed;
- Swipe Multiple Fingers: Users need to use more than one finger to swipe.

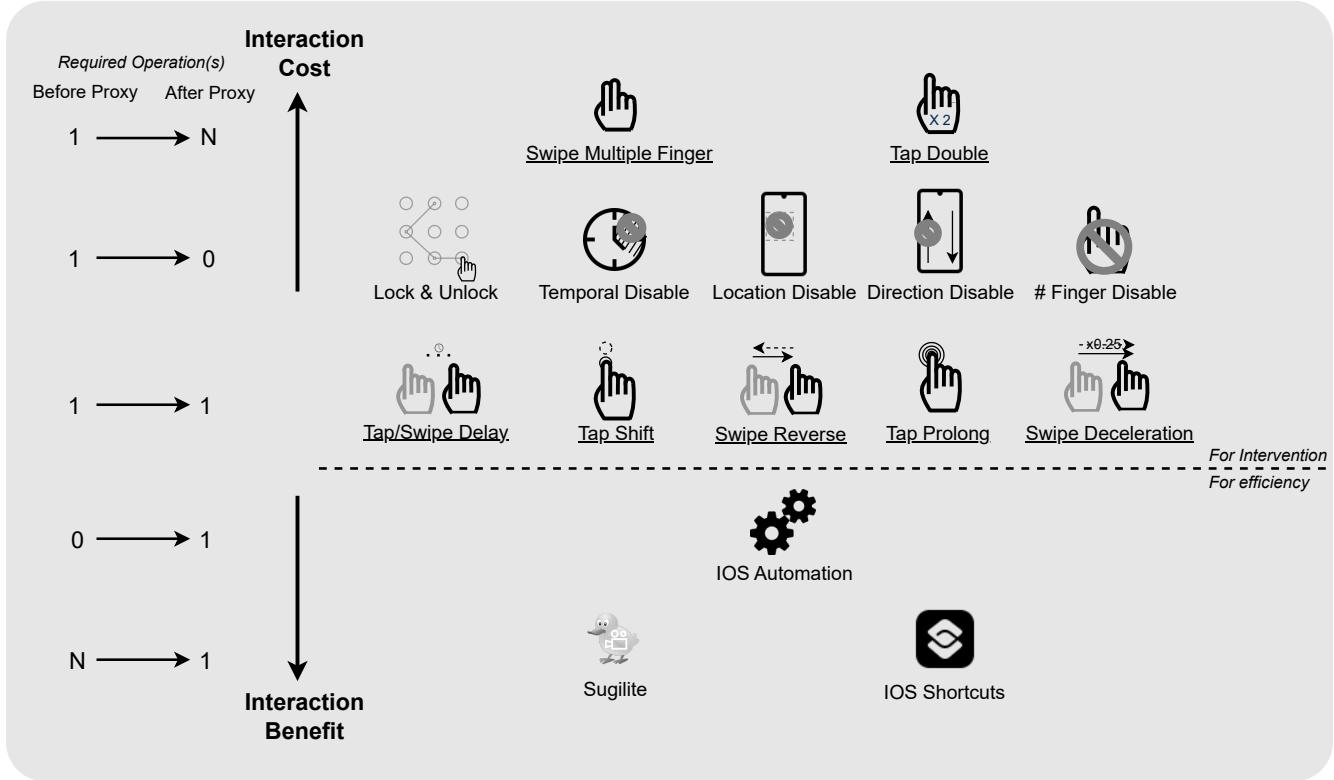


Figure 2: Summary of Remapping Methods According to the Design Space of Interaction Proxies [80]. The underlined techniques are our demonstrated implementations. “1”, “0”, and “N” here means the number of operations/interactions needed to perform a task before and after the proxy.

Leveraging Interaction Proxy remapping [80], these intervention techniques replace the original interaction with more difficult and counter-intuitive micro-tasks that require more physical or mental efforts. Figure 2 presents a summary of the remapping methods which includes our design space as well as existing remapping techniques. Note that most existing remapping techniques are designed to facilitate interaction rather than to introduce friction.

There are five remaining interventions our design space affords, including (1) Lock & Unlock: Users need to do some tasks to continue the usage, i.e., sliding a block or typing numbers; (2) Temporal Disable: The screen is disabled for a period of time; (3) Location Disable¹: Certain areas of the screen is disabled; (4) Direction Disable: Certain direction of interaction (mainly swipe) is disabled; and (5) # of Finger Disable: The operation (mainly tap) is disabled if performed with a certain number of fingers. Existing intervention techniques often combine the Temporal Disable and Lock & Unlock to block smartphone usage. Users set a time limit and the app/screen will be blocked when the usage exceed the limit (i.e., Temporal Disable). Users have to stop or follow a specific unlock procedure to continue (i.e., Lock & Unlock). In this paper, we do not cover the four disable interventions as they are overly restrictive

and would compromise user experience. We use Lock & Unlock as the baseline as it is already widely adopted by modern smartphones.

3.2 InteractOut Mobile App

Intervention Implementations. We instantiate InteractOut as an application on the Android system. The left of Figure 3 shows the case when users operate the phone normally. Users perform gestures on the screen. The phone recognizes the operation and passes it to the application layer. The application responds according to the operation (e.g., display new content with a swipe down). InteractOut inserts a proxy layer between the screen and the application layer, as shown with the blue color on the right of Figure 3. The proxy layer first consumes and recognizes user input gestures. Then, it uses the gesture information to build “virtual gestures” (e.g., a swipe with a reversed direction). Finally, it passes the virtual gestures to the application layer.

To implement the proxy layer in Android, we use the Accessibility Service [1]. We adopt a transparent Accessibility overlay as our proxy layer, and customize a Gesture Detector class to obtain the input gesture information, including the duration, location (or trajectory) and the number of fingers. To build a virtual gesture, we employ the dispatchGesture API. This API takes gesture information as parameter and performs the gesture on the current interface. Thus, we can easily manipulate gesture information to create the input interventions, as introduced in Section 3.1.

¹Note that there are also disabling techniques based on the geographical location of the smartphone [10, 32], and they are different from the “Location Disable” we present in the design space.

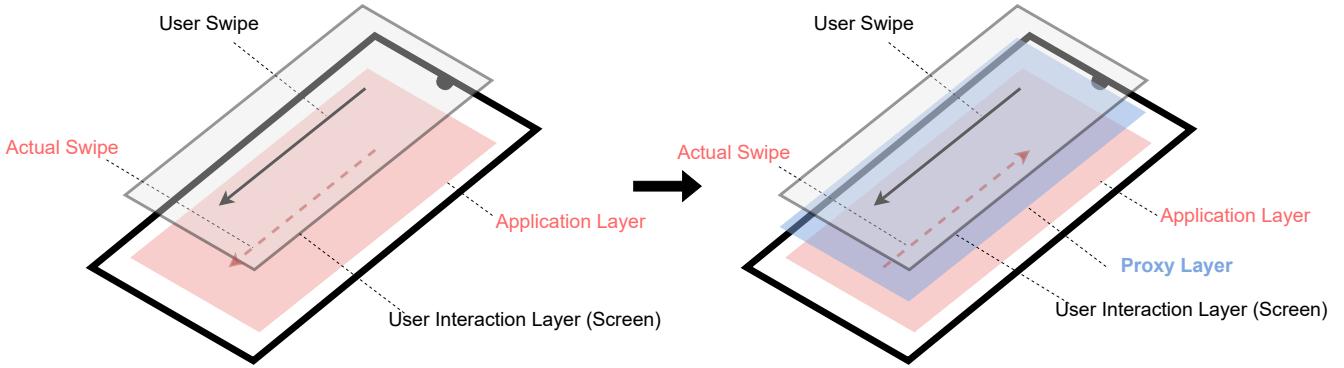


Figure 3: Structure of the InteractOut Implementation on Android. The proxy overlay takes user input and outputs the modified gesture. Swipe Reverse intervention is used for illustration.

We implement our input manipulation intervention techniques as follows:

- Tap Delay: Add a delay $T_{\text{tap-delay}}$ to execute `dispatchGesture` for taps;
- Tap Prolong: Set a threshold of tap duration $T_{\text{tap-threshold}}$ and ignore the taps whose duration is less than the threshold;
- Tap Shift: Add a pair of fixed shift values (x, y) to the coordinate of the users' single tap;
- Tap Double: Remap a double tap to a single tap, and ignore single taps. When the input is double tap, the output is single tap. Note that output cannot be double tap as the input double tap is remapped;
- Swipe Delay: Add a delay $T_{\text{swipe-delay}}$ to execute `dispatchGesture` for swipes;
- Swipe Deceleration: Multiply the duration of gesture with a factor $F_{\text{swipe-decelerate}}$ so the effect of the swipe is decelerated;
- Swipe Reverse: Replace the trajectory of a swipe with a reversed one;
- Swipe Multiple Fingers: Set a threshold of a number of fingers N and ignore the swipes whose number of fingers is fewer than the threshold.

Intervention Delivery Mechanism. Our implementation allows users to set which apps they want to intervene and the usage limit. InteractOut monitors the usage time for each app via the Accessibility API. When the limit of an application is reached, the proxy overlay will launch with selected interventions.

InteractOut is highly customizable. For interventions with thresholds, their intensity can be adjusted by changing the threshold. The configuration of the intervention and its intensity can also be customized for individual applications. It is noteworthy that for the interventions with parameters $T_{\text{tap-threshold}}$, $T_{\text{tap-delay}}$, and $F_{\text{swipe-decelerate}}$, we propose a dynamic design for the interventions to gradually take effect: the intervention intensity will start from zero and gradually increase to a target maximum value. Specifically, we apply a linear step increase on $T_{\text{tap-threshold}}$, $T_{\text{tap-delay}}$ and $T_{\text{swipe-delay}}$, and an exponential step increase on $F_{\text{swipe-decelerate}}$. The step increase can be triggered by either a

user operation (a tap, swipe, etc.) or the time elapsed since the last increase. In this way, users will perceive very subtle interventions at first, which then become increasing stronger over time.

Notification and Bypass Mechanism. InteractOut runs in the background and does not have an interface when delivered. Instead, we leverage a persistent notification in the notification drawer [4] to inform users of the intervention techniques and their intensity (if applicable). Explicit notifications are not necessary, as users will notice the input manipulations as they start to take effect. Furthermore, following the implementation of the iOS Screen Time bypass options, we provide a bypass mechanism with the same options for users to select from three options (Figure 4) “One more minute”,

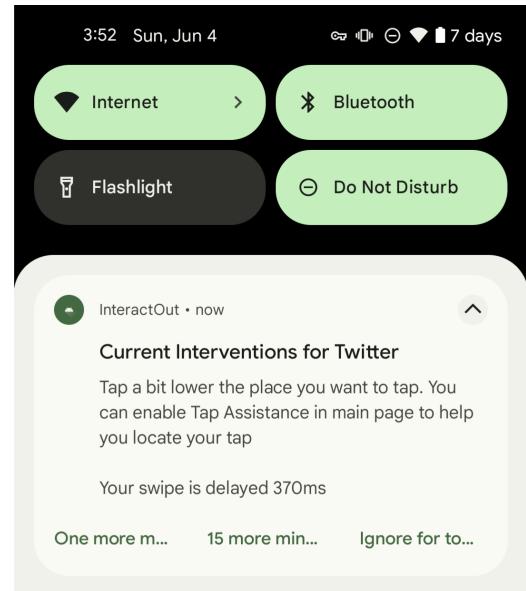


Figure 4: Bypass Option Menu in Android Notification Drawer. User can select 1 or 15 minutes or ignore the limit for the rest of the day directly in the notification.

Table 1: The Two Intensity Levels Used in the Lab Study. Note that dp is a density-independent pixel whose actual length is based on the physical screen.

| Gesture Category | Manipulation Strategy | Level 1 | Level 2 |
|------------------|-----------------------|---------|---------|
| Tap | Delay | 500ms | 1000ms |
| | Prolong | 100ms | 200ms |
| | Shift | y+100dp | y+200dp |
| | Double | yes | / |
| Swipe | Delay | 300ms | 800ms |
| | Deceleration | ×0.5 | ×0.25 |
| | Reverse | yes | / |
| | Multiple fingers | 2 | 3 |

“15 more minutes”, or “Ignore for today” directly through the notification drawer to pause the intervention. This notification is always available so that users can bypass the intervention at any time. We also offer these bypass options in the InteractOut app.

To finalize the InteractOut implementation, one key question remains: what should the intervention intensity be to effectively reduce smartphone overuse? We explore this in the next section.

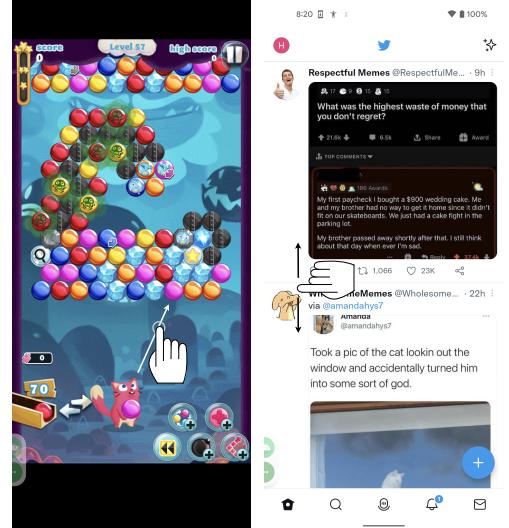
4 LAB STUDY

As introduced in Section 3, InteractOut can support various intervention implementations and intensities. This leads to the question of finding the right balance between restrictiveness (as over-restrictiveness leads to bad user experience) and flexibility (as over-flexibility leads to low engagement). To answer this question, we first conducted a pilot lab study to compare different implementations of InteractOut and their intensity levels. Our findings about the appropriate intervention and intensity selection will inform the field deployment experiment in Section 5 to evaluate InteractOut in real-world settings.

4.1 Participants

We recruited participants from our university’s mailing lists. We asked participants to fill out a screening survey to evaluate their smartphone addiction level (SAS-SV [38]) and their desire to change smartphone usage behavior (TTM [58]), and to collect their demographics information. We set a SAS-SV score threshold of 30 (indicating smartphone-additive) and a TTM threshold of beyond stage (indicating the desire to change behavior). From a pool of 112 survey respondents who met the inclusion criteria, we randomly sampled 30 participants (10 female, 20 male, aged 19 to 30).

Figure 5: The Two Mobile Applications Used in the Lab Study. (a) Bubble Mania, for tap implementations. (b) Twitter, for swipe implementations.



4.2 Design and Procedure

We employed a within-subject design with a baseline condition of no intervention and our InteractOut condition with 8 gesture intervention implementations (see Section 3.1). For each implementation, each of them was assigned two different intensity levels (see Table 1). These levels were chosen based on UI design principles (e.g., important limits of response times [48]) and early pilot studies. We chose Twitter (now known as X) [74] and Bubble Mania [63], two popular and addictive apps [18], which represented apps with swipe-heavy and tap-heavy interactions, to mimic realistic smartphone usage scenarios. We used a Google Pixel 5a running Android 12.

At the beginning of our study, participants were asked to complete a pre-study survey to collect commercial intervention products they have used. They completed an interactive tutorial to familiarize themselves with the 8 gesture intervention implementations. Then, participants were asked to test each technique for 3 minutes on the two assigned apps. The order of implementations was counterbalanced to minimize order effects. After each implementation stage, participants were asked to complete a short survey on a 7-point Likert scale regarding their perception of each implementation. The questions include annoyance level, disruptiveness, usage desire reduction, mental load, and physical load. Moreover, if applicable, we asked participants which intensity level they would prefer to use in daily life. After going through all implementations in one category (tap or swipe), they were asked to rank the four manipulation strategies based on perceived effectiveness and receptivity in reducing the desire to continue using the app. Finally, we conducted an exit interview about their experience using these implementations and general comments. Our study was approved by our institution’s IRB and lasted about 2 hours. Participants were compensated with \$30 for their time.

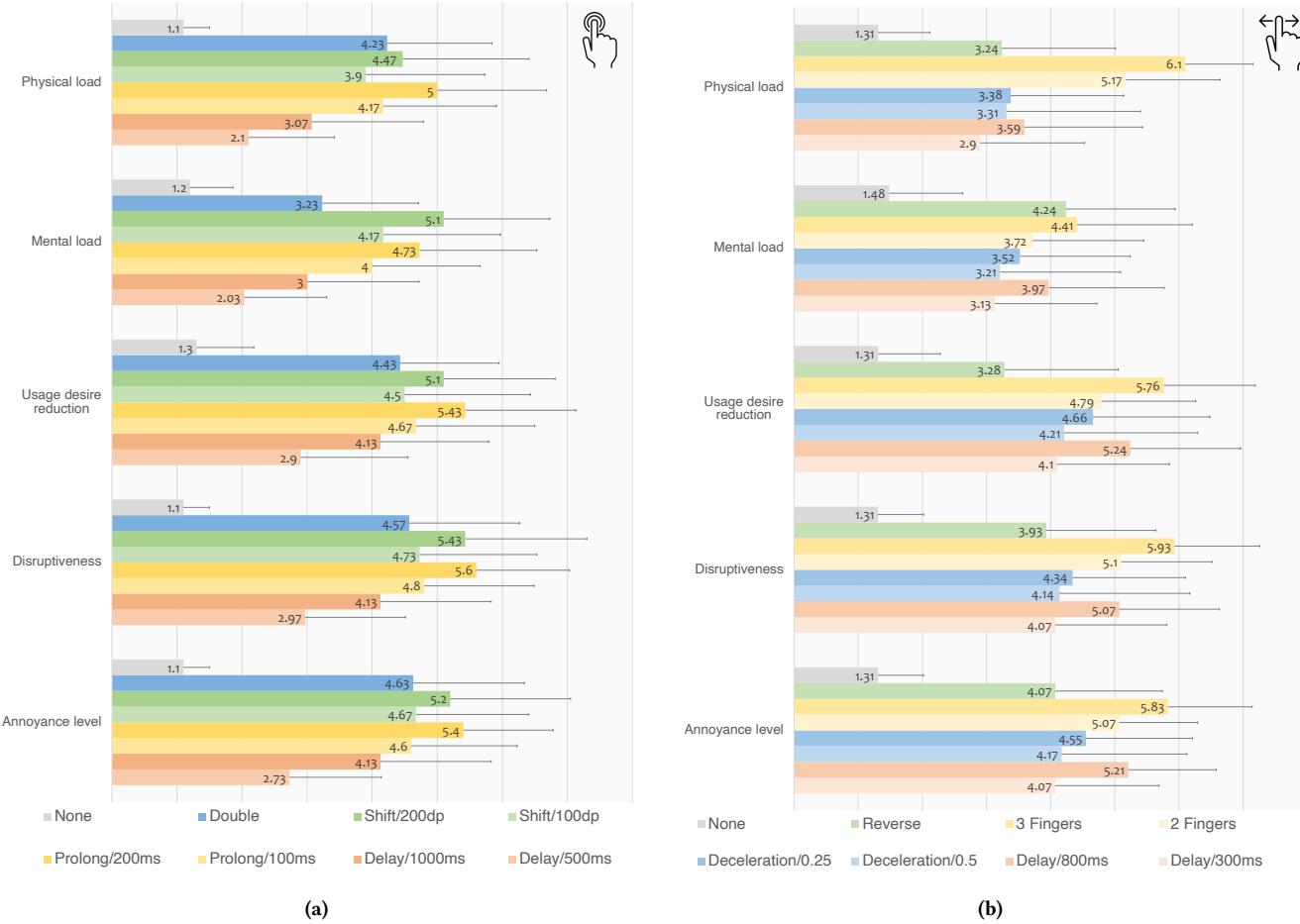


Figure 6: Participants' Likert Scale Responses on (a) Tap Implementations and (b) Swipe Implementations on a 7-point Likert Scale (1 - Strongly disagree and 7 - Strongly agree). As expected, InteractOut are significantly more annoying, more disruptive, more mentally and physically demanding than the baseline condition with no intervention.

4.3 Results

Figure 6 presents the questionnaire results, and Figure 7 summarizes the rankings results. Not surprisingly, we found that both tap and swipe implementations were significantly more effective than the baseline with no interventions (both $p < 0.0001$). We also found that participants generally preferred the higher intensity level of Tap Delay, Tap Shift, Swipe Delay, and Swipe Deceleration.

4.3.1 Implementation Intensity Levels Comparison. We found that most participants preferred the higher intensity level for most implementations (Figure 8), even though the high-intensity level could be more disruptive and demanding. Participants could easily ignore the implementations with the lower intensity level: “500ms delay is too short that I could not even tell if there is intervention or not” (P21), “I need to put extra effort to (successfully) tap the screen when it’s 200ms prolonged compared to 100ms” (P4). However, one exception was for Swipe Deceleration. Between the $\times 0.5$ and $\times 0.25$ levels, participants could hardly tell the difference, as shown in the Likert scale responses ($p = 0.42$) and preferences (9 vs. 13). In summary, as shown in Figure 6 and Figure 8, the higher intensity

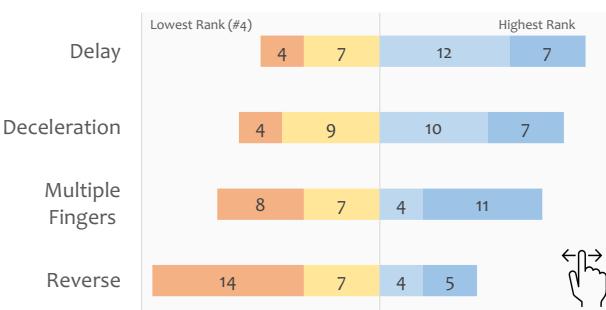
levels we chose tend to be more effective and fairly acceptable. Thus, in the next subsection, we only refer to the higher intensity level if not explicitly stated.

4.3.2 Intervention Implementations Comparison. We found that the effectiveness and receptivity of an implementations is related to its physical and mental loads. **Implementations with a higher mental load but a lower physical load were the most preferred.** They were effective because participants needed to think before taking actions, but they were not too annoying because it took no more physical effort than normal actions once participants got used to it. For example, P4 and P12 both mentioned that the Tap Shift in Bubble Mania and other similar use cases enlarges the gulf of execution by requiring additional cognitive load to think about where to activate.

Meanwhile, the less preferred implementations were those with high physical load. For instance, the Swipe Multiple Fingers was extremely physically demanding. Many participants had to adopt unnatural postures to swipe with multiple fingers. Interestingly, despite its strong annoyance, some participants still ranked it as the



(a)



(b)

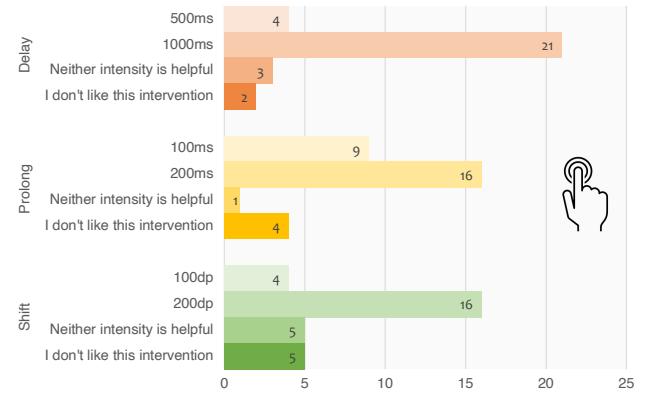
Figure 7: Participants' Perceived Effectiveness and Receptivity of Manipulation Strategies within (a) Tap Implementations and (b) Swipe Implementations. Tap Shift and Swipe Delay were the most preferred, while Tap Double and Swipe Reverse were the least preferred.

most preferred implementation because it was effective yet quite intuitive (thus a low mental load). The least effective implementation was Swipe Reverse, which has both low physical and mental loads. Most participants made several mistakes at first, but they could easily get used to it within a few trials: “*It reminds me of my first use of the MacBook trackpad*” (P22).

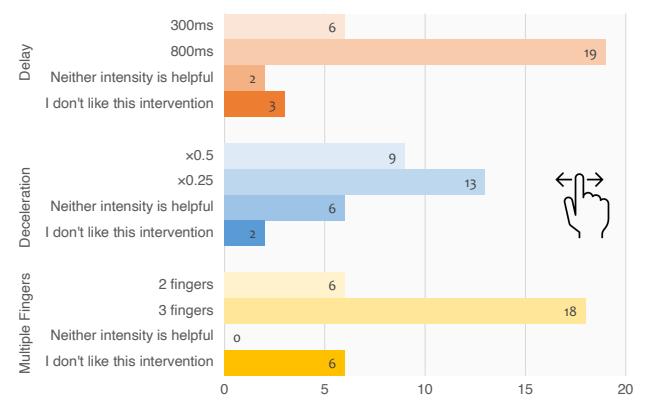
Through the lab study, we identified that among the methods we tested, the higher intensity level of Tap Delay, Tap Shift, Swipe Delay, and Swipe Deceleration are the four implementations that are better at striking the balance between effectiveness and user experience. However, due to the short experiment duration during the lab study, it is unclear whether these four implementations will still be as effective in real-world settings. Furthermore, how users balance mental and physical load could be different in lab vs. field settings. Therefore, the lab study not only informed the implementation and intensity selection and design of our intervention app, but also motivated us to deploy our app in the real world.

5 FIELD EXPERIMENT

Based on the lab study, we developed an intervention app consolidating the four preferred intervention techniques. We then conducted a five-week field experiment. The goal of the field experiment is to understand how InteractOut influences smartphone usage and user receptivity in real-world scenarios. We also compare InteractOut against one of the most common intervention methods, the Timed



(a)



(b)

Figure 8: Participants' Preferences on the Two Intensity Levels of (a) Tap Implementations and (b) Swipe Implementations. Higher intensities were always the preferred choice for all implementations.

Lockout intervention. Our results indicate that InteractOut outperforms the Timed Lockout with significantly more smartphone usage reduction and a higher intervention acceptance rate.

5.1 Intervention Techniques

We compare InteractOut against a baseline intervention technique Timed Lockout, a popular technique widely available on modern smartphones.

5.1.1 InteractOut Interventions. Our final version of the intervention includes the four implementations with the high-intensity level: Tap Delay (1000 ms), Tap Shift (y=200 dp), Swipe Delay (800 ms), and Swipe Deceleration ($\times 0.25$). All interventions except for Tap Shift follows the step increase feature we discussed in Section 3.2. Every time users touch the screen, the intervention intensity increases by a small step (10 ms for Tap/Swipe Delay and $4^{-\frac{1}{100}}$ for Swipe Deceleration). This means that the intervention will saturate within 100 touches for Tap Delay and Swipe Deceleration and 80 touches for Swipe Delay. If users do not touch the screen for one minute, the intensity will also increase by one step.

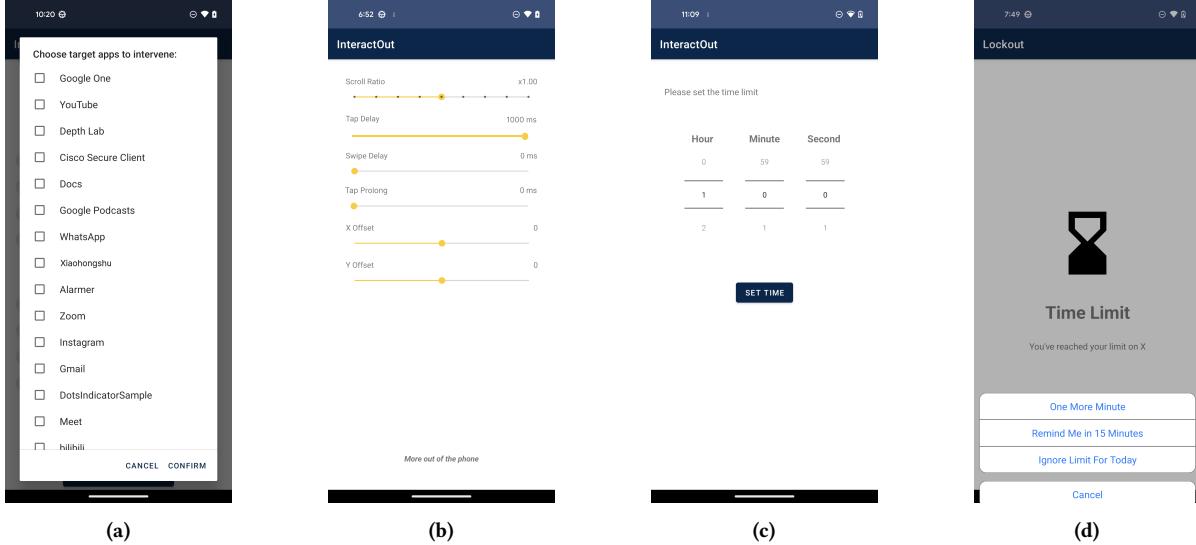


Figure 9: (a) **Target App Selection Page.** Users specify the target apps to intervene during the onboarding stage of the field experiment; (b) **InteractOut’s Configuration Page for Interventions and Intensities** and (c) **InteractOut’s Configuration Page for Total Time Limit** – Note that users cannot change these configurations for the purpose of the study. We discuss future use cases in Section 7.5; and (d) **Lockout’s Bypass Menu**, triggered by selecting “Ignore Limit” in the popup blocking screen. Note that **InteractOut’s bypass menu** is shown in Fig. 4.

Our app provides a config stage, where users can choose the apps that they want to receive intervention as well as the time budget (see Figure 9a and 9c). In our field experiment, we fix the daily time budget as one hour in total for all target apps combined to ensure consistent control in our study, and for most participants to reach the limit and experience the intervention. But in practice, users can adjust the budget based on their needs. The bypass menu of InteractOut is shown in Figure 4. It is also available on the main page of the InteractOut app.

5.1.2 Baseline: Timed Lockout. Timed Lockout is one of the most common intervention techniques supported by modern Android and iOS operating systems [2, 5]. Users can set time limits based on individual apps or groups of apps. Once the limit is reached, the Timed Lockout intervention will pop out a blocking interface notifying the user as well as providing options to bypass the intervention to request more usage (see Figure 9d). For our study, we also fixed the time limit to be one hour total, making it consistent with InteractOut. For simplicity, we use “Lockout” as an abbreviation of the Timed Lockout intervention.

5.2 Participants

We recruited a new group of participants from our university’s mailing lists. Similar to the lab study, we used the same screening survey and inclusion criteria. Moreover, we required participants to be regular Android users. From a pool of 69 survey respondents who met the inclusion criteria, a total of 42 participants joined and completed the field experiment (11 female, 31 male, aged 18 to 34). Their Android device manufacturers include Samsung (23), Google (11), OnePlus (2), Xiaomi (2), Motorola (3) and LG (1). None of them participated in the previous lab study.

5.3 Design

5.3.1 Experimental Design. During the experiment, we adopted a within-subject design. The 5-week schedule of the field experiment is shown in Figure 10. The first week was used for the base measurement of participants’ usual smartphone usage data without any intervention, followed by 2 two-week intervention stages, each including 12 intervention days and 2 break days. The break days were inserted to investigate the lasting effect of interventions (see Sec 6.1.3) and to reduce the influence between interventions. Participants used one intervention in the first intervention stage and used the other in the second stage. The order of the intervention technique was counterbalanced. For InteractOut, since combining the two tap interventions or the two swipe interventions may introduce overly high interaction friction, we randomly picked one tap and one swipe intervention for each participant, resulting in four intervention combinations.

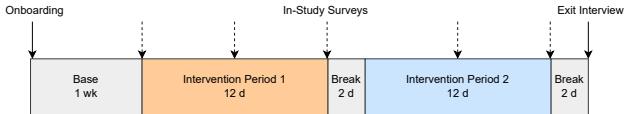


Figure 10: Field Experiment Schedule. The order of the two interventions is counterbalanced. The break days are inserted to investigate the lasting effect of interventions and to reduce the influence between interventions. There are 5 in-study surveys in total, 1 at the end of the Base week and 2 for each intervention period, with one in the middle and one at the end.

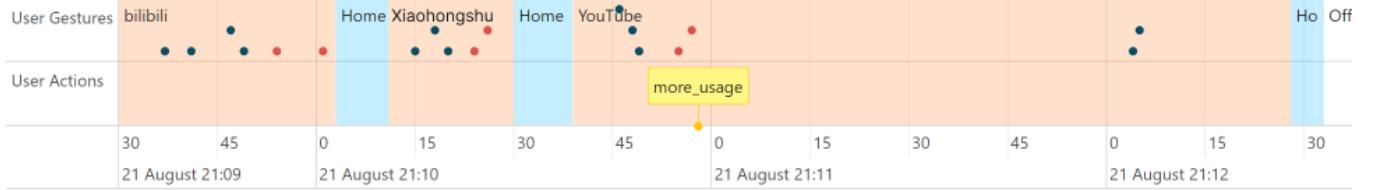


Figure 11: Web-Based Visualization Tool for App Activity Visualization. This example shows a sample app activity timeline under InteractOut. Black and red dots indicate swipes and taps during the intervention. This tool serves two purposes: (1) to provide data collection transparency to end users and help them better understand the data, and (2) to help researchers better monitor study compliance and troubleshoot potential issues. Gesture events are only logged when intervention is on.

5.3.2 Evaluation Metrics. Our evaluation metrics focus on the app usage (usage time and opening frequency) change and acceptance rate of the intervention methods. Thus, we logged app enter and exit events, as well as the instances of intervention encounter and bypass on participants' devices throughout the study. The data collection was done through the Google Firebase database system [21]. Each week, we also delivered an in-study survey to collect workload using NASA TLX [30], and subjective feedback on the two interventions. At the end of the study, we conducted an exit interview for each participants for their overall perception of 2 intervention techniques. At the end of the study, participants attended an offboarding session during which we conducted a semi-structured exit interview, and uninstalled the study app from the participants' devices.

5.3.3 Visualization Tool. Moreover, to better understand the data and control study compliance, we developed a web-based visualization tool (Figure 11) to show real-time app activities and daily activity summaries. We used this tool to keep track of participants' app activities and troubleshoot outlier patterns (e.g., a multi-day gap between two app events). During onboarding, we also used this tool to show participants their live data we were collecting, which provided them with data collection transparency.

5.4 Field Experiment Procedure

Our study was conducted on a rolling basis from March 8th to April 28th, 2023. In the onboarding session, participants came to our lab, installed our app, learned the features of their assigned gesture interventions, got familiarized with the visualization tool, and specified the target apps to apply screentime interventions. The most frequently selected target apps included Instagram (25/42 participants), YouTube (18/42), Reddit (10/42), Discord (8/42) and Twitter/X (7/42). To manage the study, researchers used a Python script to automatically monitor usage activities and switch study phases, while manually sending out surveys according to a fixed schedule (Figure 10). After five weeks of the intervention period, participants went through the exit interview and ended the study. Our study was approved by our institution's IRB. Participants were compensated with up to \$100 USD based on the number of in-study surveys they completed and the days data were uploaded.

6 FIELD EXPERIMENT RESULTS

During the five-week study, we collected a total of 852,556 app activities, 15,042 intervention encounters, 2,671 intervention bypasses, and 214,125 interactions during InteractOut intervention

active period (single tap, double tap, and swipe). We aggregated these events to analyze each participant's app usage behavior and intervention acceptance rate. We analyzed the quantitative data and the qualitative data collected via questionnaires and interviews.

6.1 App Usage Behavior

The primary goal of our intervention is to reduce app usage. Thus, we first investigate the app usage behavior: app usage time and app opening frequency. Due to the large app usage variation among individuals, we normalized each participant's data by calculating the ratio against their own data during the base week when there was no intervention. A ratio smaller or greater than 1 indicated that participants reduced or increased app usage compared to their ordinary behavior. Our results showed that participants had a significantly smaller ratio when using InteractOut compared to the Lockout baseline on their target app usage behavior.

6.1.1 App Usage Time. Both InteractOut and Lockout significantly decreased participants' target app usage compared to naturalistic usage, while InteractOut had a significantly larger time reduction. The average usage time was 149.04 ± 7.24 min in the base week and 131.20 ± 5.56 min, 103.70 ± 4.67 min for Lockout and InteractOut, respectively. We then used the normalized data for comparison. As Figure 12 shows, participants had a large usage time reduction of target apps during the InteractOut period (Mean = 0.78, Std = 0.029), which is 15.6% lower than Lockout (Mean = 0.94, Std = 0.031). A Shapiro-Wilk normality test showed that the usage time did not follow a normal distribution. Therefore, we used a General Linear Mixed Model (GLMM) with a Log-Gamma link function (based on a Kolmogorov-Smirnov test) and participant ID as a random effect. The fixed effects were *Intervention* (with levels "Natural", "InteractOut" and "Lockout"), *Order* (with levels "InteractOut first" and "Lockout first"), their interaction (*Intervention* \times *Order*), and *Implementation* (with four levels being the 2 \times 2 combinations of tap/swipe interventions, as noted in Section 5.3). Results showed that *Intervention* was the dominant factor of the usage time change ($\chi^2 = 66.2$, $p_{\text{Intervention}} < 0.001$), while all other factors did not show significance ($p_{\text{Order}} = 0.52$, $p_{\text{Intervention} \times \text{Order}} = 0.21$, $p_{\text{Implementation}} = 0.72$). It is worth noting that the four combinations of input intervention techniques achieved similar results, indicating their robustness. A follow-up pairwise post-hoc Tukey's HSD test on *Intervention* further showed that both InteractOut ($Z = -8.52$, $p < 0.001$) and Lockout ($Z = -4.03$, $p < 0.001$) lowered the usage time, while InteractOut had a significantly lower

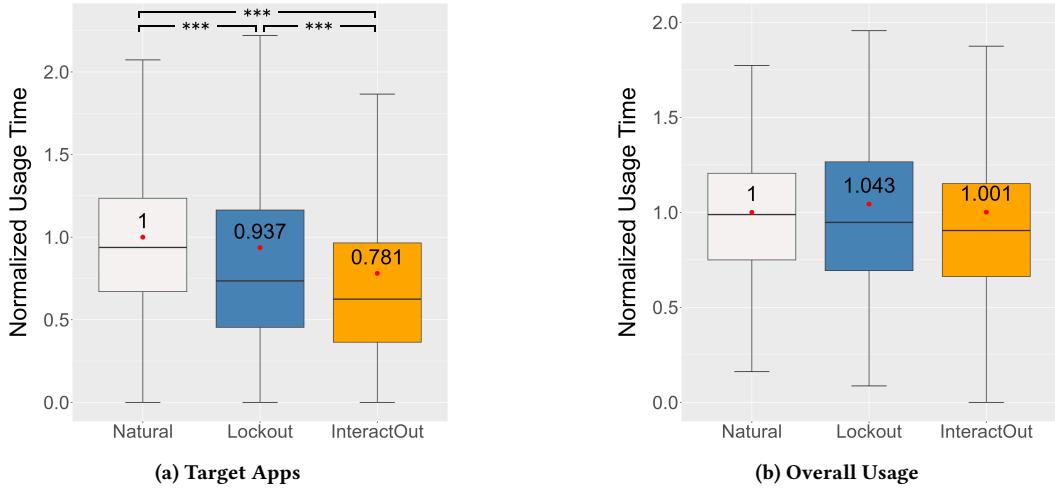


Figure 12: Normalized App Usage Time in the Intervention Conditions. It shows the normalized ratio of the usage time against the average usage time in the base week. Lower ratio means more effective reduction. In each box, the middle black line shows the median and the red point shows the mean. The same apply to other box plot figures. Note that the minimum usage time is 0 as some participants do not use their target apps in some days.

usage time than Lockout ($Z = -4.41, p < 0.001$). However, when looking at the usage time of all apps, we observed no significance ($p_{\text{Intervention}} = 0.13, p_{\text{Order}} = 0.86, p_{\text{Intervention} \times \text{Order}} = 0.78, p_{\text{Implementation}} = 0.63$), as shown in Figure 12b. One explanation could be that the decreased usage of target apps canceled out with the increased usage of non-target apps, which could include many non-overuse cases.

We further investigated the influence of app interaction intensity on the usage time change. We categorized each target app based on their average number of screen touches per minute (interactions per minute, IPM) during InteractOut's intervention-on period. The thresholds were chosen so that 3 groups had almost the same number of target apps (Low: $\text{IPM} \leq 10$; Medium: $10 < \text{IPM} \leq 18$; High: $\text{IPM} > 18$). As Figure 13 shows, we found **both Lockout and InteractOut showed the best performance in the medium-intensity group, while InteractOut showed a significantly stronger usage reduction in the medium and high-intensity group**. A GLMM on fixed effect *Intensity_Level*, *Intervention* and *Intensity_Level* \times *Intervention* showed significance for *Intensity_Level* ($\chi^2 = 10.5, p < 0.05$) and *Intensity_Level* \times *Intervention* ($\chi^2 = 6.32, p < 0.05$). We then compared Lockout and InteractOut in each intensity group. A GLMM with post-hoc Tukey's HSD test on fixed effect *Intervention* showed that InteractOut had significantly lower usage time than Lockout in the high ($p < 0.001$) and medium ($p < 0.001$) intensity group, but not the low group ($p = 0.36$). This shows that InteractOut had less impact on users' app usage in low-intensity apps. A potential reason is that the small number of input interactions limits the manipulation of user input.

6.1.2 App Opening Frequency. We observed similar results for opening frequency, where InteractOut reduced smartphone usage significantly more than Lockout. The average opening frequency was 66 in the base week and 57, 48 for Lockout and

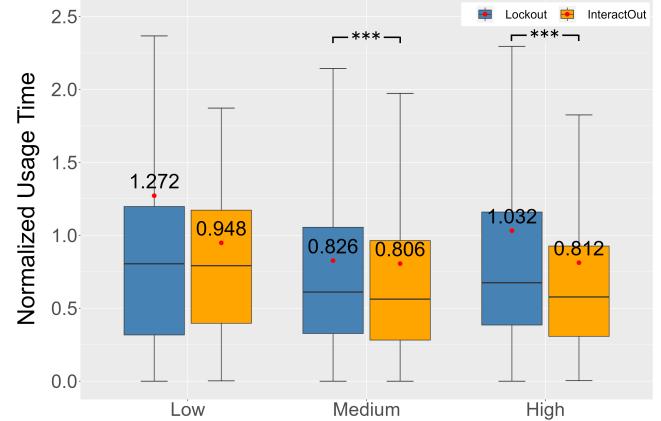


Figure 13: Normalized Target Apps Usage in 3 Levels of Interaction Intensity (Low, Medium, High). Interaction intensity was categorized based on the number of required touch interactions for each target app. It shows the normalized ratio of the usage time against the average usage time in the base week in each interaction intensity level group. Lower ratio means more effective reduction. In each box, the middle black line shows the median and the red point shows the mean. Note that the minimum usage time is 0 as some participants do not use their target apps in some days.

InteractOut, respectively. We also used the normalized data for comparison. Figure 14 shows that participants opened target apps 16.5% less in the InteractOut period (Mean = 0.78, Std = 0.023) than they did in the Lockout period (Mean = 0.95, Std = 0.024). We ran a GLMM with the same configuration on the normalized opening frequency. Results showed that *Intervention* was also the main factor

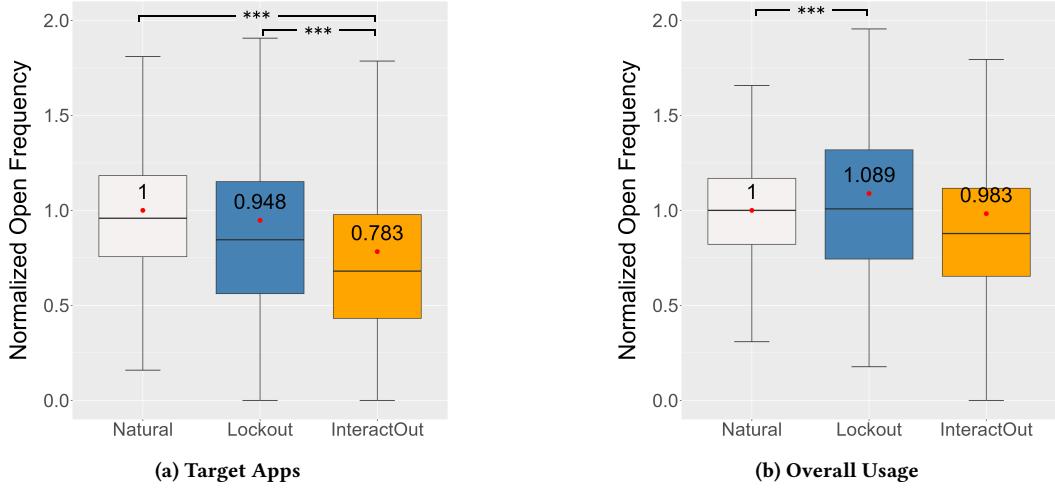


Figure 14: Normalized App Opening Frequency in the Intervention Conditions. It shows the normalized ratio of the usage time against the average usage time in the base week. A lower ratio means more effective reduction. Note that the minimum opening frequency is 0 as some participants do not use their target apps in some days.

of opening frequency changes ($\chi^2 = 56.2, p_{\text{Intervention}} < 0.001$), while the *Intervention* \times *Order* interaction also had some marginal effect ($p_{\text{Intervention} \times \text{Order}} = 0.05, p_{\text{Order}} = 0.84, p_{\text{Implementation}} = 0.26$). A pairwise post-hoc Tukey's HSD test on *Intervention* showed that InteractOut ($Z = -5.84, p < 0.001$) significantly lowered the opening frequency of target apps compared to the base week. In contrast, Lockout showed a similar level of opening frequency as participants' natural behaviors in the base week ($Z = -0.75, p = 0.73$). InteractOut showed a significantly lower opening frequency compared to Lockout ($Z = -5.35, p < 0.001$). For overall app usage, a GLMM showed the significance of *Intervention* ($\chi^2 = 20.0, p < 0.001$). The post-hoc test showed significantly higher opening frequency in the Lockout period ($Z = -2.9, p < 0.05$) and lower frequency during the InteractOut period, yet without significance ($Z = 1.3, p = 0.39$).

We also investigated the opening frequency change on different app interaction intensities. Figure 15 shows that **InteractOut had significantly lower opening frequency than Lockout in all 3 intensity groups, but there was no significance between 3 intensity groups**. A GLMM with post-hoc Tukey's HSD test showed a significantly lower opening frequency in all intensity groups ($p_{\text{low}} < 0.001, p_{\text{medium}} < 0.001, p_{\text{high}} < 0.01$).

6.1.3 Lasting Effect on App Behaviors. Both Lockout and InteractOut kept a decreased app opening frequency but did not retain the performance in usage time. We compared participants' app behaviors on break days to the data on intervention days to investigate the lasting effect of interventions. A GLMM with the same configuration showed the significance of *Intervention*. A pairwise post-hoc Tukey's HSD test on *Intervention* further showed the decreased opening frequency ($Z = 5.1, p < 0.001$) but similar usage time ($Z = 1.6, p = 0.27$) in InteractOut break days compared to the base week. Similarly, in Lockout break days, we also found a slightly decreased opening frequency ($Z = 2.9, p < 0.01$) and similar usage time ($Z = 1.3, p = 0.39$) compared to the base week.

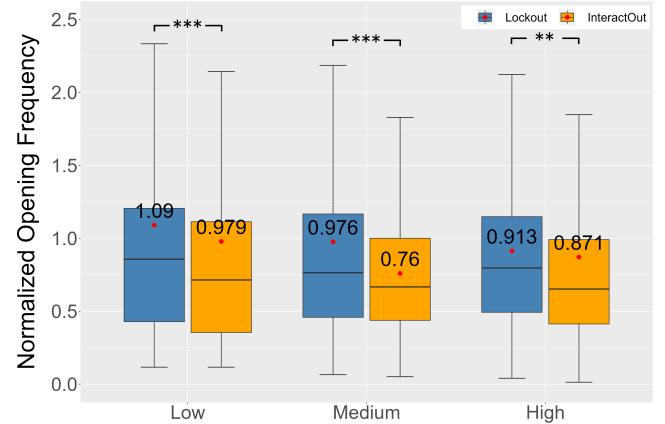


Figure 15: Normalized Target Apps Opening Frequency in 3 Levels of Interaction Intensity (Low, Medium, High). Interaction intensity was categorized based on the number of required touch interactions for each target app. It shows the normalized ratio of the opening frequency against the average opening frequency in the base week in each interaction intensity level group. Lower ratio means more effective reduction. In each box, the middle black line shows the median and the red point shows the mean. Note that the minimum usage time is 0 as some participants do not use their target apps in some days.

There was no significant difference between InteractOut and Lockout ($p = 0.95$ for usage time and $p = 0.97$ for opening frequency).

6.1.4 Results Summary. Overall, our results on app usage behavior indicated that InteractOut was significantly more effective in reducing target apps' usage time and opening frequency than Lockout. It also showed a lasting effect on the opening frequency of target

apps. However, the overall usage across all apps only showed a slight decrease (see Figure 12 and Figure 14).

6.2 Intervention Acceptance Rate

Another important metric of smartphone overuse interventions is their receptivity. We measured the receptivity of the two intervention techniques by their acceptance rate, i.e., the proportion of times participants decided to accept the intervention and did not bypass it. Higher acceptance rate means greater receptivity of the intervention by the users. Note that for both InteractOut and Lockout, the bypass options were similar (see Figure 4 and Figure 9d). **InteractOut showed a significantly higher overall and categorical acceptance rate than Lockout.** Figure 16a shows the overall acceptance rate of the two intervention techniques. InteractOut had an acceptance rate of 61.6%, while Lockout was 36.3%. A Kolmogorov-Smirnov test found that the acceptance rate followed a Gaussian distribution. Thus, we ran a GLMM with Gaussian link function on fixed effects *Intervention*, *Order*, *App_Category*, and *Intervention* × *App_Category*. Results showed the significance of *Intervention* ($F = 5.61, p < 0.05$) and *App_Category* ($F = 2.80, p < 0.05$). The interaction of *Intervention* and *App_Category* also showed a marginal significance ($p = 0.097$). The post-hoc Tukey's HSD test on *Intervention* showed a significantly higher acceptance rate for InteractOut than Lockout ($Z = 7.6, p < 0.001$).

The acceptance rates of each app category are shown in Figure 16b. A post-hoc Tukey's HSD test on *App_Category* did not show a statistical difference among app categories. This indicated that InteractOut outperformed Lockout consistently across different app categories.

6.2.1 Behaviors After Accepting Interventions. We also investigated participants' behaviors after encountering the intervention in a target app. We focused on the immediate events after encountering an intervention (possible events include entering a non-target app, another target app, or closing the screen), and calculated the proportion of each kind of event. Our results showed that for InteractOut, participants entered a non-target app 50.8% of the time, entered another target app 27.1% of the time, and closed the screen 22.1% of the time. In the Lockout condition, entering a non-target app was also the first option (60.1%), followed by closing the screen (19.8%) and entering another target app (18.9%). Overall, participants had similar behaviors after interventions.

6.2.2 Results Summary. Participants showed higher receptivity to InteractOut than Lockout regardless of the categories of their target app choices. But after accepting the intervention, they turned to an intervention-free non-target app most of the time, with a 20% chance to end smartphone usage. This after-acceptance behavior explains the similar usage time and opening frequency increase of overall usage in Section 6.1.

6.3 Subjective Measurements

We also collected participants' subjective feedback through weekly in-study surveys and exit interviews. Thematic analysis was used to analyze the interview data [15], in order to learn about user perceptions of the interventions, including the workload, effectiveness,

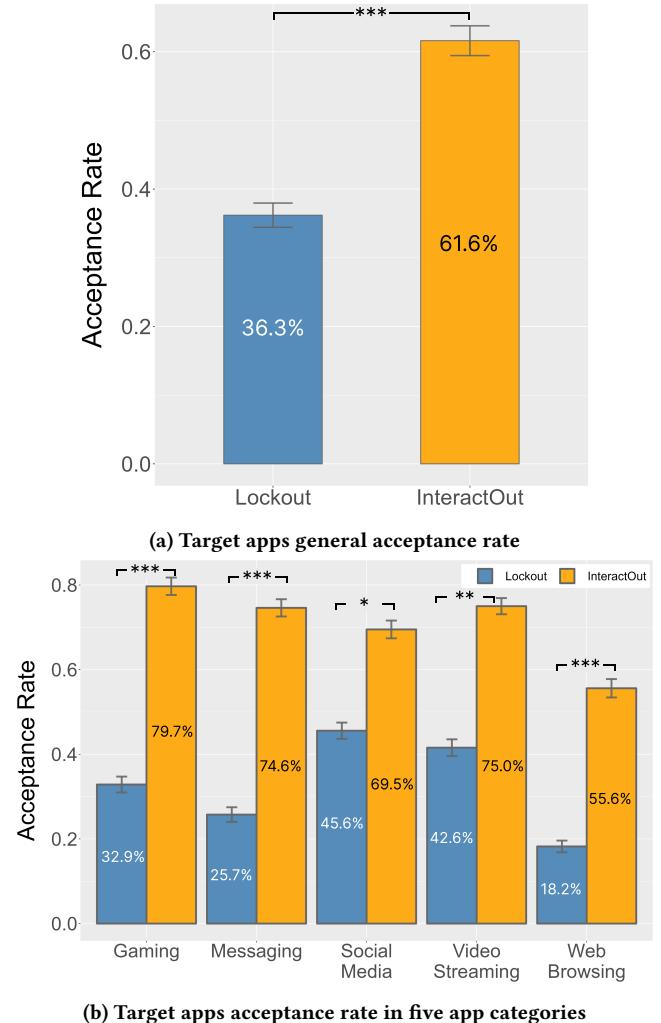


Figure 16: Intervention Acceptance Rate, i.e., the proportion of times that the intervention was not bypassed. Higher rate means greater intervention receptivity. InteractOut outperformed Lockout consistently across different app categories.

and receptivity. Our results showed that InteractOut was more mentally acceptable than Lockout even though InteractOut introduced a higher cognitive load.

6.3.1 Subjective Workload. We measured the subjective workload of the interventions using the NASA TLX questionnaire in our weekly in-study surveys. We asked participants to score the workload of smartphone use under the current condition (i.e., Natural, InteractOut, or Lockout). We summarize the average overall weighted scores and sub-scores in Figure 17. Expectedly, InteractOut had a higher overall workload than Lockout. A GLMM with Gaussian link function confirmed our observation with the significance on *Intervention* ($\chi^2 = 166, p < 0.0001$) but not others ($p_{Order} = 0.61, p_{Intervention \times Order} = 0.64$).

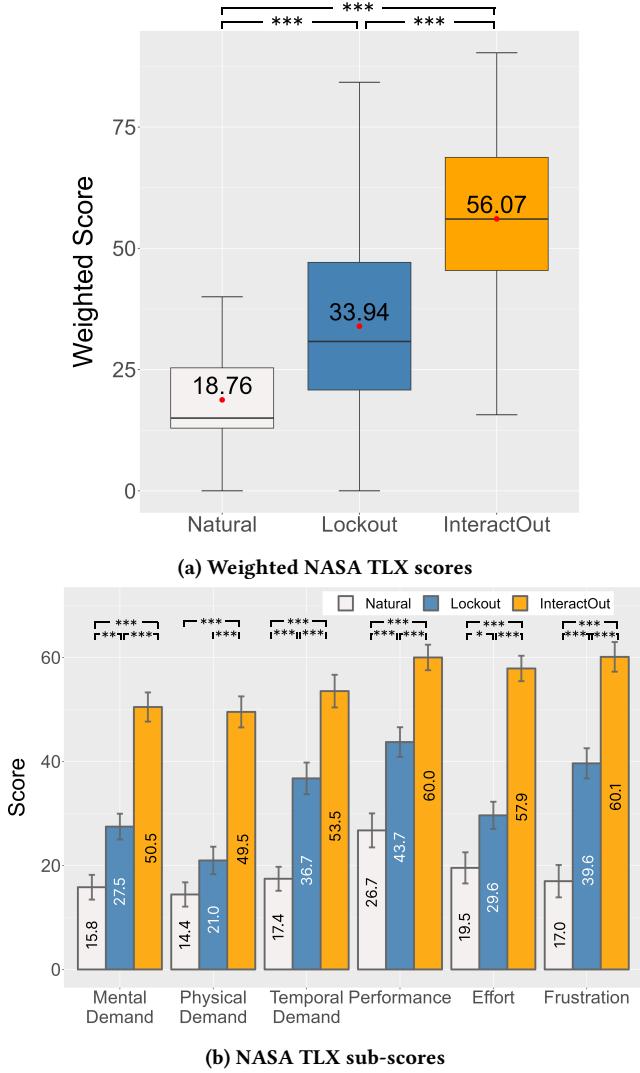


Figure 17: NASA TLX Responses in In-Study Surveys. InteractOut has a significantly higher overall and each individual workload.

6.3.2 Perceived Effectiveness and Receptivity. We summarize the main advantages of InteractOut based on participants' comments.

- **Intervene with appropriate flexibility.** Most participants appreciated the “no-force” intervention InteractOut provided. They were usually in the middle of an activity using the smartphone when encountering the intervention. InteractOut allowed them to finish what they were doing but also acted as a gentle reminder. For example, “[InteractOut] is more comfortable because it allows me to complete my current task such as reading a news article.” (P26) “InteractOut is more gentle. I do not have to snooze. It is more user-friendly and constantly reminds me to put down my phone.” (P35) Participants generally found such soft and implicit interventions to be more acceptable.

- **Provide user agency.** In traditional lockout interventions, users often have no choice but to stop or bypass the limit. Even if they choose to stop, they may feel little agency in making this decision. However, with InteractOut, participants could decide when to stop, which is more natural than a restrictive lockout. This enabled them to finish their current task and have a graceful stop. *“It slows down the app, and I start to think if I want to use the app anymore, then I feel like I do not need it and I will stop.”* (P25) *“(InteractOut) is better as the overlay at least allows me to finish reading the article.”* (P3) The decision-making process was controlled by the users themselves and presented continuously, thus leading to better effectiveness and user experience.
- **Be always-on.** Participants experienced the intervention consistently once they exceeded the time limit. Being always on can act as a constant reminder for users. *“It is constantly there and I am always aware of it.”* (P33) Such a user experience change could sometimes be “misperceived” as a change on the app itself. *“When I use the app [in InteractOut], I found I do not like using the app, whereas the blocking [Lockout] does not change how the app works after I ignore the limit.”* (P21) In this case, participants’ mental model towards the app changed due to constant intervention.
- **Introduce a gulf of gratification.** Although InteractOut’s intervention is implicit, it could introduce a gulf of gratification, meaning users needed to spend more effort to gain entertainment from smartphones. Participants reported that InteractOut introduced gaps in simple operation logic such as scrolling up (e.g., when using social media or video streaming apps). *“[InteractOut] slowing down the scroll speed, [and]... The scrolling is also not real-time... I lose interest of the app.”* (P39) This indicates that manipulating input can effectively trigger System 2 in smartphone overuse scenarios [7, 64].

6.3.3 Critiques. We also received some negative feedback about InteractOut. First, InteractOut made some precise and real-time operations difficult. This was important for non-casual games, such as role-play, first-person shooting, and racing games. For example, P27 picked two games for intervention, but could not play the game with InteractOut. Second, InteractOut was not as effective at intervening in scenarios with sparse interactions, such as watching a long video or reading an e-book. *“In Netflix, there is not too much operation, I cannot sense the intervention.”* (P33) In these cases, participants spent most of the time looking at the screen without performing many gestures, making InteractOut less effective. However, we envision that InteractOut could be combined with other existing intervention techniques. We will discuss some potential cases in Section 7.3.

6.3.4 Mismatch between Real & Perceived Effectiveness. Moreover, we found an interesting mismatch between the real and perceived effectiveness of interventions for some participants. For example, P31 considered InteractOut as ineffective, but the data showed that their usage time on target apps decreased by 17.9% (from 373.1 minutes to 306.1 minutes). A similar mismatch was also observed in four other participants. This suggests that there may exist a subtle and implicit mental effect of input manipulation interventions. Users

may not feel the effectiveness explicitly, yet their objective phone usage behavior has changed. However, the sample of 5 participants for this observation is too small for an in-depth investigation. Future work with a larger-scale deployment could enable us to further investigate this, which we discuss in Section 7.5.

6.3.5 Subjective Results Summary. InteractOut was perceived as being effective and acceptable for most participants due to its subtle and continuous manipulation of input. This property shows a significant advantage over traditional disruptive and restrictive interventions. Users are more likely to develop an active control of reducing smartphone usage, which is one of the favorable and ultimate goals of smartphone overuse intervention. Meanwhile, we acknowledge the disadvantage of InteractOut for scenarios where smartphone usage is more consumption-focused without heavy input interactions.

7 DISCUSSION AND FUTURE WORK

In this work, we propose a set of new input manipulation-based intervention techniques InteractOut. Our lab study showed the effectiveness of these techniques, while highlighting four promising interventions. This pointed us to our field experiment to further evaluate these four interventions' performance in real-world scenarios. Through our lab study and field experiment, we showed the advantage of InteractOut. In this section, we discuss the results from our study, the potential combination of InteractOut and existing intervention methods, the implications for various stakeholders, as well as future directions for larger-scale real-life deployments. We also discuss the limitations of our work.

7.1 Workload and Subjective Preference

As shown in Section 6.3, InteractOut introduced a higher perceived workload for users when operating their smartphones. However, the majority of participants favored InteractOut as their daily smartphone intervention. The primary reason was that InteractOut allowed them to maintain a desirable level of flexibility and control over their smartphone usage. Despite the increased workload, participants were willing to bear additional interaction costs to maintain control. This result aligns with prior research emphasizing the limitations of forceful interventions, which often lead to user abandonment [6].

This finding also shows that InteractOut strikes a delicate balance between intervention effectiveness and user acceptance. However, the current workload of InteractOut may still not be low enough, which require future investigations. The challenge lies in determining the optimal balance at which workload serves as a constructive motivator for behavior change without overwhelming users. This balance between intervention subtlety and user engagement needs further exploration in the development of future smartphone overuse interventions.

7.2 Comparison with Existing Interventions

Although we did not formally compare with additional baselines in our field study, here we compare InteractOut with some existing works using the respective statistics reported in the results sections. Compared to the respective baseline usage without interventions,

InteractOut had a higher reduction (-21.9%) in usage time than LocknType [35] (-7.8%), Good Vibration [50] (-15.7%), and TypeOut [78] (-20.3%). InteractOut also had a higher acceptance rate (61.3%) than TypeOut [78] (58.3%). On the other hand, InteractOut had a lower reduction in opening frequency (-21.7%) than TypeOut [78] (-27.8%). These preliminary comparisons show the potential of InteractOut to have a promising performance compared to prior works, while further study is needed to provide quantitative evidence, and it is difficult to directly compare their results because of the different study populations, durations, and other factors. Future work could also devise benchmarks and guidelines [76, 77] to enable such comparisons across research studies, which is often difficult for human-subject experiments.

7.3 Compatibility and Complementarity with Existing Interventions

In considering the potential applications of InteractOut within the landscape of smartphone overuse interventions, we foresee that it could offer compatibility and complementarity to both explicit and implicit intervention techniques. InteractOut could be seamlessly integrated with existing explicit interventions, such as Timed Lockout. This integration would involve applying InteractOut during the extra time granted after users have reached their predefined smartphone usage limit.

Furthermore, in the realm of implicit interventions, InteractOut exhibits compatibility with reminder-based techniques. This synergy allows for a multi-pronged approach to combining InteractOut's subtle input manipulation and periodic reminders, both aimed at redirecting users' attention away from their devices. Additionally, InteractOut's adaptability extends to other implicit interventions, such as Good Vibrations [50], or potential user interface and user experience manipulations designed to gently discourage excessive smartphone usage.

Since InteractOut mainly focuses on the input aspect, it could serve as a valuable addition to other intervention strategies. Its compatibility with existing techniques and potential for synergy offer researchers, developers, and users a promising pathway to more effectively manage smartphone usage, while maintaining a user-centered approach to intervention design.

7.4 Implications for Users, Platform Owners and App Developers

As highlighted in Section 6.3, some participants perceived InteractOut as an alteration of their usual app interactions. On the one hand, it raises concerns about the potential emergence of a dark pattern [22]. The implicit manipulation of user input could be misused and considered deceptive. Users may perceive this as a violation of their autonomy and control, potentially leading to frustration and mistrust. We also recognize the possibility that users could mis-recognize the intervention as device malfunction. A potential solution is to provide a small indicator, such as a notification or a visual indicator to remind users of the consented intervention. On the other hand, this implicit nature of the intervention could also pose challenges for platform owners like iOS and Android, particularly considering the constraints imposed on third-party apps

accessing accessibility services by platforms like Google [69]. Platform owners and app developers, who often prioritize fine-tuning the user experience to maximize usability and engagement, may also have reservations about interventions that modify their behavior of their systems and apps. InteractOut, in its current form, functions as an external third-party tool, sidestepping the need for explicit buy-in from platform owners and app developers.

However, an avenue for future research lies in investigating the perceptions and reactions of these stakeholders towards interventions like InteractOut. Understanding their perspectives could inform refinements to make the intervention more compatible with platform guidelines and potentially gain direct support within individual apps. Balancing the goals of reducing smartphone overuse with the interests of platform owners and app developers presents an intriguing challenge for the evolution of interventions like InteractOut. Further explorations in this direction could pave the way for more harmonious integration into the broader digital ecosystem.

7.5 Large-Scale Longitudinal Deployments & Future Directions

Both our lab study and field experiment only involved a limited group of users (30 and 42, all university students), with a limited period of time for deployment. Here, we list a few promising directions when large-scale longitudinal deployment is possible.

User-Defined Intervention Configuration and Personalized InteractOut. In our experiment, we pre-determined the time budget and intervention combinations for the purpose of the study. But in real-life deployment, users should have the agency to control the intervention. For example, our config interface (similar to the one shown in Figure 9b) could be extended to enable users to pick their own preferred combination. They could also set their own intensity and the way intensity changes over time. Such a customized configuration could better fit individuals' preferences. Moreover, future work could investigate the potential of making interventions intelligent and adaptive based on AI models to detect risk behavior with sensor data (e.g., [75]), moving toward the vision of just-in-time adaptive intervention (JITAI) [46, 53].

App Feature-Level Interventions. Our current version of InteractOut mainly focuses on intervening at the system-level. It is triggered in a specific set of target apps, but the intervention remains the same across different apps. This sometimes can cause issues, especially when an app is complicated, and not all functions need to be intervened. A beneficial intervention should support meaningful use (i.e. messaging) and limit meaningless use (i.e. endless news/posts feed) [6]. We also believe that the real-perceived effectiveness mismatch mentioned in 6.3.4 could be addressed by adding app feature-level support to InteractOut. In larger-scale deployments, we could extend our exploration to certain apps' specific feature-level interventions [52]. For example, InteractOut may only intervene in the posting and browsing features of a social media app, but not influence its messaging function, since the latter actively involves social communication. We aim to explore how such interventions may be harnessed to augment user receptivity and enhance the overall effectiveness of the intervention strategy.

Input Manipulation beyond Smartphones. Unlike traditional interventions on the screen output, InteractOut manipulates user inputs to create interventions. Although InteractOut is implemented

on smartphones, this idea could be generalized to other devices beyond smartphones with different input modalities. For example, on desktops with keyboard and mouse control, we could design similar delay interventions on mouse pointer movement, double clicking interventions on mouse clicking, and key remapping interventions on the keyboard. On tablets with a mixture of smartphone and desktop operation logic, we could apply either gesture or keyboard interventions when appropriate. We also envision the potential of input manipulations for future devices such as augmented reality and virtual reality (AR/VR). Our InteractOut design space has the potential to be applied and extended to various scenarios to generate more input interventions.

These future research directions with large-scale deployments not only enrich our understanding of the multifaceted potential of InteractOut but also contribute to the ongoing evolution of input manipulation-based intervention methodologies.

7.6 Limitations

There are also some limitations in our work. First, our study population mainly focuses on college students in a local university. The results may not be representative or generalizable to other population groups. Moreover, in the current implementation, participants needed to release their fingers to receive feedback. For swipe interactions, this means that participants could only see the effect of the swipe after the completion of the trajectory. For long swipes, this would cause a recognizable delay. This problem could potentially be resolved in the MotionEvent level [3], which act on each motion event and preserve the fluency of the whole gesture. However, by the time we wrote this paper, we had yet to find an API to provide this fine-grained manipulation. To enable this, Android could provide new APIs to developers to manipulate the global motion events, or incorporate a hardware-level implementation of InteractOut that can directly control feedback of the touchscreen.

8 CONCLUSION

In this paper, we proposed a new perspective on smartphone overuse intervention. We designed a suite of input manipulation-based intervention techniques called InteractOut. Through a pilot lab study we identified four promising InteractOut implementations and their appropriate intensity settings. We then conducted a 5-week field experiment with a traditional Timed Lockout baseline. InteractOut showed 15.6% lower usage time, 16.5% lower open frequency, and 25.3% higher acceptance rate on target apps with a better receptivity. InteractOut suggested the long-term benefit of forceless intervention with a precisely controlled user experience degradation, which should be a joint effort between platform owners, app developers and intervention designers. We envision InteractOut to supplement current smartphone interventions and inspire a wide range of input-based intervention techniques for interactive devices.

ACKNOWLEDGMENTS

We thank our participants for their contribution to our lab and field experiments, and the reviewers for their valuable feedback and suggestions. This research is supported in part by a Google Cloud Platform Credit Award.

REFERENCES

- [1] 2023. AccessibilityService. <https://developer.android.com/reference/android/accessibilityservice/AccessibilityService>
- [2] 2023. Manage how you spend time on your Android phone with Digital Wellbeing - Android Help. <https://support.google.com/android/answer/9346420?hl=en>
- [3] 2023. MotionEvent. <https://developer.android.com/reference/android/view/MotionEvent>
- [4] 2023. Notifications Overview. <https://developer.android.com/develop/ui/views/notifications>
- [5] 2023. Set up Screen Time for yourself on iPhone. <https://support.apple.com/guide/iphone/set-up-screen-time-for-yourself-iphbfa595995/ios>
- [6] 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. <https://doi.org/10.2196/31730>
- [7] Adam L Alter, Daniel M Oppenheimer, Nicholas Epley, and Rebecca N Eyer. 2007. Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of experimental psychology: General* 136, 4 (2007), 569.
- [8] Muhammad Anshari, Yabit Alas, and Exzayrani Sulaiman. 2019. Smartphone addictions and nomophobia among youth. *Vulnerable Children and Youth Studies* 14, 3 (2019), 242–247. <https://doi.org/10.1080/17450128.2019.1614709>
- [9] Forest APP. 2023. *Stay focused, be present*. <https://www.forestapp.cc/>
- [10] Apple. 2023. *Shortcuts*. <https://support.apple.com/guide/shortcuts/welcome/ios>
- [11] Aswin Mohan. 2023. *LessPhone*. <https://play.google.com/store/apps/details?id=me.aswinmohan.nophone>
- [12] Fayiga Ahmed Bakir and Srinivasan Subramanian Grandee. 2020. Impact of the COVID-19 lockdown on digital device-related ocular health. *Indian journal of ophthalmology* 68, 11 (2020), 2378.
- [13] Sebastian Boring, David Ledo, Xiang 'Anthony' Chen, Nicolai Marquardt, Anthony Tang, and Saul Greenberg. 2012. The Fat Thumb: Using the Thumb's Contact Size for Single-Handed Mobile Interaction. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services* (San Francisco, California, USA) (*MobileHCI '12*). Association for Computing Machinery, New York, NY, USA, 39–48. <https://doi.org/10.1145/2371574.2371582>
- [14] Andrew Bragdon, Eugene Nelson, Yang Li, and Ken Hinckley. 2011. Experimental Analysis of Touch-Screen Gesture Designs in Mobile Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). Association for Computing Machinery, New York, NY, USA, 403–412. <https://doi.org/10.1145/1978942.1979000>
- [15] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [16] Nancy A. Cheever, Larry D. Rosen, L. Mark Carrier, and Amber Chavez. 2014. Out of sight is not out of mind: The impact of restricting wireless mobile device use on anxiety levels among low, moderate and high users. *Computers in Human Behavior* 37 (2014), 290–297. <https://doi.org/10.1016/j.chb.2014.05.002>
- [17] Hyunsung Cho, DaEun Choi, Donghwi Kim, Wan Ju Kang, Eun Young 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.
- [18] Xiang Ding, Jing Xu, Guanling Chen, and Chenren Xu. 2016. Beyond Smartphone Overuse: Identifying Addictive Mobile Apps. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI EA '16*). Association for Computing Machinery, New York, NY, USA, 2821–2828. <https://doi.org/10.1145/2851581.2892415>
- [19] Jonathan S BT Evans and Keith E Stanovich. 2013. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science* 8, 3 (2013), 223–241.
- [20] Sónia P. Gonçalves and Joana Vieira dos Santos. 2022. Smartphone Use Side-by-Side with Burnout: Mediation of Work-Family Interaction and Loneliness. *International Journal of Environmental Research and Public Health* 19, 11 (2022). <https://doi.org/10.3390/ijerph19116692>
- [21] Google. 2023. Firebase. <https://firebase.google.com/>
- [22] Colin M Gray, Kubilay 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*, 1–14.
- [23] Joshua Harwood, Julian J. Dooley, Adrian J. Scott, and Richard Joiner. 2014. Constantly connected – The effects of smart-devices on mental health. *Computers in Human Behavior* 34 (2014), 267–272. <https://doi.org/10.1016/j.chb.2014.02.006>
- [24] Alexis Hiniker, Sungsoo (Ray) 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* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 4746–4757. <https://doi.org/10.1145/2858036.2858403>
- [25] Daniel Hintze, Philipp Hintze, Rainhard D Findling, and René Mayrhofer. 2017. A large-scale, long-term analysis of mobile device usage characteristics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 1–21.
- [26] Wilhelm Hofmann, Malte Fries, and Fritz Strack. 2009. Impulse and Self-Control From a Dual-Systems Perspective. *Perspectives on Psychological Science* 4, 2 (March 2009), 162–176. <https://doi.org/10.1111/j.1745-6924.2009.01116.x>
- [27] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. 1985. Direct Manipulation Interfaces. *Human–Computer Interaction* 1, 4 (1985), 311–338. https://doi.org/10.1207/s15327051hci0104_2
- [28] IFTTT. 2023. *IFTTT*. <https://ifttt.com/>
- [29] iMobiApp. 2023. *RealizD*. <https://www.realizd.com/>
- [30] The NASA TLX Tool: Task Load Index. 2019. *TLX @ NASA Ames - Home*. <https://humansystems.arc.nasa.gov/groups/tlx/>
- [31] Reynol Junco and Sheila R. Cotten. 2011. Perceived Academic Effects of Instant Messaging Use. *Comput. Educ.* 56, 2 (feb 2011), 370–378. <https://doi.org/10.1016/j.compedu.2010.08.020>
- [32] Inyeop Kim, Gyuwon Jung, Hayoung Jung, Minsam Ko, and Uichin Lee. 2017. Let's FOCUS: Mitigating Mobile Phone Use in College Classrooms. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 63 (sep 2017), 29 pages. <https://doi.org/10.1145/3130928>
- [33] Jaejeung Kim, Chiwoo Cho, and Uichin Lee. 2017. Technology Supported Behavior Restriction for Mitigating Self-Interruptions in Multi-Device Environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 64 (sep 2017), 21 pages. <https://doi.org/10.1145/3130932>
- [34] Jaejeung Kim, Hayoung Jung, Minsam Ko, and Uichin Lee. 2019. GoalKeeper: Exploring Interaction Lockout Mechanisms for Regulating Smartphone Use. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1, Article 16 (mar 2019), 29 pages. <https://doi.org/10.1145/3314403>
- [35] Jaejeung Kim, Joonyoung Park, Hyunsoo Lee, Minsam Ko, and Uichin Lee. 2019. LocknType: Lockout Task Intervention for Discouraging Smartphone App Use. 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. <https://doi.org/10.1145/3290605.3300927>
- [36] Minsam Ko, Seungwoo Choi, Koji Yatani, and Uichin Lee. 2016. Lock n' LoL: Group-Based Limiting Assistance App to Mitigate Smartphone Distractions in Group Activities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 998–1010. <https://doi.org/10.1145/2858036.2858568>
- [37] Minsam Ko, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Dahee Shin, Koji Yatani, Junehwa Song, and Kyong-Mee Chung. 2015. NUGU: A Group-Based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) (*CSCW '15*). Association for Computing Machinery, New York, NY, USA, 1235–1245. <https://doi.org/10.1145/2675133.2675244>
- [38] Min Kwon, Dai-Jin Kim, Hyun Cho, and Soo Yang. 2014. The Smartphone Addiction Scale: Development and Validation of a Short Version for Adolescents. *PLOS ONE* 8, 12 (12 2014), 1–7. <https://doi.org/10.1371/journal.pone.0083558>
- [39] Toby Jia-Jun Li, Amos Azaria, and Brad A. Myers. 2017. SUGILITE: Creating Multimodal Smartphone Automation by Demonstration. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (*CHI '17*). Association for Computing Machinery, New York, NY, USA, 6038–6049. <https://doi.org/10.1145/3025453.3025483>
- [40] Hajin Lim, Ian Arawao, Yaxian Xie, Negar Khojasteh, and Susan R Fussell. 2017. Distraction or life saver? The role of technology in undergraduate students' boundary management strategies. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–18.
- [41] Gadi Lissak. 2018. Adverse physiological and psychological effects of screen time on children and adolescents: Literature review and case study. *Environmental Research* 164 (2018), 149–157. <https://doi.org/10.1016/j.envres.2018.01.015>
- [42] Markus Löchtefeld, Matthias Böhmer, and Lyubomir Ganev. 2013. AppDetox: Helping Users with Mobile App Addiction. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (Luleå, Sweden) (*MUM '13*). Association for Computing Machinery, New York, NY, USA, Article 43, 2 pages. <https://doi.org/10.1145/2541831.2541870>
- [43] Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What makes smartphone use meaningful or meaningless? *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–26.
- [44] Ulrik Lyngs, Kai Lukoff, Petr Slovák, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt. 2019. Self-control in cyberspace: Applying dual systems theory to a review of digital self-control tools. In *proceedings of the 2019 CHI conference on human factors in computing systems*, 1–18.
- [45] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2017. Just-in-Time Adaptive Interventions (JTAs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine* 52, 6 (12 2017), 446–462. <https://doi.org/10.1007/s12160-016-9830-8>
- [46] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2018. Just-in-time adaptive interventions (JTAs) in mobile health: key components and design principles

- for ongoing health behavior support. *Annals of Behavioral Medicine* (2018), 1–17.
- [47] Chauncey J Neyman. 2017. A survey of addictive software design. (2017).
- [48] Jakob Nielsen. 1993. Response Times: The 3 Important Limits. <https://www.nngroup.com/articles/response-times-3-important-limits/>
- [49] Donald A Norman and Jakob Nielsen. 2010. Gestural interfaces: a step backward in usability. *interactions* 17, 5 (2010), 46–49.
- [50] 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* (Barcelona, Spain) (*MobileHCI '18*). Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages. <https://doi.org/10.1145/3229434.3229463>
- [51] Fabian Okeke, Michael Sobolev, and Deborah Estrin. 2018. Towards A Framework for Mobile Behavior Change Research. In *Proceedings of the Technology, Mind, and Society* (Washington, DC, USA) (*TechMindSociety '18*). Association for Computing Machinery, New York, NY, USA, Article 27, 6 pages. <https://doi.org/10.1145/3183654.3183706>
- [52] 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 (2023), 1–30.
- [53] 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 2024 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–18. <https://doi.org/10.1145/3613904.3642747>
- [54] 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* (Yokohama, Japan) (*CHI '21*). Association for Computing Machinery, New York, NY, USA, Article 702, 17 pages. <https://doi.org/10.1145/3411764.3445489>
- [55] Joonyoung Park, Jin Yong Sim, Jaejeung Kim, Mun Yong Yi, and Uichin Lee. 2018. Interaction Restraint: Enforcing Adaptive Cognitive Tasks to Restrain Problematic User Interaction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI EA '18*). Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3170427.3188613>
- [56] Martin Pielot and Llu Rello. 2017. Productive, Anxious, Lonely: 24 Hours without Push Notifications. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) (*MobileHCI '17*). Association for Computing Machinery, New York, NY, USA, Article 11, 11 pages. <https://doi.org/10.1145/3098279.3098526>
- [57] Benjamin Poppinga, Alireza Sahami Shirazi, Niels Henze, Wilko Heuten, and Susanne Boll. 2014. Understanding Shortcut Gestures on Mobile Touch Devices. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (Toronto, ON, Canada) (*MobileHCI '14*). Association for Computing Machinery, New York, NY, USA, 173–182. <https://doi.org/10.1145/2628363.2628378>
- [58] James O. Prochaska and Wayne F. Velicer. 1997. The Transtheoretical Model of Health Behavior Change. *American Journal of Health Promotion* 12, 1 (1997), 38–48. <https://doi.org/10.4278/0890-1171-12.1.38> PMID: 10170434.
- [59] Julie Rico and Stephen Brewster. 2010. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 887–896.
- [60] Sensor Tower. 2023. *ActionDash*. <https://actiondash.com/>
- [61] Claude M. Steele. 1988. The Psychology of Self-Affirmation: Sustaining the Integrity of the Self. *Advances in Experimental Social Psychology*, Vol. 21. Academic Press, 261–302. [https://doi.org/10.1016/S0065-2601\(08\)60229-4](https://doi.org/10.1016/S0065-2601(08)60229-4)
- [62] Matthew J Stollak, Amy Vandenberg, Andie Burklund, and Stephanie Weiss. 2011. Getting Social: The Impact of Social Networking Usage on Grades Among College Students. In *Proceedings from ASBBS annual conference*, Vol. 18. 859–865.
- [63] Storm8 Studios. 2023. *Bubble Mania*. <https://play.google.com/store/apps/details?id=com.teamlavabubble>
- [64] Nassim Nicholas Taleb. 2014. *Antifragile: Things that gain from disorder*. Vol. 3. Random House Trade Paperbacks.
- [65] Sandra Tor and Simon Ekman von Huth. 2021. The Impact of Gesture Navigation on Mobile Usage.
- [66] M. K. Umami, P. Arezes, and A. M. Sampaio. 2015. Identifying finger postures when interacting with the touch screen of mobile devices.
- [67] Chad J Valasek. 2022. Disciplining the Akratic user: Constructing digital (un) wellness. *Mobile Media & Communication* 10, 2 (2022), 235–250.
- [68] Deedra Vargo, Lin Zhu, Briana Benwell, and Zheng Yan. 2021. Digital technology use during COVID-19 pandemic: A rapid review. *Human Behavior and Emerging Technologies* 3, 1 (2021), 13–24. <https://doi.org/10.1002/hbe2.242>
- [69] Manuel Vonau. 2022. Android 13 blocks unwarranted use of accessibility services by further restricting sideloaded apps. <https://www.androidpolice.com/android-13-blocks-accessibility-services-sideloaded-apps/>
- [70] Brooke E. Wagner, Amanda L. Folk, Samantha L. Hahn, Daheja J. Barr-Anderson, Nicole Larson, and Dianne Neumark-Sztainer. 2021. Recreational Screen Time Behaviors during the COVID-19 Pandemic in the U.S.: A Mixed-Methods Study among a Diverse Population-Based Sample of Emerging Adults. *International Journal of Environmental Research and Public Health* 18, 9 (2021). <https://doi.org/10.3390/ijerph18094613>
- [71] Anqi Wang, Zhen Wang, Ya Zhu, and Xuliang Shi. 2022. The Prevalence and Psychosocial Factors of Problematic Smartphone Use Among Chinese College Students: A Three-Wave Longitudinal Study. *Frontiers in Psychology* (2022), 1476. <https://doi.org/10.3389/fpsyg.2022.877277>
- [72] Chee Wai Wong, Andrew Tsai, Jost B. Jonas, Kyoko Ohno-Matsui, James Chen, Marcus Ang, and Daniel Shu Wei Ting. 2021. Digital Screen Time During the COVID-19 Pandemic: Risk for a Further Myopia Boom? *American Journal of Ophthalmology* 223 (2021), 333–337. <https://doi.org/10.1016/j.ajo.2020.07.034>
- [73] Ruolan Wu, Chun Yu, Xiaole Pan, Yujia Liu, Ningning Zhang, Yue Fu, Wang Yuhan, Zheng Zhi, Chen Li, Jiang Qiaolei, Xuhai Xu, and Yuanchun Shi. 2024. MindShift: Leveraging Large Language Models for Mental-States-Based Problematic Smartphone Use Intervention. In *Proceedings of the 2024 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–18. <https://doi.org/10.1145/3613904.3642790>
- [74] X. 2023. X. <https://twitter.com>
- [75] Xuhai Xu, Prerna Chikarsal, Afsaneh Doryab, Daniella K Villalba, Janine M Dutcher, Michael J Tumminia, Tim Althoff, Sheldon Cohen, Kasey G Creswell, J David Creswell, et al. 2019. Leveraging routine behavior and contextually-filtered features for depression detection among college students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–33.
- [76] Xuhai Xu, Xin Liu, Han Zhang, Weichen Wang, Subigya Nepal, Yasaman Sefidgar, Woosuk Seo, Kevin S Kuehn, Jeremy F Huckins, Margaret E Morris, et al. 2023. GLOBEM: Cross-Dataset Generalization of Longitudinal Human Behavior Modeling. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–34.
- [77] Xuhai Xu, Han Zhang, Yasaman Sefidgar, Yiyi Ren, Xin Liu, Woosuk Seo, Jennifer Brown, Kevin Kuehn, Mike Merrill, Paula Nurius, Shwetak Patel, Tim Althoff, Margaret Morris, Eve Riskin, Jennifer Mankoff, and Anind Dey. 2022. GLOBEM Dataset: Multi-Year Datasets for Longitudinal Human Behavior Modeling Generalization. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 24655–24692. https://proceedings.neurips.cc/paper_files/paper/2022/file/9c7e8a0821dfcb58a9a33cbd37cc8131-Paper-Datasets_and_Benchmarks.pdf
- [78] Xuhai Xu, Tianyuan Zou, Han Xiao, Yanzhang Li, Ruolin Wang, Tianyi Yuan, Yuntao Wang, Yuanchun Shi, Jennifer Mankoff, and Anind K Dey. 2022. TypeOut: Leveraging Just-in-Time Self-Affirmation for Smartphone Overuse Reduction. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (*CHI '22*). Association for Computing Machinery, New York, NY, USA, Article 442, 17 pages. <https://doi.org/10.1145/3491102.3517476>
- [79] Zafy Intelligence. 2023. *AntiSocial*. <https://antisocial.io/>
- [80] Xiaoyi Zhang, Anne Spencer Ross, Anat Caspi, James Fogarty, and Jacob O. Wobbrock. 2017. Interaction Proxies for Runtime Repair and Enhancement of Mobile Application Accessibility. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (*CHI '17*). Association for Computing Machinery, New York, NY, USA, 6024–6037. <https://doi.org/10.1145/3025453.3025846>