A Mixed-Method Exploration into the Mobile Phone Rabbit Hole

NAĐA TERZIMEHIĆ*, LMU Munich, Germany
FLORIAN BEMMANN*, LMU Munich, Germany
MIRIAM HALSNER, LMU Munich, Germany
SVEN MAYER, LMU Munich, Germany

Smartphones provide various functions supporting users in their daily lives. However, the temptation of getting distracted and tuning out is high leading to so-called rabbit holes. To quantify rabbit hole behavior, we developed an Android tracking application that collects smartphone usage enriched with experience sampling questionnaires. We analyzed 14,395 smartphone use sessions from 21 participants, collected over two weeks, showing that rabbit hole sessions are significantly longer and contain more user interaction, revealing a certain level of restlessness in use. The context of rabbit hole sessions and subjective results revealed different triggers for spending more time on the phone. Next, we conduct an expert focus group (N=6) to put the gained insights into perspective and formulate a definition of the mobile phone rabbit hole. Our results form the foundation for predicting and communicating the mobile phone rabbit hole, especially when prolonged smartphone use results in regret.

CCS Concepts: • **Human-centered computing** → **Smartphones**; *Mobile computing*.

Additional Key Words and Phrases: human computer interaction, smartphone use, experience sampling, digital wellbeing, mobile phone rabbit hole

ACM Reference Format:
Nađa Terzimehić, Florian Bemmann, Miriam Halsner, and Sven Mayer. 2023. A Mixed-Method Exploration into the Mobile Phone Rabbit Hole. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 194 (September 2023), 29 pages. https://doi.org/10.1145/3604241

1 INTRODUCTION

In Carroll's [9] *"Alice's Adventures in Wonderland,"* a girl named Alice follows a strange humanoid rabbit and subsequently falls down a hole – coining the term *falling into a rabbit hole*. More than a century and a half after publishing the book, the expression *rabbit hole* continuously pops up in conversations especially revolving around one topic: when sharing experiences on using online digital technology (e.g., [12, 17, 38, 45, 72, 79]). In this context, the rabbit hole describes rather over-the-top and, at times, prolonged digital content consumption compared to the user's initial intention [12, 45]. Given that the smartphone has become an ever-present companion, it is now feasible to, accidentally or not, drop into a digital rabbit hole at any given time and place. Until

*Both authors contributed equally to this research.

Authors' addresses: Nađa Terzimehić, LMU Munich, Frauenlobstr. 7a, Munich, 80337, Germany, nadja.terzimehic@ifi.lmu.de; Florian Bemmann, LMU Munich, Frauenlobstr. 7a, Munich, 80337, Germany, florian.bemmann@ifi.lmu.de; Miriam Halsner, LMU Munich, Frauenlobstr. 7a, Munich, Germany, miriam.halsner@campus.lmu.de; Sven Mayer, LMU Munich, Frauenlobstr. 7a, Munich, 80337, Germany, info@sven-mayer.com.

2573-0142/2023/9-ART194 $15.00
https://doi.org/10.1145/3604241

now, both research (e.g., [45, 46]) and society (e.g., [14, 72, 78]) have mostly debated the negative rabbit hole-like effects, in particular in the digital well-being research area (e.g., [64, 77]); however, the HCI community has yet to define and understand the term *mobile phone rabbit hole* (MPRH). Recently, Cho et al. [12] considered an MPRH to be *"the act of deviation from the original purpose of use,"* caused by *"following a continuous chain of viewing just a bit more."* We start exploring this understanding to formalize and expand our knowledge of MPRHs. For example, how long does a typical MPRH session last? By what apps, websites, or contexts are they triggered? What is the user's emotional state when drawn into an MPRH? What everyday life contexts are particularly prone to MPRH? How do MPRHs differ compared to *"normal,"* that is, non-rabbit hole use sessions?

In this paper, we perform a mixed-method analysis of smartphone sessions and behaviors to address the above questions. First, we ran a two-week in-the-wild study ($N = 21$) collecting 14,395 labeled smartphone sessions. In detail, we develop an Android app that tracks quantitative users' mobile phone data and utilizes experience sampling to collect situations where smartphone use deviates from its intended purpose, as defined by [12]. The experience sampling probes the user's intended smartphone use goal at unlock and lock; we ask whether the users fulfilled their intention and whether additional smartphone activities followed. To gain a more nuanced, context-aware picture of the MPRH, we asked users after smartphone use about their perceived awareness of their surroundings, sense of agency, and emotional state. While we derived our in-the-wild study from related work, we uncovered unexpected patterns we could not connect to prior findings. Thus, we opted to explore a broader definition of MPRH and took a step back by performing an expert focus group ($N = 6$). We aimed to understand the MPRHs more deeply from a qualitative perspective and ways of communicating an on-going MPRH session, particularly in potential prevention and intervention scenarios.

Our quantitative analysis of the in-the-wild study shows a longer duration of MPRH-sessions compared to non-MPRH sessions, increased gaming, and visual entertainment app usage, and more frequent home screen visits. Furthermore, MPRH sessions occurred more often in the evening and in relatively steady situations. We found that in at least 17% of instances, users used their phones without a concrete intention to use; following Cho et al. [12], these are not MPRH sessions. On the other hand, when a user gets triggered by, e.g., notifications to engage in other activities, the sessions get labeled as MPRH even though the deviating activity might be productive (e.g., doing work emails). As our in-the-wild study uncovered these patterns, which seem to conflict with the idea of MPRH, we conducted a focus group to revisit the initial definition. The focus group yielded 1) a more broad definition of the MPRH, including triggers, implications, and differentiation between positive and negative MPRHs, and 2) a set of design guidelines to prevent and intervene in negative MPRHs, where applicable.

We primarily contribute to the digital well-being research area [77] in deeply understanding behaviors within MPRH sessions, allowing future research to implement potential prediction and mitigation strategies. Our insights reveal key factors which describe the user behavior when the user falls into the rabbit hole. Finally, we reflect on the definition of the MPRH and discuss UI implications on how to communicate MPRHs, both in an intervening and preventive way.

2 RELATED WORK

Our work is strongly rooted in the digital well-being area [77], as our aim is to understand and eventually predict a negative smartphone use pattern for the sake of improving user's well-being.

2.1 Understanding the Digital Rabbit Hole

In this work, we aim to understand and inspect the *mobile phone rabbit hole* more closely. Previous research investigated the rabbit hole in other content and device contexts. Fox et al. [22] explored the rabbit hole phenomena with the goal of informing information search tools. Piccardi et al. [59] investigated long Wikipedia reading sessions independent of the device context. Their results suggest that users are more likely to fall into a rabbit hole starting from articles about entertainment, sports, politics, and history – and tend to stay focused on one topic in a rabbit hole. Staying within a similar topic, but also jumping to a completely new topic applies to watching YouTube videos [82]. Anecdotes or blog-based definitions of digital rabbit holes (e.g., [14, 72, 78]) refer to rather downsides of an MPRH: time waste, the disability to enjoy (a boring) reality or the easiness of falling down a spiral of misinformation, or even harmful behaviors (e.g., right extremism [31], body dysmorphia [25]). Yet, Wigmore [79] emphasize the positives of the journey down the rabbit hole, stating that *"the [rabbit hole] path often leads to serendipitous discoveries. [...] the meandering path may eventually turn out to be more productive than a more direct one."* The article by Schulz [68] differs between three categories of rabbit holes: 1) *incremental*, which refers to a sequence of distractions that begin with a specific intention, get interrupted by minor distractions (such as receiving a text-message), and eventually end up consuming more time than intended; 2) *exhaustive*, which occurs when the user sets out with the goal of learning about a particular topic but ends up accumulating a vast amount of information that is difficult to comprehend; and 3) *associative*, which happens when the user searches for one thing, ends up finding something remotely related, then gets sidetracked again by something even more tangential, and so on. In the context of smartphones, Cho et al. [12] describe the smartphone rabbit hole as *"the act of deviation from the original purpose of use"*, as a result of *"following a continuous chain of viewing just a bit more"*. The listed definitions and conceptualizations of the MPRH share the commonality of causing the user to deviate from their initial purpose or intention for using their digital devices. They all involve a sequence of distractions that lead the user down a path of unintended usage, resulting in an extended duration of use beyond what was originally intended. We thus start our exploration by following the understanding of an MPRH by Cho et al. [12], being one of the rare works in HCI to pin down an understanding of the MPRH-phenomena. As such, our aim is to set a common ground for understanding the expression *"falling down a rabbit hole"* in HCI, in particular when engaged in smartphone interaction.

2.2 Prolonged Smartphone Use

Expressions such as mobile engagement- [19], compulsion- [39], dopamine- or dopamine-driven-feedback-loops [26, 80] describe the usage loop of endlessly scrolling through content on various social media apps or news feeds. The search for information, the anticipation of finding interesting content, or the rewards in the form of likes or reactions keep the users scrolling for a long time. One reason for such behavior can be procrastination, described as the voluntary delay of urgent tasks, which may result in negative consequences [1, 40]. Prolonged use might also be a self-control failure [54], i.e., when people fail to resist a temptation to use their smartphones while having other important obligations at hand [16]. In this case, a 5-minutes social media break can easily turn into prolonged usage and interrupt the user's other tasks – due to either user's lack of motivation or their inability to stop themselves from consuming more media [11, 66]. This signals that short-term rewards are preferred over longer-term gratifications [57]. The gratifications users seek from phone use can be divided into instrumental and ritualistic or habitual use [28]. Instrumental use describes using the smartphone to fulfill a specific purpose. On the other hand, ritualistic use describes aimlessly exploring content, e.g., out of boredom or to pass time. Sub-terms of ritualistic use include compulsive or habitual phone use [28, 36, 56, 75]. Compulsive or habitual use can be defined as *"a brief, repetitive inspection of a dynamic content quickly accessible on the device"* [56], e.g., when users pull out their phone to check in on different apps or messages [75]. Likewise, communication with others via the smartphone can become a contributor to habitual smartphone use [7, 56].

Habitual phone use might induce regret [12], e.g., when the habitual checking does not yield positive rewards or when users feel distracted from their original intent. Similarly, absent-minded smartphone use [49]) refers to frequent use of the phone without a specific purpose, which may include habitually checking the phone and thus compulsively making calls, as well as endlessly scrolling through content or indiscriminately exploring or switching apps.

According to Reinecke and Hofmann [63], a person's negative emotional and physical state (e.g., stress or being tired) can lead towards recreational smartphone use and media consumption. Boredom or a lack of stimulation can decrease productivity and motivate individuals to use their phones for distraction [56]. People may also use their phones to escape negative emotions [46] or stressful situations [32, 33].

Furthermore, today's mobile applications are designed to keep the user's attention [53], to encourage immersion and continual interaction [1, 18, 44, 81]. Monge Roffarello and De Russis [53] discuss *"attention-capture dark design patterns"* (e.g., notifications that alert the user of new rewarding information [3, 56, 66]; post-play or autoplay functions and algorithmic curation [29]; recommendations, click-bait or infinite scrolling [45, 86]) to distract user's from their initial goal. This can undermine user's sense of agency and lead to a perceived sense of loss of time and surroundings, and ultimately regret. An exploration by Yan et al. [84] found common chains-of-use of apps, with notifications triggering these chains. All of this can contribute to behavioral patterns, such as fragmentation and habituation [3], which can build an easier pathway for the user to indulge into prolonged use.

Finally, the user's current context can also influence the likelihood of prolonged use. Contexts such as the time of the day, user's current location [20] or performed activities [41, 51] can highly influence user's smartphone behavior. When users experience downtime or a phase of loss of motivation in these contexts, they turn towards their phones [75] seeking rewards. Further triggers include unoccupied time, a tedious task, social awkwardness, an expectation of social or informational reward [75]. However, at 89% of the time [27], there is simply no external reason that leads users to pick up and unlock their phones.

Summarizing the current efforts on prolonged smartphone use, we seek to understand and detect the MPRH by taking into account several contextual information and usage patterns to expand the presented existing research.

2.3 Quantitatively Understanding and Predicting Smartphone Use Behaviors

Predicting different smartphone use patterns can help improve the user experience and help developers design better mobile apps, for example, to build more meaningful applications [46] or create a smoother user experience for users [69, 83, 84]. Cao and Lin [8] reported an extended review of works that mine smartphone usage data and predict different app usage patterns.

A recent study [65] aimed to create a better digital wellbeing app by extracting smartphone habit patterns for individual users. They used a clustering algorithm with extracted features like time, app usage, notifications, physical activity, and location. This helped discover complex smartphone habits such as context habits, application habits, and app context habits, which were applied to their digital wellbeing app *Socialize*, effectively assisting smartphone users in reducing their unwanted smartphone use with more awareness. Similarly, Do and Gatica-Perez [15] extracted ten representative daily smartphone usage patterns with a bag-of-apps model.

Hiniker et al. [28] predict ritualistic or instrumental phone use based on the gratification theory. Their model is a combination of Decision Trees, and Naïve-Bayes can correctly classify the use types with an accuracy of 77% and up to 97% with a sliding confidence threshold. They found that users spent more time on their phones when seeking ritualistic gratification, and users which seek ritualistic gratifications are more likely to browse social networks or play mobile games.

Shin and Dey [70] explored specific users that exhibit problematic phone use. The most relevant features for indicating overuse were the number of apps used per day or session, or the length of non-event-initiated sessions. Their unsupervised learning model could detect problematic phone use with an 89.6% accuracy.

Pielot et al. [62] investigated if boredom can be detected in users and examined if boredom affects phone use. With the most important features being the recency of communication, the intensity of recent use, and the general usage intensity, they developed a user-independent machine learning model that can infer boredom with an accuracy of 82.9%. They found that the state of boredom correlates with longer phone use and that boredom is perceived differently for different demographics. On the topic of boredom, Matic et al. [50] developed a machine learning model that classifies users' smartphone usage into high or low boredom proneness with over 80% accuracy.

Yan et al. [84] investigated which apps tend to trigger sessions and which apps are more likely to be follower apps in a session, finding that communication and social media notifications, as well as the Web Browser are among the most popular triggers across all users. Huang et al. [30] used contextual features such as time, location, user profile, or last used app to predict the next applications and found strong dependency between apps. Another approach is the deep reinforcement learning framework called *DeepAPP*. This framework is context-aware and a general network that is then personalized to optimize predicting the next application the user will choose [69, 83]. Baeza-Yates et al. [2] approached the problem of predicting the next app as a personalized classification problem with features regarding usage sessions and app usage patterns. We contribute to the listed body of research by developing our predictor of the MPRH, for preventive and interventive scenarios.

2.4 Research Questions

Based on the listed related work, we extract our two main research questions. Cho et al. [12] described a general understanding of MPRH sessions. However, specific triggers that would allow us to quantify the sessions are unclear yet, restricting us from building a classifier for in-the-wild deployment. To overcome this issue, we formulate the first research question as:

RQ1 How do MPRH sessions differ from non-MPRH smartphone use sessions, and what are the potential factors that contribute to this difference?

While this helps in building prediction models to counteract potential MPRH sessions actively, we still lack a contextual understanding of the external factors that contribute to MPRH sessions occurring in the first place. Such an understanding allows addressing challenges of MPRH sessions before they even occur, rather than mitigating them while they are happening or have already happened. Thus, our second research question is as follows:

RQ2 Are there any particular everyday life contexts or activities that are more prone to causing MPRH sessions, and if so, how can we design interventions to prevent or mitigate them?

3 IN-THE-WILD *RABBIT HOLE* COLLECTION STUDY

We conducted an in-the-wild study to collect labeled data allowing us to understand *mobile phone rabbit hole* behavior. Thus, the study targets two goals: 1) understanding *rabbit hole* behavior from a data perspective, and 2) enabling predicting *rabbit hole* behavior. For this, we used our new *Rabbit-Hole-Tracker* (RHT) app. The app collects user's smartphone use data and the context of smartphone use, such as time and location via WiFi network. Moreover, we collected labels for the different sessions with the app to annotate if the participants classified the smartphone session as *rabbit hole* session using the experience sampling method (ESM).

Table 1. Overview of the different usage event types collected with the *rabbit holeTracker*. All data is recorded with timestamps.

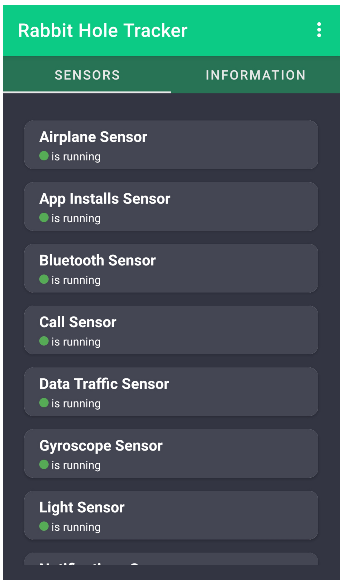
Sensor	Description	Sampling Rate	Example Ref.
Smartphone Sensors			
Accelerometer	The current values of the accelerometer of the phone	On Sensor Change	[7, 21, 35, 52, 60]
Gyroscope	The current values of the gyroscope of the phone	On Sensor Change	[52, 60]
Proximity	Proximity to the phone	On Sensor Change	[21, 50, 60, 61]
Light	Ambient light level as provided by the phone light sensor	On Sensor Change	[35, 52]
Usage Events			
Foreground App	Information on the currently open application	On Event Received	[7, 28, 34, 35, 42, 46, 50, 51, 58, 61, 65, 70, 76, 87]
Accessibility Event	Click and scroll events and screen changes	On Event Received	[70]
Android Usage Event	E.g., App changes, unlocks, reboot	750ms Sampling Rate	[50, 51, 61, 65, 76, 76]
Internet URLs	Visited URLs in the browser	On Event Received	[42]
Data usage	Transmitted and received bytes of a mobile or WLAN internet connection	750ms Sampling Rate	[34]
Notifications	Posted or removed notification with priority level	On Event Received	[35, 50, 51, 61, 65]
Phone State			
Airplane Mode	Whether the phone is in airplane mode	750ms Sampling Rate	[61, 70]
Bluetooth	If Bluetooth is enabled and if the phone is connected to a specific device	On Mode Change	[5, 6]
Ringer Mode	Silent, vibrate or normal ringer mode	On Mode Change	[21, 34, 35, 50, 58, 60, 61]
Screen Orientation	Portrait or landscape mode	On Mode Change	[35, 50, 61, 70]
Screens State	Screen turned on or off, and if the phone is locked or unlocked	On Event Received	[34, 35, 50, 61, 70, 76]
Battery Level	Current battery level and if the phone is charging	750ms Sampling Rate	[34, 35, 35, 58, 60, 61, 70, 76]
Internet connection	If the phone is connected via WLAN or mobile or disconnected	750ms Sampling Rate	[15, 35, 50, 58, 61, 70, 70, 71]
Communication Events			
Calls	Outgoing, ringing or income calls with duration	On Event Received	[5, 6, 15, 21, 34, 42, 58, 60, 70, 71]
SMS	Timestamp of outgoing or incoming SMS	On Event Received	[5, 6, 15, 42, 58, 61, 70]

3.1 *Rabbit-Hole-Tracker*: Mobile Tracking App

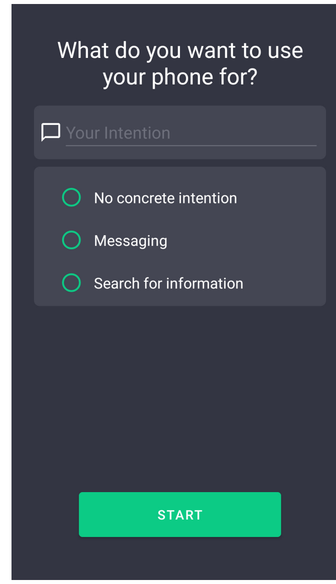
To only record necessary data out of performance and privacy issues [37], we first surveyed related work on which smartphone sensors are a potentially valuable source to understand the MPRH. Based on that, we implemented the RHT app. The app combines data tracking with the user's input on unlock and lock using the ESM [76]. We informed which data to collect through related work on detecting and predicting various smartphone behaviors listed in Section 2.3. Table 1 lists the smartphone data we tracked.

3.1.1 Technical Implementation.

Data Tracking. We followed the guide by Bemmann et al. [4] to give the user an overview and control over the different sensor recordings. Thus, the user needs to grant all the necessary



(a) Sensor-tracking overview screen



(b) ES-questionnaire at unlock

Fig. 1. The main screen of the *Rabbit-Hole-Tracker* App lists all the sensors that are being tracked between unlocking and locking.

permissions for the application to work properly, see Figure 1a. Special permissions and services, such as the Android accessibility service for the app, the notification listener, and access to Android's usage statistics, need to be manually activated in the smartphone's settings by the user. Following, the main screen presents an overview of all sensors that are being tracked. The sensor recordings start automatically when the user enters the main overview screen for the first time, see Figure 1a.

Each usage event type is implemented in one tracking sensor. The different usage information types can be grouped into four types: 1) *smartphone sensor* data, which are internal smartphone sensors such as the accelerometer or proximity sensor; 2) *usage events*, such as the accessibility service events or Android app usage events; 3) *smartphone state information*, which refers to, e.g., the current ringer mode, the screen state or the internet connection availability and source; and 4) *smartphone events*, which include phone calls' or SMS' received.

To focus on smartphone usage during active phone usage, we collected data only when the smartphone was unlocked and stopped logging when the user locked their phone again. Data logging was performed using two different sampling methods: either at a 750-millisecond sampling rate or when a change was detected, such as an android sensor event. We opted for a slightly higher sampling rate than the 500-millisecond rate used by Böhmer et al. [7] to minimize power consumption, transmitted internet data, and privacy concerns. This higher rate still allowed us to capture accurate switching events, including changes in the ringer mode.

We cluster the data into sessions (such as a *rabbit hole* session). A session is generally defined by to the screen turning on until the screen turns off. Thus, it starts with screen[1] and boot[2] events: A *ON_USERPRESENT* screen event indicates the start of a session, an event of either screen event *OFF_LOCKED* or *OFF_UNLOCKED* or the boot event *SHUTDOWN* marks the end of a session.

[1]https://developer.android.com/reference/android/content/Intent#ACTION_USER_PRESENT, last accessed 2023–07–17
[2]https://developer.android.com/reference/android/content/Intent#ACTION_BOOT_COMPLETED, last accessed 2023–07–17

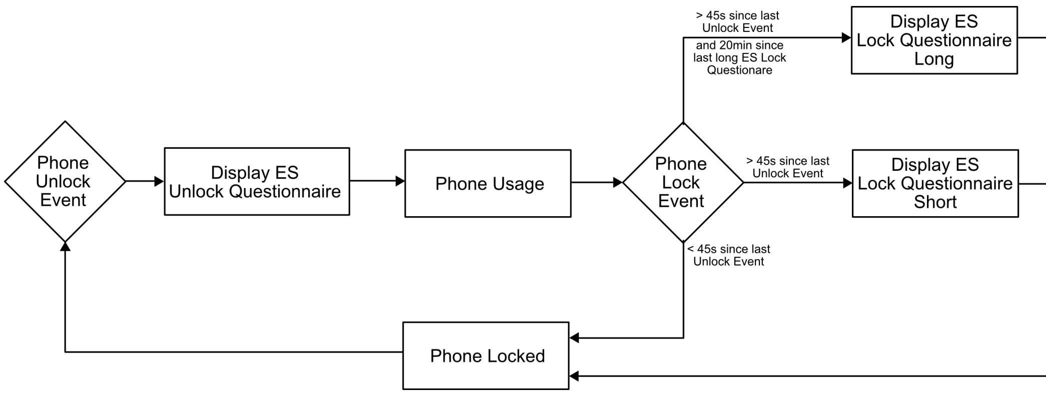
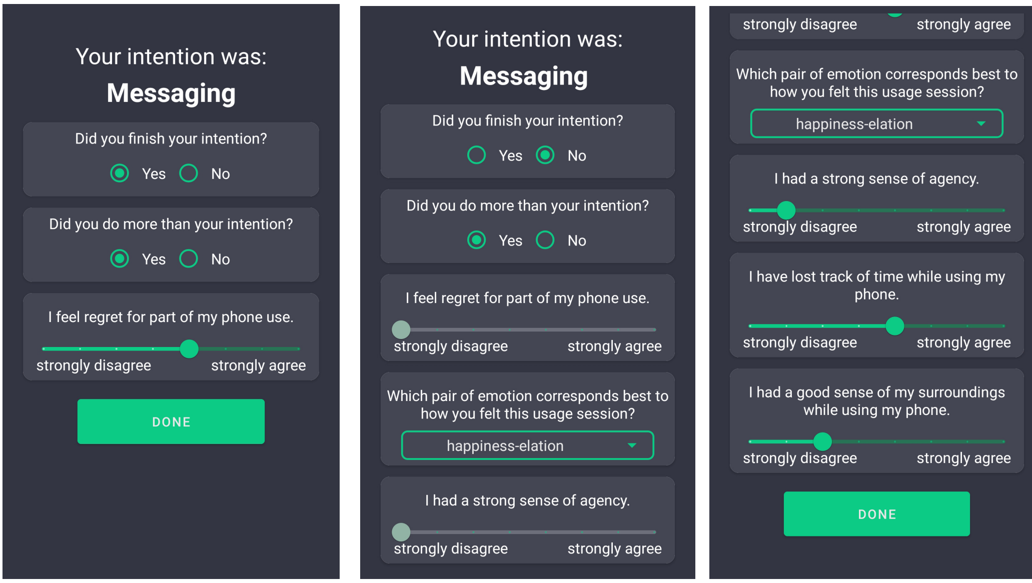


Fig. 2. Sampling protocol for the ESM of the application.

Experience Sampling. Besides data-tracking, the RHT app prompts the user with experience sampling (ES) micro-surveys about their smartphone use intentions and perceived usage at every unlock and lock event: one unlock survey and two types of surveys at lock. Figure 2 summarizes the app's ES scheduling, which we further explain.

We did so to have as many usage sessions labeled as possible, as opposed to some related work, where the questionnaires are spread out more throughout the day [20, 50, 62]. However, this meant that the user got surveys at *every single* lock and unlock. Given the high occurrence frequency of the unlock and lock events, this decision bore a high potential for annoying and burdening the user. Thus, we designed the surveys to be as unobtrusive and quick to answer as possible: the ES unlock questionnaire only asks for users' intention of smartphone use for that particular usage session (see Figure 1b). The user could either input their answer in an autocomplete text field, which prompts previously entered intentions that the user can select from after the two first letters, or they can choose from a list of the last entered intention and preset answers. These preset answers were *"No concrete intention," "Messaging,"* or *"Searching for information."* We selected these as messaging is one of the most common intentions for phone use. Furthermore, *"no concrete intention"* represent a more habitual intention to use, whereas *"search for information"* represent an instrumental intention [28]. The questionnaire is dismissed as soon as the user offers an answer. This was implemented to understand different smartphone use intentions and remind the user of their intention at the end of a usage session (i.e., at lock).

At lock, either a short or a long ES lock survey could appear (see Figure 3), with short being the default (see Figure 3a). The survey was triggered if the last lock event was more than 45 secs ago. We informed this threshold from [76], who recommend using the threshold to minimize the error of session identification. The short version of the ES lock questionnaire displays the entered intention at phone unlock and asks whether the user finished their intention or did more than their intention, as well as whether the user felt regret for any part of their session use. The long ES lock questionnaire (see Figure 3b) replaces the short lock questionnaire if the previous long questionnaire appeared more than 20 minutes ago. Besides including all questions of the short questionnaire, it further contains questions about users' emotions, user's perceived sense of agency, sense of (loss of) time, and sense of surroundings after being engaged in the session. The question types are either answered with an eight-point slider, a yes or no selection, or a drop-down selection. If *"no concrete intention"* was selected in the ES unlock questionnaire, the ES lock questionnaire did not display questions about finishing the intention.



(a) Short ES-questionnaire at lock

(b) Long ES-questionnaire at lock

Fig. 3. The ES questionnaires that were displayed at the phone's lock event.

The tracked data is saved using the Firebase Realtime Database[3]. Our Rabbit Hole Tracker app is openly available for researchers on GitHub. Please see https://github.com/mimuc/mobilehci23-mobile-phone-rabbit-hole.

3.2 Study Procedure

We invited participants to the study via our university mailing list and convenience sampling with the following text: *"Have you ever found yourself spending more time on your phone than you initially planned, that is, falling down a mobile rabbit hole of smartphone use? [...] We are conducting a two-week study to investigate this mobile rabbit hole and detect it based on different usage features."*

The only prerequisite for taking part in the study was to own an Android smartphone. After signing up, we sent participants a Firebase distribution link, with an invitation to install our RHT application. The app then sent a notification to the initial survey consisting of an overview of the study, a consent form, demographics and questionnaires on participant's general (SUQ-G) and absentminded (SUQ-A) smartphone use [49]. These scales probe users' dual nature of smartphone use, with absentminded use being linked to inattention in everyday life. The SUQ-A and SUQ-G questionnaires contain 10 items on a 7-point Likert-scale on absentminded and general smartphone use, respectively, such as *"How often do you find yourself checking your phone without realizing why you did it?"*. The questionnaires calculate a score as the result. We chose these scales to nuance our participants' pool based on their smartphone use patterns. After finishing the initial survey, participants could start using the RTH app, where they gave all required permissions. We asked participants to use their smartphones for the next two weeks as much as they would normally do, complemented by the experience sampling questionnaires at smartphone lock and unlock events for the study duration of two weeks. After two weeks, the application prompted a notification

[3]https://firebase.google.com/products/realtime-database, last accessed 2023–07–17.

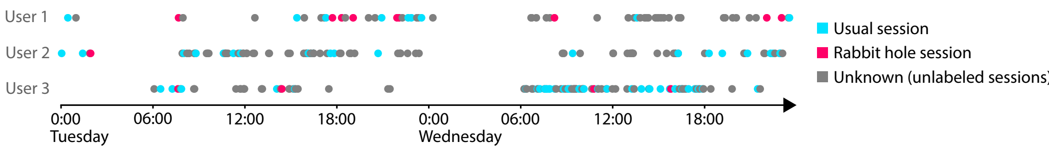


Fig. 4. Smartphone usage sessions of three exemplary participants throughout two days visualized as scatter plot. Each square donates a usage session, where sessions labeled as rabbit-hole sessions are red, sessions labeled as usual as blue, and unlabeled sessions are grey.

containing the link to the final survey. The final survey contained again the SUQ-A and SUQ-G questionnaires. We also included questions on participants' experiences with the RHT app, in particular how using the RHT app influenced their overall smartphone use behavior. Finally, we probed participants' perceived reasons for prolonged smartphone use in general. Upon completion, participants could uninstall the app. We compensated participants with 30€.

3.3 Participants

From 28 participants who used the app for two weeks, we had to exclude seven participants who either did not answer the initial or final questionnaire, or did not fill out any experience sampling questionnaire in the app. The final pool of 21 participants have a mean age of 27 years (SD = 9.32). 14 participants identify themselves as female, and 7 as male. 19 participants have either a completed university degree or are currently obtaining one (n = 17).

Participants reported an average score of 4.47 (SD = .99) on the SUQ-A scale and 4.76 (SD = .88) on the SUQ-G scale, placing both between *"occasionally"* and *"frequently."* These numbers describe our participants' pool as rather frequent smartphone users. After using the RTH app for two weeks, participants reported an average score of 4.07 (SD = 1.13) on the SUQ-A scale and 4.56 (SD = .75) on the SUQ-G scale. Neither scores for the SUQ-A or the SUQ-G scale were statistically different before and after (p = .1 and p = .5, respectively), meaning that our tool did not produce any significant behavior change effects within our participants' pool.

4 RABBIT HOLE COLLECTION STUDY: ANALYSIS & RESULTS

By utilizing the labeled data from our in-the-wild user study, we foster an understanding of what a MPRH is and whether we can predict MPRH behavior. Accordingly, we first pre-process the dataset and characterize the dataset to draft an initial understanding of factors describing MPRH behavior, see Section 4.4. Then, we build models to classify MPRH sessions. The prediction models grant us insights into the importance of the many features extracted from the recorded data.

4.1 Pre-Processing

We regard sessions longer than 3 h 20 m as outliers and exclude them from our dataset (76 of our 16,676 sessions were removed) by doing a Z-score analysis over the log-transformed session length. Datasets from users who did not finish both questionnaires, as well as users who did not answer a sufficient amount of experience sampling questionnaires, were excluded from the final dataset to be analyzed. This results in 21 participants having a total of 14,395 smartphone usage sessions. For the data analysis, we aggregated the smartphone events per session, i.e., we transformed the one-line-per-event dataset into an aggregated one-line-per-session format.

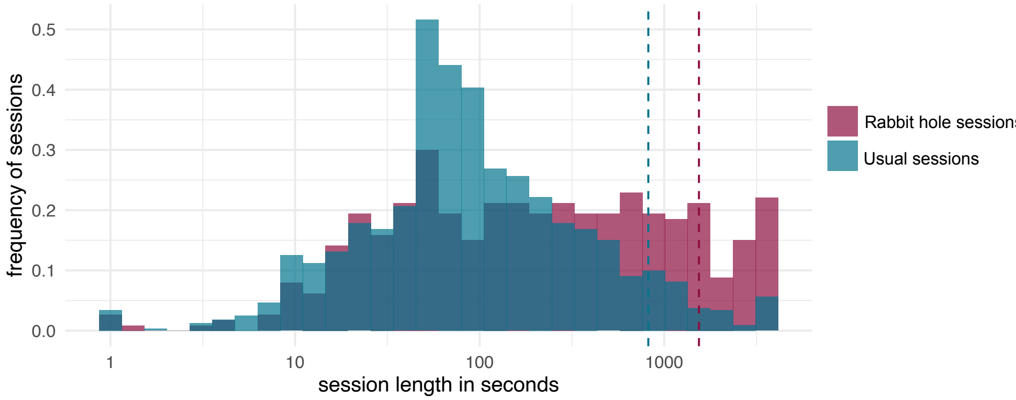


Fig. 5. Rabbit hole sessions are, on average, twice as long as usual usage sessions. To enhance the readability of the differences, the x-axis is scaled logarithmically. All sessions shorter than 1 second and longer than 1 hour were grouped in the first, respectively last bin. The dashed line represents the mean of each type of session.

We incorporated 146 aggregated session features in our dataset, categorized into six groups:

- **Time-based features** include the session length (see Figure 5), hour of the day, and day of the week, as well as features describing the previous sessions like counts of sessions in the past 1/2/3 hours, time passed and glances that happened since the previous session.
- **Smartphone settings.** Internet status and ringer mode.
- **Questionnaire data.** The smartphone session data was joined with each user's questionnaire data. For the demographics questionnaire, we kept the values of age and gender.
- **App Usage.** For each used app, we tracked the number of times it was used in a session, spent time, number of clicks, and scrolls. The data was encoded with one column per app and feature. To reduce the thereby introduced huge dimensionality of our dataset we categorized the apps by the app categorization of Schoedel et al. [67], resulting in 27 distinct app categories that we observed.
- **Time-spent features normalized.** Of features on session length, time spent in an app category
- **Biometrics.** We counted the number of clicks and scrolls in each session (respectively situation) and included features of both absolute counts and time-relative frequencies.

4.2 [RQ1] Rabbit Hole Labeling

Each participant completed on average 479 sessions (SD = 410). An average session endures 4.43 minutes (SD = 12.45), with 4.54 apps used during a session (SD = 7.77). We labeled sessions as *rabbit hole* or *non-rabbit hole* by using the yes and no answer respectively to the *"Did you do more than your intention?"* item collected with ES at lock. This results in 1738 labeled smartphone usage sessions, 25.8% of which are labeled as rabbit hole (N_{RH} = 449, N_{noRH} = 1289). The unlabeled sessions were discarded for the following analyses. The item *"I feel regret for part of my phone use."* further splits the rabbit hole sessions into sessions with an either positive or negative user experience. Users regret 15.1% of the rabbit hole sessions, while 84.9% were perceived positively ($N_{RH\_positive}$ = 302, $N_{RH\_negative}$ = 49, for 125 MPRH sessions the regret item was unavailable). The

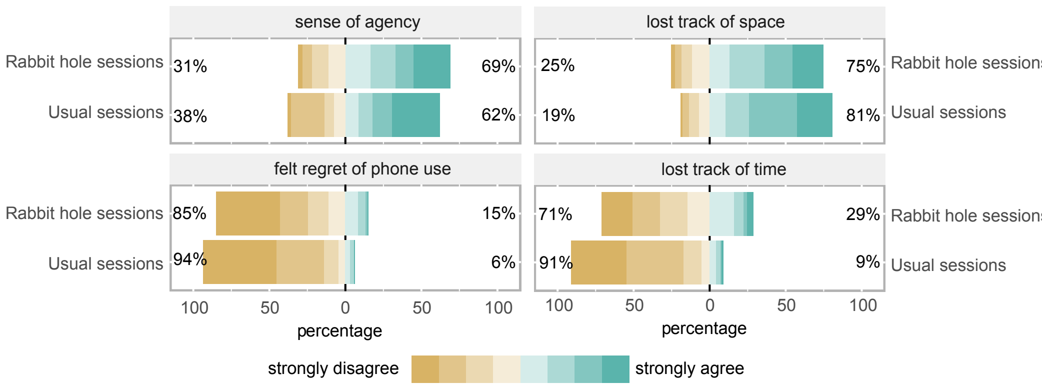


Fig. 6. The distribution of the answers on our ESM items. After rabbit hole sessions users reported a slightly higher sense of agency, loss of track of time, and regretted the use more often. They lost track of space less often.

collected data is depicted in Figure 4, where a subset of sessions of 3 exemplary users is visualized with their MPRH labels.

4.3 [RQ1] ESM Results

In the unlock ES questionnaire, we were asking for the user's intention to use their smartphone. The predefined options *messaging* (3035 times), *no concrete intentions* (2036 times), and *search for information* (1644 times) were selected most often. We grouped other intentions that were mentioned by our participants, the most common intention groups were *listening to music or podcasts* (373 times) and using *social media* (188 times). 6232 sessions were left unlabeled.

On lock, i.e., after the smartphone usage session, we asked the user whether they actually did what they intended. In 1949 sessions, the user stated that they finished their intention, in 238 they stated the opposite (12208 sessions were not labeled). For 449 sessions, they stated that they did even more than intended, while in 1289 they did not (12657 unlabeled). Furthermore, we had four questions on how they perceived the past usage session (7-point Likert scale). For rabbit hole sessions, users reported a slightly higher sense of agency ($M_{RH} = 6$, $M_{no\_RH} = 6$ $N = 960$, $H = 0.16$, $1 d.f.$, $p = .69$) and loss of track of time ($M_{RH} = 3$, $M_{no\_RH} = 2$, $N = 937$, $H = 75.90$, $1 d.f.$, $p < .001$). They regretted their phone use more often ($M_{RH} = 2$, $M_{no\_RH} = 2$, $N = 1387$, $H = 21.58$, $1 d.f.$, $p < .001$), and lost track of space less often ($M_{RH} = 6$, $M_{no\_RH} = 7$, $N = 1246$, $H = 12.00$, $1 d.f.$, $p < .001$), see Figure 6. H-statistics were calculated with Kruskal Wallis Chi-Squared tests.

4.4 [RQ1] Quantitative Understanding of the Rabbit Hole

We conducted paired Wilcoxon Signed Rank Tests on user averages to test for differences in the feature in rabbit hole compared to non-rabbit hole sessions. For comparison, we grouped the session data by users.

Rabbit hole sessions are significantly longer than smartphone usage sessions where the intention has been strictly followed ($session\_length_{RH} = 16.41\,min$, $session\_length_{noRH} = 4.15\,min$, $V = 185$, $p < .001$). Users use apps at a lower frequency, but spend more time in them ($apps\_per\_minute_{RH} = 2.05$, $apps\_per\_minute_{noRH} = 2.62$, $V = 21$, $p < .01$). Specifically, *gaming* and *system* apps were used for longer periods of time, while those associated with *social media* were also given greater attention.(see Figure 7, left subfigure). When considering longer sessions, we observe that only *gaming* apps experience significantly greater usage times, whereas *social media* apps exhibit a

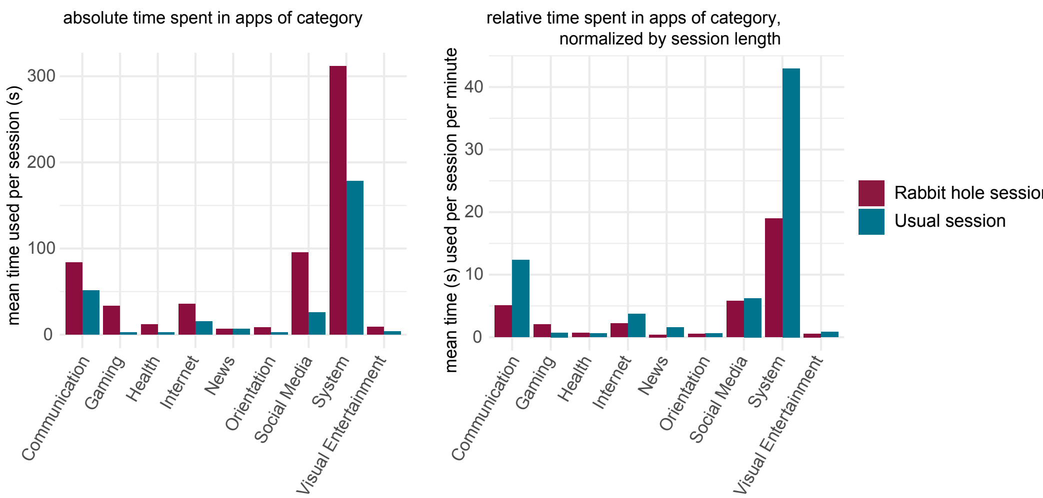


Fig. 7. In rabbit hole sessions, users spent more time, especially in gaming, social media, and system apps. However, regarding this in context of the session length, we see disproportional higher time spent in one specific app category only for Gaming apps.

Feature	group means RH	group means no RH	SMD	paired Wilcoxon Signed Rank Test V	paired Wilcoxon Signed Rank Test p	CI (95%) lower	CI (95%) upper	r
session length (seconds)	985.09	248.52	1.15	185	0.0001	320.09	1092.58	-0.71
scroll frequency (per minute)	11.42	11.55	0.08	101	0.8288	-1.30	1.26	-0.71
click frequency (per minute)	5.95	6.45	−0.04	36	0.01597	-1.97	-0.31	-0.71
time spent in Gaming apps (seconds)	20.54	5.58	0.58	14	0.1056	8.41	98.08	-0.34
time spent in Visual Entertainment apps (sec.)	63.75	15.46	0.06	30	0.8385	-60.41	479.28	0.10
time spent in System apps (seconds)	342.45	160.9	0.21	150	0.02582	2.60	296.35	0.71
time with wifi connected	1057.57	829.32	0.02	162	0.005329	115.35	880.95	0.71
time in ringer mode vibrate	242.43	156.2	0.17	27	0.6248	-115.92	733.23	-0.94
number of apps used per session	11.77	5.16	0.88	180	0.000164	2.03	11.65	-0.71
frequency of apps used (per minute)	2.05	2.62	−0.35	21	0.001694	-1.01	-0.40	-0.71

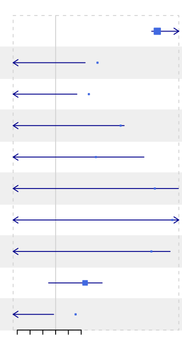


Fig. 8. Characteristics of rabbit hole and non-rabbit hole sessions compared in a forest plot. The means and standardized mean difference (SMD) compare both groups, a paired Wilcoxon Signed Rank test indicates whether a characteristic behaves significantly differently. Rabbit hole sessions are significantly longer, while the frequency of used apps is lower. On the very right we provide 95% confidence intervals and effect sizes.

proportional increase in usage with session duration. The used *system* apps are mostly *launcher* apps (e.g., *com.android.systemui*, *huawei.android.launcher*), which reveals users to be more often on the home screen, yet, leaving it quickly. The extended duration of rabbit hole sessions significantly contributes to the large amount of time spent on the home screen or in *system* apps. Ruling the higher session duration out, the relative amount of time spent is actually smaller.

To give an impression of the differences in user behavior between rabbit hole and non-rabbit hole sessions, we show the used apps of three randomly chosen sessions each in Figure 10. While users behave more focused in non-rabbit hole sessions, i.e., directly going from app to app, we

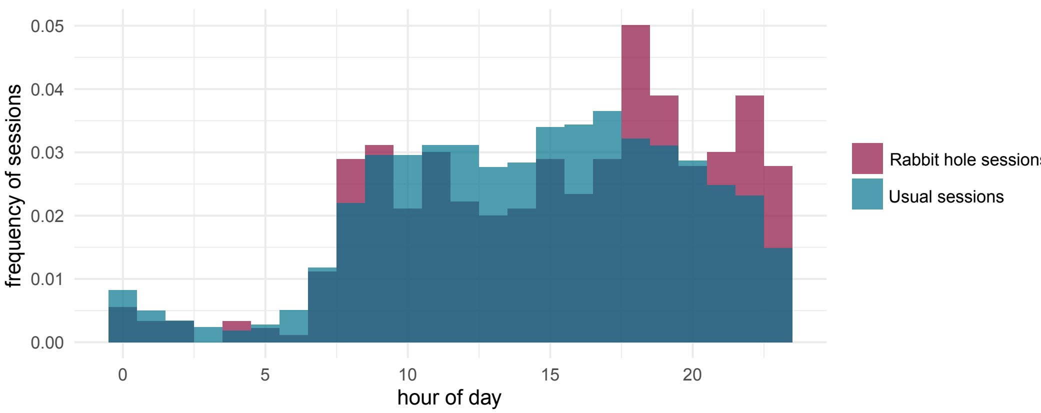


Fig. 9. Rabbit hole sessions tend to happen rather later in the evening, and less frequently during the day.

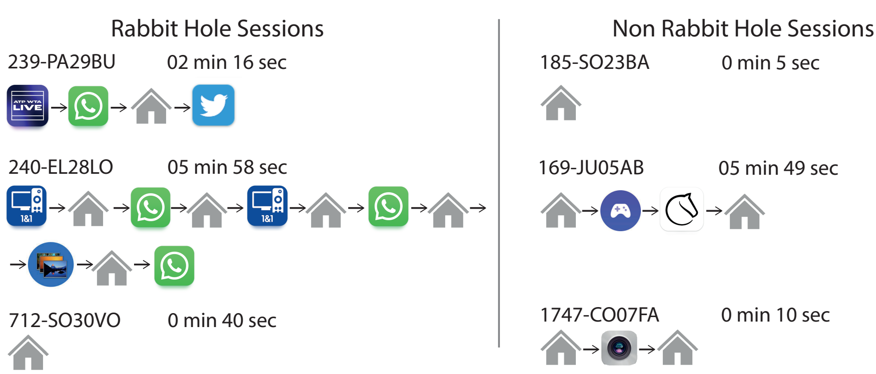


Fig. 10. app sequences (sequence of icons) of 3 randomly sampled sessions. (a) for RH sessions, (b) for no RH sessions

found many visits of launcher apps and a usage of more unrelated apps. Launcher app usages in usual sessions seem more targeted, for example to reach the camera in example 6.

In a rabbit hole session, the smartphone is longer connected to a WiFi network ($time\_wifi_{RH}$ = 17.63 min, $time\_wifi_{no\_RH}$ = 13.82 min, $V = 162$, $p < .01$) and we observe a lower click frequency ($click\_freq_{RH}$ = 5.95 $clicks/min$, $click\_freq_{no\_RH}$ = 6.45 $clicks/min$, $V = 36$, $p < .05$).

In the time before a rabbit hole session (we regarded between up to 1 and up to 3 hours before), fewer smartphone usage sessions happen than usual ($sessions\_previous\_hour_{RH} = 3.45$, $sessions\_previous\_hour_{no\_RH} = 4.21$, $V = 35.5$, $p < .05$; $sessions\_previous\_three\_hours_{RH} = 8.34$, $sessions\_previous\_three\_hours_{no\_RH} = 10.00$, $V = 36$, $p < .05$). Rabbit hole sessions happened more often later in the day, i.e., between 6 p.m. and midnight, and less frequently during the day (see Figure 9).

4.5 [RQ2] Predicting the Rabbit Hole

We predict single rabbit hole sessions, i.e., develop an algorithm that can detect an ongoing rabbit hole right at a time or afterward. Therefore, we have split our dataset into a train, validation, and test dataset by participants, i.e., we assigned 15 users to the train set, 3 to the test dataset, and 3 to

Table 2. The model parameters and their optimization that was tried by a grid search. The values for session prediction are underlined identifying the best value of each parameter.

Model Parameter	Optimization Range
n_estimators	5, <u>10</u>, 100, 200, 500, 700
max_features	<u>sqrt</u>, log2, None
max_depth	4, 5, 6, 7, <u>8</u>, 10, 12, 14, None
min_samples_leaf	<u>1</u>, 2, 4
min_samples_split	2, 5, <u>10</u>
criterion	<u>gini</u>, entropy, log_loss

validation. Due to the huge class imbalance, we applied SMOTE oversampling [10]. After the initial model investigation, we selected a random forest as the model best-performing model. Then, we optimized the hyperparameters using with a grid search, tuning the parameters listed in Table 2. The model training with the identified best parameter configuration takes approximately 4 seconds on a commodity notebook. We implemented the prediction model using Python's sklearn library[4].

We investigated to predict whether a smartphone usage session is a rabbit hole or not, i.e., treating each session as an observation and the users' label on whether they did more than intended as the target variable. The chosen model optimization parameters, which we identified through a grid search, are underlined in Table 2.

On the training dataset, with 15 participants, the model reached an accuracy score of 87.97%, and on the test dataset, with 3 participants, 64.97%. On the validation dataset, where we tested the model with 3 more yet unseen participants, the model's performance is 72.41%, which we consider as the model's actual performance. The precision for rabbit hole sessions thereby was higher (77%) than that of usual sessions (69%).

Analyzing our model's feature importance, we find that the features contributing most to the prediction of a rabbit hole session are related to app usage, precisely usage of apps of the categories *social media*, *system*, and *communication* (descending order by the impact on the model output magnitude). The number of clicks and click frequency show high importance, and also the number of sessions that happened beforehand in the previous 3 hours. Device settings rank rather low, only the wifi status *connected* ranks high. Other settings, such as the ringer mode, show even less importance to the prediction. Calculating a conversion to time-relative features is beneficial for some features, esp. the number of usages of *system* apps and time spent in *social media* apps.

To understand how these features contribute to the models' prediction result, we created SHAP values and plots [47]. The SHAP values allow for a more in-depth analysis of how feature values influence the result. In Figure 11, we visualize the top 20 features for the prediction model as a SHAP plot. The distribution of the colored dots conveys how each feature value is distributed and which values contribute to which direction (i.e., push the prediction decision towards the result *rabbit hole* or *non-rabbit hole*. Spending time in *social media* apps is a strong predictor for rabbit hole sessions, same for *system* apps (red dots are shown on the right side only). A high number of scrolls in *system* apps and clicks in general also argue towards a rabbit hole. However, a generally high scroll frequency argues rather against a rabbit hole.

[4]https://scikit-learn.org/stable/index.html, last accessed 2023–07–17.

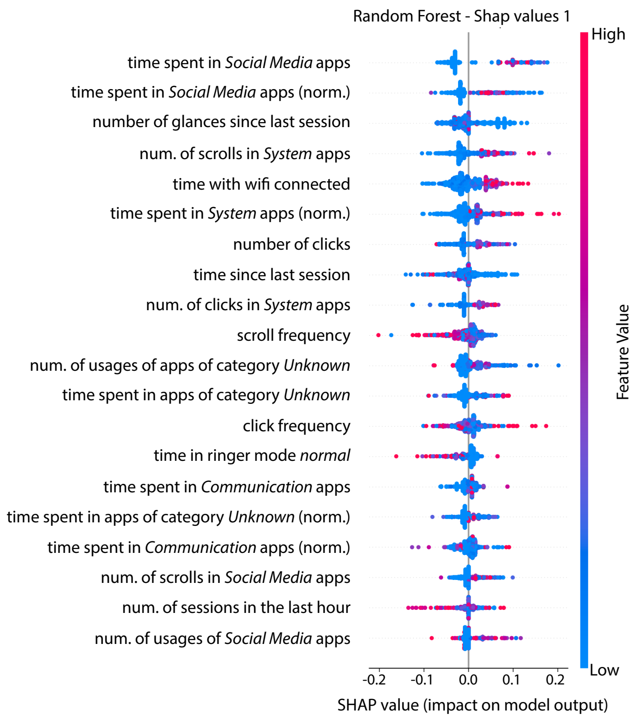


Fig. 11. Beeswarm SHAP plot, visualizing how the top 20 features contribute to our session prediction model. Each point indicates how an observation contributes to the model's output. A positive impact value pushes the prediction result towards deciding on a rabbit hole and a negative one against it. Features with the suffix *(norm.)* are normalized by the session length.

4.6 Field Study Limitations

Before starting the smartphone use session, the intention of the user is asked. As such, the application and the study setting itself may have biased user behavior toward making the smartphone use session shorter. Albeit our participants did record a slight decrease in the frequency of their smartphone use by means of the SUQ-A and SUQ-G questionnaires, these have shown to be statistically insignificant. Furthermore, a recent study showed the effects of asking the smartphone use intention beforehand to not statistically decrease the overall screen time [74]. However, as we conducted the recruiting online and included all volunteering participants, the final sample of participants may have been biased by self-selection.

Furthermore, our classification approach is based entirely on user's self-reported feedback. As such, our data may be unreliable due to factors such as inaccurate responses (i.e., the number of non-labeled sessions) and the influence of the experience sampling's wording. Yet, for the latter, 90% of participants have or are obtaining a university degree, meaning that they possess high knowledge of English.

Finally, a significant number of use sessions did not involve a specific purpose or goal beforehand, meaning that, in many instances, users opened their smartphones without any particular intention or the intention was to tame boredom. Therefore, these instances were not considered as MPRH based on the adopted MPRH understanding from Cho et al. [12]. Likewise, situations in which a user intends to perform a specific action, such as checking the weather, but gets distracted by a notification and ends up reading incoming messages, could be categorized as rabbit holes based on the definition used, although *intuitively* being no rabbit holes. For this reason, we conduct a focus group and report its finding in the following section.

5 FOCUS GROUP: QUALITATIVE UNDERSTANDING OF THE MOBILE PHONE RABBIT HOLE

In the previous section, we conducted an in-the-wild study and identified specific patterns that differed from the findings of existing research. As a result, we chose to take a step back and enhance our quantitative findings by incorporating a qualitative approach through an expert focus group consisting of six participants. Our objective was to delve deeper into the concept of the MPRH and develop a more comprehensive definition. Additionally, we sought ideas on how to effectively communicate the detection of an ongoing MPRH to the user, as well as brainstorm potential concepts for prevention and intervention. The goal of the focus group is thus two-fold: firstly, we aim to gather insights into experts' conceptualizations of an MPRH and its relation to mobile digital technology. Second, we are interested in ways of communicating that falling into an MPRH is likely to happen to affected smartphone users.

5.1 Procedure

In a short introduction round, we carefully explained the aim of the focus group with particular attention not to bias the experts' own understanding of an MPRH. The focus group consisted of two parts. In the first part, we asked participants to silently reflect on and write down their understanding of an MPRH and the terms they relate to it. The reflection phase ensured deliberate contemplation and recall to first form their own conception. We then asked each expert to read out their own notes to the group. Only after we encouraged a discussion among the participants and asked them to identify the factors that influence their feeling of falling down an MPRH. We thereby established a mutual understanding or an awareness of what it means to others, i.e., its various meanings, facets, and factors.

In the second part of the focus group, we aimed at gathering guidelines or solutions on how to communicate the mobile rabbit hole to the user that they might be falling down an unwanted rabbit hole: *"Assuming your smartphone would be able to (a) predict or (b) detect you falling into the rabbit hole, we would now like to discuss guidelines or solutions on how to present such findings to the user."* Participants then split up into three groups, with the groups receiving one of the following scenarios: **1) Prediction:** *How could we let the user know that an MPRH session is likely to happen?*, **2) Detection:** *What solutions could help in dealing with unwanted MPRH sessions once they have been identified?*, and **3) Wrongful Prediction/Detection** *How might users respond in the case of inaccurate or inadequate prediction/detection?* Each group discussed potential solutions or guidelines for the given scenario. They optionally sketched the solutions on paper. Next, they presented their solutions and discussed them in the big round.

The focus group lasted for one hour. We recorded the session and obtained participants' consent beforehand to do so. The session was finally transcribed for further analysis.

5.2 Participants

We recruited six HCI experts (four female, two male, mean age = 28.6 years), i.e., people who do teaching and research in HCI. Table 3 provides an overview of the experts. All experts use a smartphone daily.

Table 3. Demographics overview of our focus group participants.

ID	Expert's Research Topic	Gender	Age
E1	Health & Wellbeing at Work	M	31
E2	Mental Health & Wellbeing	F	27
E3	Mental Health & Wellbeing	F	32
E4	Technology-Mediated Communication	F	28
E5	Usable Privacy & Security	F	26
E6	Human-Robot-Interaction	M	28

5.3 Analysis

First, we transcribed the audio recordings. The first and second authors independently analyzed the transcribed data set of the first question of the focus group, identifying emerging themes. We then subsequently discussed any existing doubt until an agreement on a common definition was reached. Next, we thematically analyzed the experts' statements on their understanding and experience of a mobile phone rabbit hole. We applied a bottom-up, open-coding process using affinity diagramming [24].

5.4 Results

We present the results for the focus group to understand the MPRH from a qualitative perspective. We accompany the described factors and facets of a MPRH, as pictured in Figure 12, with applicable experts' quotes. Concluding, we list the experts' design suggestions to communicate mobile phone rabbit holes to the user.

5.4.1 [RQ1] Understanding the MPRH. Before a rabbit hole occurs, our participants' statements reveal that there is typically a quick and practical initial task to start smartphone use, such as checking the weather, checking an incoming notification, or seeking information. Here, E1 stated: *"I look at my phone for some practical reason, usually to check the weather, but then I scroll here and there, and I've read it for 20 minutes now. And then, when I close my phone, I realize I did not check the weather."* At time though, there is no clear intention before on what to do. Instead, users intend on how long to use the smartphone – i.e., they want to kill a rather short amount of time. This factor is covered as *initial situation* in our derived user-centered definition in Figure 12. Following, experts explain to deviate from their initial intention of doing something or spending a certain amount of time as *"something (else) lures them in(to the rabbit hole)."* According to our participants, this something can be an internal or external *trigger*. Internally, it might be an established habit, raised interest they picked up along the way of the rabbit hole, or procrastination. Named external triggers were algorithms displaying similar content or recommendations. E6 pictured it as a *"silenced recommendation-based jump in."*

Once in the rabbit hole, participants report several *behaviors* they experienced. They dive deep into a topic they might or might not be interested in, jump between content randomly or per recommendation, do doom scrolling, or engage in meaningless, that is, aimless interaction. For example, E4 reported: *"[You] dive deep into a topic [you] did not know you were interested beforehand."*

The depicted behaviors are followed by a *state* of the participant, which we categorize as active and passive states. In the active state, experts report remaining deeply focused and interested as they continue going deeper down the rabbit hole, especially with content they like or feel entertained about. E6 commented: *"I actually like [the rabbit hole], especially with Youtube, I enjoy doing that."* Thereby the users are in active control on starting with and continuing in the rabbit hole *"if I don't*

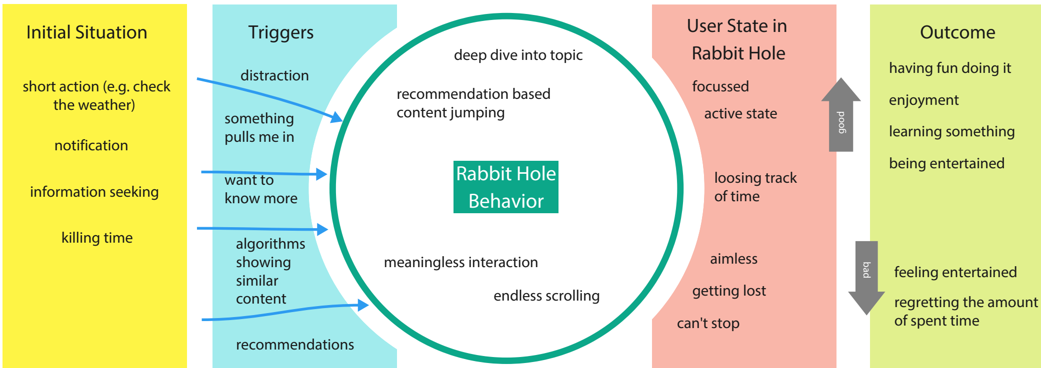


Fig. 12. The derived user-centered definition of a mobile phone rabbit hole.

have time, I don't do it" (E4). E6 had a similar vision. On the contrary, in the passive state, E1, E2, E3 and somewhat E5 note to subconsciously continue their smartphone interaction. At times, they might become aware of them being in a rabbit hole, but note that they cannot stop it. In both states though, users report losing track of time: *"20 minutes appear to be 2 minutes"* E1.

As a rabbit hole outcome, we identified enjoyment, entertainment, and knowledge gain as *positives*. *"You just enjoy it, like spending time entertaining yourself"* (E4). In some cases, though, the irrelevance of the topic can still cause a negative feeling in the aftermath; such behavior was explained by E1: *"Sometimes, I'll go deep into something and [...] in the end, I will be, I learned something new, this is positive, I'm interested. [But then], I still don't care about that."* E3 complements this statement: *"[Even though] I end up seeking information and get really deep into a topic, it's very exhausting."* Thus, the passive state induces an exclusively *negative* outcome, e.g., feeling regret for the wasted time.

Finally, participants discussed some common scenarios, e.g., contexts or content. They recall incidents when they themselves have fallen down the rabbit hole. Three experts mentioned social networks, in particular YouTube or Reddit. E5 and E6 noted the scenario of shopping inventory they are not particularly familiar with. As such, E6 commented *"I still want to maximize my choice, [...] the thing I buy, and optimize my spending."*

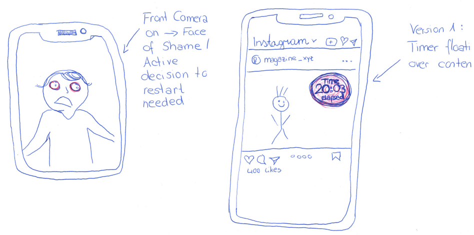
5.4.2 [RQ2] Design Suggestions to Communicate the MPRH. In the second part of the focus group, we asked experts to brainstorm ideas on how to communicate the likeliness of the user falling down an unwanted MPRH. We differentiated between the detection and prediction scenario, as well as the wrongful detection scenario.

Given the conclusion of the first part of the focus group, namely that most unwanted rabbit holes happen subconsciously, the goal of informing the user should be primarily to raise awareness that they are about to fall into a rabbit hole; E3 described it as *"a trigger for self-reflection."*

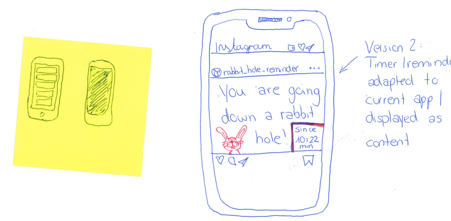
The experts' group agreed that sudden notifications and pop-ups that consume the whole screen, such as iOS' screen-time notifications[5], should *not* be the way to go – these tend to become swiftly annoying and ignored by the user. Thus, four experts (E3-E6) proposed a reminder banner that blends in the user's current content context. For example, E3 proposed *"[a] reminder there so that you cannot miss it, but you can also easily just ignore it."*

Blending in was an important discussion point – the rabbit hole is a continuum, and as such, the user should not be suddenly informed of the rabbit hole, but rather gradually, as the rabbit hole

[5]https://support.apple.com/en-gb/guide/iphone/iphbfa595995/ios



(a) *Left:* A user terrified of their image in the black mirror, after being in a MPRH. *Right:* Floating timer.



(b) *Left:* Quick shutting down of the screen once MBRH was detected. *Right:* Timer blended in content, as Instagram post.

Fig. 13. Some resulting sketches from our experts' focus group on the question of communicating MBRH to the user.

develops. This could give the user time to prepare ending the interaction with their smartphone. Concrete examples suggested were presenting a timer (see Figure 13a) or turning the screen gradually off (see Figure 13b), similar to rendering a *tunnel* on the smartphone's screen.

In the case of a textual prompt, the prompt could ask the user questions beyond their smartphone behavior. E3 and E4 proposed asking about user's emotional state, how they feel about this use session; reminding them of their initial intention and whether they have fulfilled it – if not, the smartphone could *"jump"* to the app of the initial intention; or proposing a contextually appropriate alternative activity instead, to exit the digital tunnel. E1 reflected at this point: *"How do we actually get out of rabbit holes now? [...] It's usually like, some time critical thing comes up that you have to do or you, like, finally convince yourself that, like, I could be doing something better every time."*

The smartphone could proactively and beforehand suggest an alternative activity either from user's daily ToDo List (E2) or randomly (E6) before the MPRH. Further discussion revolved around balancing effectiveness and annoyance – the more drastic ways are more effective (e.g., turning off the screen), but also more frustrating when they're wrong (e.g., the user might think the smartphone is broken). If an incorrect detection occurs or if the user wishes to explore further, certain participants suggested that the user should be given the option to continue their exploration. However, this would require the user to actively respond: *"[If] I wanted to be in the rabbit hole, I will just turn on my smartphone, which is okay, but I do have to do something active [as response]"* (E6).

Another proposal involves a *"cute"* visualization of a rabbit – in order to reduce the annoyance factor – that needs to be taken out of a hole. E2 suggested the option for the user to indicate intentional rabbit hole desire beforehand, in order to prevent wrongful detection. In any case, experts enhanced the importance of the system learning individual patterns of use and adapting to those, as *"clearly, we all have, like, different [patterns of use]"* (E1).

6 DISCUSSION

Previously, we explored a MPRH from both quantitative and qualitative points of view. We first merge these mixed-methods findings and propose a *mobile phone rabbit hole* definition. We then re-run the quantitative analysis with the new definition to conclude by answering RQ1. Our deepened understanding of the MPRH allows us to suggest an on-device detection and prediction of a rabbit hole, accompanied by UI implications for such a predictor. We wrap up with a discussion on ethics and empowerment.

6.1 [RQ1] A Mobile Phone Rabbit Hole Definition

Related work informed us to perceive a mobile phone rabbit hole primarily through the lens of deviated intention (e.g., [12, 48]). Thus, our chosen target label to treat a smartphone use session as a MPRH was whether users engaged in more smartphone activities than they intended in that particular session. In more than one-quarter of the labeled sessions, this showed to be the case. Moreover, after almost 85% of MPRH sessions, users expressed regret. These findings align with experts' statements, as well as related anecdotal evidence that MPRHs induce a rather negative emotional state. However, qualitative findings and the remaining 15% of not-regretted MPRH object to the idea of a strictly negative perception of MPRH. Our results suggest entertainment and enjoyment in the session without disturbing the user's real-world obligations to be the case – although the same loss of track of time happens as with negative MPRH.

Indeed, the 25% of sessions labeled to be a MPRH were, on average, significantly longer than non-MPRH sessions for almost 8 minutes. This aligns with findings that 90% of smartphone sessions are shorter than four minutes [84]. The duration of a MPRH overlaps with a findings from a recent qualitative study [73] that users express regret around the 10-minute mark. However, our mixed-methods results suggest the duration of the session to not be the sole indicator of a MPRH, as users might be enjoying a movie or reading an e-book. The high occurrence of system apps, especially launcher apps, in our data, expresses that users often visit the home screen in a MPRH. This might be to get inspired on what to do next, without any predefined intention, further extending the duration of the session.

If no concrete intention existed for the use session, the ES questionnaire at lock did not probe whether the use exceeded the intention. Following the established perception of deviated intention, these sessions ($n = 2,123$) could not be classified as MPRH sessions. However, our qualitative results reveal that users at times actively engage in smartphone interaction with no intention on what to do, but *how long* to do it – i.e., to kill time – but then they get lured into a MPRH, and they spend longer on their phone than planned. With that being said, we propose an extensive definition of the ***mobile phone rabbit hole*** in Figure 14.

Definition

MPRH *The mobile phone rabbit hole refers to the subjective feeling of being drawn into smartphone activities that divert from the user's original intention or intended duration of use.*
Implication *Engaging in a MPRH leads to prolonged screen time and diminished awareness of surroundings or time.*
Negative MPRH *If the user loses awareness of their smartphone use while in the rabbit hole, it is a negative rabbit hole.*

$$\textbf{RH:}\quad {}_{before_{RH}}(!intention \quad OR \quad !intended\ time) \quad AND \quad {}_{during_{RH}}!(context\ awareness) \tag{1}$$

$$\textbf{negative RH:}\quad RH \quad AND \quad {}_{during_{RH}}!(interaction\ awareness) \tag{2}$$

Fig. 14. The deduced definition of a MPRH, emerged from our mixed-method exploration.

The formalization is to be read as follows:

(1) *"If user's intended intention OR planned time to spend AND their awareness of either time or surroundings is violated, the user is in a rabbit hole."*
(2) *"If user's awareness of smartphone use behaviors in the rabbit hole is violated, it results in a negative outcome."*

6.2 [R1] Quantitative Analysis: Re-Run with New Definition

With the emerged MPRH definition from the focus group, we re-run the quantitative analysis regarding RQ1, meaning what MPRH-sessions make different from non-MPRH-sessions.

6.2.1 Rabbit Hole Labeling. We re-did the labeling of smartphone use sessions into rabbit hole and non-rabbit hole again. Now, however, using the new definition above (see Figure 14). Given 14395 recorded sessions, this results in 152 rabbit hole sessions and 1165 non-rabbit hole sessions (13078 unknown). Of the rabbit hole sessions, 68 were perceived positively and 54 negatively (30 unknown). In comparison to the labeling with the old definition, the percentage of rabbit hole sessions is now lower (11.5% vs. 25.8%). However, the share of negatively perceived rabbit hole sessions increased (44.3% vs. 15.1%).

6.2.2 Quantitative Understanding of the Rabbit Hole. We re-ran the prediction model training with target labels based on the new definition, using the same parameters as in Table 2. Overall, we could not find any significant differences among the contributing features. *Social media* app usage features still show the highest feature importance and thus have the largest impact on the models decision for or against a session being a rabbit hole. The previously high ranking features number of *glances since last session*, *scrolls in system apps*, and *time with connected wifi* dropped in their contribution. Instead, *clicks in health* and *communication* apps became more relevant.

The new model performs slightly worse, reaching accuracy scores of 93.36% (compared to 87.97%) in the training phase, 69.51% in the test phase (compared to 64.97%), and 61.18% (72.41%) in the validation phase which we regard as its actual performance. The precision gap between rabbit hole sessions 82% (77%) and usual sessions 57% (69%) thereby increased. Thus the new model performs slightly worse, due to over-fitting on rabbit hole sessions.

6.3 [RQ2] On-Device Detection and Prediction of Rabbit Holes

Besides gaining a quantitative and qualitative understanding of the MPRH, Section 4.5 demonstrated that the MPRH can be detected and predicted with mobile sensing data. Our first model faces some limitations, esp. regarding its score of around 70% and the tedious labeling approach with experience sampling. In the following paragraphs, we discuss how future work might overcome the current limitations.

Contextual and Situational Data. Our RHT app (c.f. Section 3.1) focused primarily on smartphone usage behavior, i.e., what users do on their phones. The focus group informed us that it takes an initial situation of the user (e.g., desire to kill time, intent to look up contextual information briefly) combined with a trigger (e.g., distracting notifications, recommendation algorithms) to lead to a rabbit hole session. To encompass these findings, future work could investigate sensing data with the aim of extracting rabbit hole-prone user contexts and situations. By combining data on location and time, the system could detect scenarios raised in the focus group, for example, killing time instead of going to sleep. We see potential in the prediction of rabbit hole-prone *situations*, rather than individual *sessions*. With our definition, we argue that the observed smartphone usage behavior (i.e., rabbit-hole behavior in Figure 12) is a property of the MPRH. In a rabbit hole-prone state (e.g., desire to kill time), users are prone to be caught by recommended content, which we comprehend as the high usage of entertainment and gaming apps in our data. The high occurrence of *system* apps in a MPRH, particularly launcher apps, expresses users often visiting the home screen. This might be due to the state of boredom and lack of inspiration on what to do next – in pursuit for the next dopamine shot.

Refined Experience Sampling. After improving the understanding of the MPRH with our quantitative and qualitative studies, we suggest tailoring the ES questionnaires, used to label smartphone sessions, to our new definition of the MPRH. This includes adding the time-to-kill intention and whether the time spent has stayed within the planned time frame. Furthermore, we propose excluding the questionnaire at unlock as it might bias users towards using their smartphones shorter in a session (i.e., have a behavior change effect). Although future systems might start with pre-trained models to avoid the user labeling huge amounts of smartphone sessions, a model personalization phase might be needed. Smartphone behaviors and the perception of the MPRH are individual, so future prediction models could follow a personalized (e.g., Lu et al. [43]) or federated learning approach (e.g., Google applies to improve the Android keyboard [85]) to accommodate these individual differences and adapt the model accordingly.

Raise Awareness (Beforehand) or Subtly Intervene? Scenarios towards fine-grained prediction or detection of an ongoing MPRH imply future research applications to investigate both reflective and corrective strategies on communicating smartphone use behaviors, including MPRHs, in the wild.

Interruptions and self-monitoring measures are common strategies to communicate (unwanted) smartphone use behaviors. Yet, sudden interruptions can raise a high level of frustration and be easily ignored. Contrary, our findings suggest an MPRH as no binary state in which one clearly is in or not, but rather as a continuum of time spent or content consumed. Here, researchers and developers could explore fostering in-the-moment awareness via on- or off-screen subtle indicators such as the proposed visualizations in Section 5.4.2 or ambient indicators that blend with the smartphone's surrounding world (e.g., [13, 55]). Our findings additionally suggest context-aware reflective prompts – compared to current solutions such as Screen Time's App Limits [6] that focus on the activities on the device solely ('Give me one more minute!), these could ask questions such as *"'How do enjoy this session?"'* (e.g., user's inner context) or *"'You started {here}, do you want to go back?"'* (e.g., regaining agency over the sessions flow).

With a greater dataset, our predictor model could enable more fine-grained situated monitoring of smartphone use, i.e., how smartphone use is situated in different everyday life contexts. These might build on top of existing self-monitoring visualization that, again, solely display smartphone use indicators ignoring the interplay of smartphones and users' everyday activities outside of it. Thus, future research could employ and compare the suggested approaches in its efficacy.

6.4 Awareness Empowers Users - Paternalism Does Not

Raising awareness of unfavorable behavior is the first important step to avoiding these behaviors in the future. Users themselves define what is *unfavorable* for them. Developers should only deploy machine learning models that work with user data while fostering awareness of the underlying mechanisms [23]. Potentially resulting privacy concerns, skepticism towards technology, and a feeling of technology paternalism might lead users not to adopt the system.

Besides intervention and prevention strategies, future work could use our MPRH-prediction approach to raise awareness among the users about behaviors when in a MPRH and situations that favor the likeliness of MPRH to develop. Interface concepts from the domain of user-centered explainable AI could be incorporated to explain why the system detected a MPRH.

[6] https://support.apple.com/en-gb/guide/iphone/iphbfa595995/ios

A rabbit hole prevention or intervention approach must avoid technology paternalism, that is, not force the user to do something, e.g., by forcing app usage time limits or completely stopping phone usage. Instead, users should be given suggestions only and control over all measures with the option to break out of the limitations at every time.

Finally, such prediction might induce privacy considerations if users obtain the feeling that their device knows more about themselves than they do without understanding why they know more. The presented sensing, detection, and prediction concepts are backed by sensitive user data, that should be handled with care. Transparency on what data is collected and how conclusions are drawn, as well as giving the users the sense of being in control of their data, are crucial [4].

7 CONCLUSION

In this paper, we investigated the MPRH by employing a mixed-method approach. First, we ran a two-week, in-the-wild, smartphone use tracking study of 21 users to understand their habits and reasons for falling into the rabbit hole. We then conduct an experts' focus group to derive a MPRH definition and identify key factors that draw users into MPRH. With this, we built a classifier to predict rabbit hole smartphone use behavior. In the future, we envision that our classifier will inspire many new user interfaces which support the user to get out of the rabbit hole or prevent them from falling into them. Our RHT app and prediction models are openly available for researchers on GitHub[7].

REFERENCES

[1] Abdulaziz Alblwi, John McAlaney, Dena Ahmed S. Al Thani, Keith Phalp, and Raian Ali. 2021. Procrastination on social media: predictors of types, triggers and acceptance of countermeasures. *Social Network Analysis and Mining* 11, 1 (2021), 1–18. https://doi.org/10.1007/s13278-021-00727-1

[2] Ricardo Baeza-Yates, Di Jiang, Fabrizio Silvestri, and Beverly Harrison. 2015. Predicting The Next App That You Are Going To Use. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (Shanghai, China) *(WSDM '15)*. Association for Computing Machinery, New York, NY, USA, 285–294. https://doi.org/10.1145/2684822.2685302

[3] Joseph B. Bayer, Scott W. Campbell, and Rich Ling. 2016. Connection Cues: Activating the Norms and Habits of Social Connectedness. *Communication Theory* 26, 2 (2016), 128–149. https://doi.org/10.1111/comt.12090

[4] Florian Bemmann, Maximiliane Windl, Jonas Erbe, Sven Mayer, and Heinrich Hussmann. 2022. The Influence of Transparency and Control on the Willingness of Data Sharing in Adaptive Mobile Apps. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 189 (sep 2022), 26 pages. https://doi.org/10.1145/3546724

[5] Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex (Sandy) Pentland. 2014. Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits. In *Proceedings of the 22nd ACM International Conference on Multimedia* (Orlando, Florida, USA) *(MM '14)*. Association for Computing Machinery, New York, NY, USA, 477–486. https://doi.org/10.1145/2647868.2654933

[6] Andrey Bogomolov, Bruno Lepri, and Fabio Pianesi. 2013. Happiness recognition from mobile phone data. In *2013 International Conference on Social Computing*. IEEE, New York, NY, USA, 790–795. https://doi.org/10.1109/SocialCom.2013.118

[7] Matthias Böhmer, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Bauer. 2011. Falling Asleep with Angry Birds, Facebook and Kindle: A Large Scale Study on Mobile Application Usage. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (Stockholm, Sweden) *(MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 47–56. https://doi.org/10.1145/2037373.2037383

[8] Hong Cao and Miao Lin. 2017. Mining smartphone data for app usage prediction and recommendations: A survey. *Pervasive and Mobile Computing* 37 (2017), 1–22. https://doi.org/10.1016/j.pmcj.2017.01.007

[9] Lewis Carroll. 1865. *Alice's Adventures in Wonderland*. Macmillan, New York, NY, USA.

[10] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357. https://doi.org/10.1613/jair.953

[11] Francesco Chiossi, Luke Haliburton, Changkun Ou, Andreas Martin Butz, and Albrecht Schmidt. 2023. Short-Form Videos Degrade Our Capacity to Retain Intentions: Effect of Context Switching On Prospective Memory. In *Proceedings*

[7] https://github.com/mimuc/mobilehci23-mobile-phone-rabbit-hole

of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 30, 15 pages. https://doi.org/10.1145/3544548.3580778

[12] Hyunsung Cho, DaEun Choi, Donghwi Kim, Wan Ju Kang, Eun Kyoung Choe, and Sung-Ju Lee. 2021. Reflect, Not Regret: Understanding Regretful Smartphone Use with App Feature-Level Analysis. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 456 (oct 2021), 36 pages. https://doi.org/10.1145/3479600

[13] Seungwoo Choi, Hayeon Jeong, Minsam Ko, and Uichin Lee. 2016. LockDoll: Providing Ambient Feedback of Smartphone Usage within Social Interaction. 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, 1165–1172. https://doi.org/10.1145/2851581.2892445

[14] Roger Collier. 2016. Mental health in the smartphone era. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 188, 16 (2016), 1141–1142. https://doi.org/10.1503/cmaj.109-5336

[15] Trinh-Minh-Tri Do and Daniel Gatica-Perez. 2010. By Their Apps You Shall Understand Them: Mining Large-Scale Patterns of Mobile Phone Usage. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/1899475.1899502

[16] Jie Du, Guido M. van Koningsbruggen, and Peter Kerkhof. 2018. A brief measure of social media self-control failure. *Computers in Human Behavior* 84 (2018), 68–75. https://doi.org/10.1016/j.chb.2018.02.002

[17] Ivaylo Durmonski. 2022. Internet Rabbit Holes: Why They Are Bad and How To Make Use Of Them. Website. https://durmonski.com/well-being/internet-rabbit-holes/ [Online; accessed 03.05.2023].

[18] Nir Eyal. 2014. *Hooked: How to Build Habit-Forming Products*. Portfolio, New York.

[19] Fabianne Rico. 2016. The Good, the Bad, and the Infinite: Mobile Engagement Loops. https://medium.com/@beimpekable/the-good-the-bad-and-the-infinite-mobile-engagement-loops-a181e7831cc. [Online; accessed 18.05.2022].

[20] Denzil Ferreira, Jorge Goncalves, Vassilis Kostakos, Louise Barkhuus, and Anind K. Dey. 2014. Contextual Experience Sampling of Mobile Application Micro-Usage. 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, 91–100. https://doi.org/10.1145/2628363.2628367

[21] Robert Fisher and Reid Simmons. 2011. Smartphone interruptibility using density-weighted uncertainty sampling with reinforcement learning. *Proceedings - 10th International Conference on Machine Learning and Applications, ICMLA 2011* 1 (2011), 436–441. https://doi.org/10.1109/ICMLA.2011.128

[22] Amyrae Fox, Robert Kaufman, and James D Hollan. 2020. Falling Down the Rabbit Hole: Exploratory Search and The Case for Personal Information Environments. (2020). https://hci.ucsd.edu/220/RabbitHole.pdf

[23] Gabriella M Harari. 2020. A process-oriented approach to respecting privacy in the context of mobile phone tracking. *Current opinion in psychology* 31 (2020), 141–147.

[24] Gunnar Harboe and Elaine M. Huang. 2015. Real-World Affinity Diagramming Practices: Bridging the Paper-Digital Gap. In *Proc. 33rd Annual ACM Conf. Human Factors in Computing Systems*. ACM, New York, NY, USA, 95–104. https://doi.org/10.1145/2702123.2702561

[25] Jennifer A Harriger, Joshua A Evans, J Kevin Thompson, and Tracy L Tylka. 2022. The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms. *Body Image* 41 (2022), 292–297.

[26] Trevor Haynes. 2018. Dopamine, Smartphones and You: A battle for your time. https://sitn.hms.harvard.edu/flash/2018/dopamine-smartphones-battle-time/. [Online; accessed 18.05.2022].

[27] Maxi Heitmayer. 2021. "It's Like Being Gone For A Second": Using Subjective Evidence-Based Ethnography to Understand Locked Smartphone Use Among Young Adults. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 49, 17 pages. https://doi.org/10.1145/3447526.3472026

[28] Alexis Hiniker, Shwetak N. Patel, Tadayoshi Kohno, and Julie A. Kientz. 2016. Why Would You Do That? Predicting the Uses and Gratifications behind Smartphone-Usage Behaviors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 634–645. https://doi.org/10.1145/2971648.2971762

[29] Tanya Horeck, Mareike Jenner, and Tina Kendall. 2018. On binge-watching: Nine critical propositions. *Critical Studies in Television* 13, 4 (2018), 499–504. https://doi.org/10.1177/1749602018796754

[30] Ke Huang, Chunhui Zhang, Xiaoxiao Ma, and Guanling Chen. 2012. Predicting Mobile Application Usage Using Contextual Information. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 1059–1065. https://doi.org/10.1145/2370216.2370442

[31] Adrian Rauchfleisch Jonas Kaiser. 2019. Website. Retrieved January 24, 2023 from https://policyreview.info/articles/news/implications-venturing-down-rabbit-hole/1406.

[32] Daniel Kardefelt-Winther. 2014. The moderating role of psychosocial well-being on the relationship between escapism and excessive online gaming. *Computers in Human Behavior* 38 (2014), 68–74. https://doi.org/10.1016/j.chb.2014.05.020

[33] Daniel Kardefelt-Winther. 2017. Conceptualizing Internet use disorders: Addiction or coping process? *Psychiatry and Clinical Neurosciences* 71, 7 (2017), 459–466. https://doi.org/10.1111/pcn.12413

[34] Kleomenis Katevas, Ioannis Arapakis, and Martin Pielot. 2018. Typical Phone Use Habits: Intense Use Does Not Predict Negative Well-Being. 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 11, 13 pages. https://doi.org/10.1145/3229434.3229441

[35] Kleomenis Katevas, Ilias Leontiadis, Martin Pielot, and Joan Serrà. 2017. Practical Processing of Mobile Sensor Data for Continual Deep Learning Predictions. In *Proceedings of the 1st International Workshop on Deep Learning for Mobile Systems and Applications* (Niagara Falls, New York, USA) *(EMDL '17)*. Association for Computing Machinery, New York, NY, USA, 19–24. https://doi.org/10.1145/3089801.3089802

[36] Patrick Keeffe. 2019. Compulsively Check Your Smartphone? Knowing Why Can Help You Stop. https://www.healthline.com/health-news/compulsive-about-checking-your-smartphone-heres-how-to-kick-the-habit [Online; accessed 05.07.2023].

[37] Asma Khatoon and Peter Corcoran. 2017. Privacy concerns on Android devices. In *2017 IEEE International Conference on Consumer Electronics (ICCE'17)*. IEEE, New York, NY, USA, 149–152. https://doi.org/10.1109/ICCE.2017.7889265

[38] Larry Kilham. 2016. *The Digital Rabbit Hole: How we are becoming captive in the digital universe and how to stimulate creativity, education, and recapture our humanity.* CreateSpace Independent Publishing Platform, Scotts Valley, California, US.

[39] Joseph Kim. 2014. What Is The Compulsion Loop? https://gameanalytics.com/blog/the-compulsion-loop-explained/. https://gameanalytics.com/blog/the-compulsion-loop-explained/ [Online; accessed 18.05.2022].

[40] Katrin B. Klingsieck. 2013. Procrastination when good things don't come to those who wait. *European Psychologist* 18, 1 (2013), 24–34. https://doi.org/10.1027/1016-9040/a000138

[41] 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

[42] Robert LiKamWa, Yunxin Liu, Nicholas D. Lane, and Lin Zhong. 2013. MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services* (Taipei, Taiwan) *(MobiSys '13)*. Association for Computing Machinery, New York, NY, USA, 389–402. https://doi.org/10.1145/2462456.2464449

[43] Hong Lu, Denise Frauendorfer, Mashfiqui Rabbi, Marianne Schmid Mast, Gokul T. Chittaranjan, Andrew T. Campbell, Daniel Gatica-Perez, and Tanzeem Choudhury. 2012. StressSense: Detecting Stress in Unconstrained Acoustic Environments Using Smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 351–360. https://doi.org/10.1145/2370216.2370270

[44] Kai Lukoff, Alexis Hiniker, Colin M. Gray, Arunesh Mathur, and Shruthi Sai Chivukula. 2021. What Can CHI Do About Dark Patterns?. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 102, 6 pages. https://doi.org/10.1145/3411763.3441360

[45] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J. Vera Liao, James Choi, Kaiyue Fan, Sean A. Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. 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 368, 17 pages. https://doi.org/10.1145/3411764.3445467

[46] 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 (March 2018), 22:1–22:26. https://doi.org/10.1145/3191754

[47] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[48] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotka, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2020. *'I Just Want to Hack Myself to Not Get Distracted': Evaluating Design Interventions for Self-Control on Facebook.* Association for Computing Machinery, New York, NY, USA, 1–15. https://doi-org.emedien.ub.uni-muenchen.de/10.1145/3313831.3376672

[49] Jeremy Marty-Dugas, Brandon C. W. Ralph, Jonathan M. Oakman, and Daniel Smilek. 2018. The relation between smartphone use and everyday inattention. *Psychology of Consciousness: Theory, Research, and Practice* 5, 1 (March 2018), 46–62. https://doi.org/10.1037/cns0000131

[50] Aleksandar Matic, Martin Pielot, and Nuria Oliver. 2015. Boredom-Computer Interaction: Boredom Proneness and the Use of Smartphone. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 837–841. https://doi.org/10.1145/2750858.2807530

[51] Abhinav Mehrotra, Sandrine R. Müller, Gabriella M. Harari, Samuel D. Gosling, Cecilia Mascolo, Mirco Musolesi, and Peter J. Rentfrow. 2017. Understanding the Role of Places and Activities on Mobile Phone Interaction and Usage Patterns. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 84 (sep 2017), 22 pages. https://doi.org/10.1145/3131901

[52] Emiliano Miluzzo, Michela Papandrea, Nicholas D Lane, Hong Lu, and Andrew T Campbell. 2010. Pocket, Bag, Hand, etc. - Automatically Detecting Phone Context through Discovery. In *Proceedings of the 2010 Int. Workshop Sensing for App Phones (PhoneSense'10)*. 21–25.

[53] Alberto Monge Roffarello and Luigi De Russis. 2022. Towards Understanding the Dark Patterns That Steal Our Attention. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 274, 7 pages. https://doi.org/10.1145/3491101.3519829

[54] Mark Muraven and Elisaveta Slessareva. 2003. Mechanisms of self-control failure: Motivation and limited resources. *Pers Soc Psychol Bull* 29, 7 (July 2003), 894–906. https://doi.org/10.1177/0146167203029007008

[55] Elizabeth L. Murnane, Xin Jiang, Anna Kong, Michelle Park, Weili Shi, Connor Soohoo, Luke Vink, Iris Xia, Xin Yu, John Yang-Sammataro, Grace Young, Jenny Zhi, Paula Moya, and James A. Landay. 2020. Designing Ambient Narrative-Based Interfaces to Reflect and Motivate Physical Activity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376478

[56] Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita. 2012. Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing* 16, 1 (2012), 105–114. https://doi.org/10.1007/s00779-011-0412-2

[57] Elliot Panek. 2014. Left to Their Own Devices: College Students' "Guilty Pleasure" Media Use and Time Management. *Communication Research* 41, 4 (2014), 561–577. https://doi.org/10.1177/0093650213499657

[58] Shwetak N. Patel, Julie A. Kientz, Gillian R. Hayes, Sooraj Bhat, and Gregory D. Abowd. 2006. Farther than you may think: An empirical investigation of the proximity of users to their mobile phones. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4206 LNCS (2006), 123–140. https://doi.org/10.1007/11853565_8

[59] Tiziano Piccardi, Martin Gerlach, and Robert West. 2022. Going Down the Rabbit Hole: Characterizing the Long Tail of Wikipedia Reading Sessions. In *Companion Proceedings of the Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 1324–1330. https://doi.org/10.1145/3487553.3524930

[60] Martin Pielot. 2014. Large-Scale Evaluation of Call-Availability Prediction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 933–937. https://doi.org/10.1145/2632048.2632060

[61] Martin Pielot, Rodrigo de Oliveira, Haewoon Kwak, and Nuria Oliver. 2014. Didn't You See My Message? Predicting Attentiveness to Mobile Instant Messages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3319–3328. https://doi.org/10.1145/2556288.2556973

[62] Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When Attention is Not Scarce - Detecting Boredom from Mobile Phone Usage. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 825–836. https://doi.org/10.1145/2750858.2804252

[63] Leonard Reinecke and Wilhelm Hofmann. 2016. Slacking Off or Winding Down? An Experience Sampling Study on the Drivers and Consequences of Media Use for Recovery Versus Procrastination. *Human Communication Research* 42, 3 (2016), 441–461. https://doi.org/10.1111/hcre.12082

[64] Alberto Roffarello and Luigi De Russis. 2019. The Race Towards Digital Wellbeing: Issues and Opportunities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow, Scotland Uk, 1–14. https://doi.org/10.1145/3290605.3300616

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

[66] Anna Schnauber-Stockmann, Adrian Meier, and Leonard Reinecke. 2018. Procrastination out of Habit? The Role of Impulsive Versus Reflective Media Selection in Procrastinatory Media Use. *Media Psychology* 21, 4 (2018), 640–668. https://doi.org/10.1080/15213269.2018.1476156

[67] Ramona Schoedel, Michelle Oldemeier, Léonie Bonauer, and Larissa Sust. 2022. Dataset for: Systematic categorisation of 3091 smartphone applications from a large-scale smartphone sensing dataset. https://psycharchives.org/en/item/37c9a04d-4785-4f0f-8e3d-088f040613b5. https://doi.org/10.23668/psycharchives.5362 [Online; accessed 17.05.2022].

[68] Kathryn Schulz. 2015. Falling Down the Rabbit Hole: Exploratory Search and The Case for Personal Information Environments. (2015). https://www.newyorker.com/culture/cultural-comment/the-rabbit-hole-rabbit-hole

[69] Zhihao Shen, Kang Yang, Wan Du, Xi Zhao, and Jianhua Zou. 2019. DeepAPP: A Deep Reinforcement Learning Framework for Mobile Application Usage Prediction. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems* (New York, New York) *(SenSys '19)*. Association for Computing Machinery, New York, NY, USA, 153–165. https://doi.org/10.1145/3356250.3360038

[70] Choonsung Shin and Anind K. Dey. 2013. Automatically Detecting Problematic Use of Smartphones. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 335–344. https://doi.org/10.1145/2493432.2493443

[71] Jeremiah Smith, Anna Lavygina, Jiefei Ma, Alessandra Russo, and Naranker Dulay. 2014. Learning to Recognise Disruptive Smartphone Notifications. 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, 121–124. https://doi.org/10.1145/2628363.2628404

[72] Andrew Sullivan. 2016. Website. Retrieved January 24, 2023 from https://nymag.com/intelligencer/2016/09/andrew-sullivan-my-distraction-sickness-and-yours.html.

[73] Nađa Terzimehić and Sarah Aragon-Hahner. 2022. I Wish I Had: Desired Real-World Activities Instead of Regretful Smartphone Use. In *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia* (Lisbon, Portugal) *(MUM '22)*. Association for Computing Machinery, New York, NY, USA, 47–52. https://doi.org/10.1145/3568444.3568465

[74] Nađa Terzimehić, Luke Haliburton, Philipp Greiner, Albrecht Schmidt, Heinrich Hussmann, and Ville Mäkelä. 2022. MindPhone: Mindful Reflection at Unlock Can Reduce Absentminded Smartphone Use. In *Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1818–1830. https://doi.org/10.1145/3532106.3533575

[75] Jonathan A. Tran, Katie S. Yang, Katie Davis, and Alexis Hiniker. 2019. Modeling the Engagement-Disengagement Cycle of Compulsive Phone 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, 1–14. https://doi.org/10.1145/3290605.3300542

[76] Niels van Berkel, Chu Luo, Theodoros Anagnostopoulos, Denzil Ferreira, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. 2016. A Systematic Assessment of Smartphone Usage Gaps. 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, 4711–4721. https://doi.org/10.1145/2858036.2858348

[77] Mariek M P Vanden Abeele. 2020. Digital Wellbeing as a Dynamic Construct. *Communication Theory* 31, 4 (10 2020), 932–955. https://doi.org/10.1093/ct/qtaa024

[78] Alicia Wanless. 2016. TED Talk: We Are All Alice Now: Falling Down a Digital Rabbit Hole. https://www.linkedin.com/pulse/we-all-alice-now-falling-down-digital-rabbit-hole-alicia-wanless

[79] Ivy Wigmore. 2013. What is: rabbit hole? https://www.techtarget.com/whatis/definition/rabbit-hole. Retrieved January 24, 2023.

[80] Ivy Wigmore. 2019. Dopamine-driven feedback loop. https://whatis.techtarget.com/definition/dopamine-driven-feedback-loop. https://whatis.techtarget.com/definition/dopamine-driven-feedback-loop [Online; accessed 18.05.2022].

[81] James Williams. 2018. *Stand out of our light: freedom and resistance in the attention economy.* Cambridge University Press, Cambridge, United Kingdom.

[82] Kaitlin Woolley and Marissa A Sharif. 2022. Down a Rabbit Hole: How Prior Media Consumption Shapes Subsequent Media Consumption. *Journal of Marketing Research* 59, 3 (2022), 453–471.

[83] Tong Xia, Yong Li, Jie Feng, Depeng Jin, Qing Zhang, Hengliang Luo, and Qingmin Liao. 2020. DeepApp: Predicting Personalized Smartphone App Usage via Context-Aware Multi-Task Learning. *ACM Trans. Intell. Syst. Technol.* 11, 6, Article 64 (oct 2020), 12 pages. https://doi.org/10.1145/3408325

[84] Tingxin Yan, David Chu, Deepak Ganesan, Aman Kansal, and Jie Liu. 2012. Fast App Launching for Mobile Devices Using Predictive User Context. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services* (Low Wood Bay, Lake District, UK) *(MobiSys '12)*. Association for Computing Machinery, New York, NY, USA, 113–126. https://doi.org/10.1145/2307636.2307648

[85] Timothy Yang, Galen Andrew, Hubert Eichner, Haicheng Sun, Wei Li, Nicholas Kong, Daniel Ramage, and Françoise Beaufays. 2018. Applied federated learning: Improving google keyboard query suggestions. *arXiv preprint arXiv:1812.02903* (2018). https://doi.org/10.48550/arXiv.1812.02903

[86] Mingrui Ray Zhang, Kai Lukoff, Raveena Rao, Amanda Baughan, and Alexis Hiniker. 2022. Monitoring Screen Time or Redesigning It? Two Approaches to Supporting Intentional Social Media Use. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 60, 19 pages. https://doi.org/10.1145/3491102.3517722

[87] Sha Zhao, Julian Ramos, Jianrong Tao, Ziwen Jiang, Shijian Li, Zhaohui Wu, Gang Pan, and Anind K. Dey. 2016. Discovering Different Kinds of Smartphone Users through Their Application Usage Behaviors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 498–509. https://doi.org/10.1145/2971648.2971696

Received January 2023; revised May 2023; accepted June 2023