Multiple Device Users' Actual and Ideal Cross-Device Usage for Multi-Stage Notification-Interactions: An ESM Study Addressing the Usage Gap and Impacts of Device Context

Fang-Ching, Tseng
r11922012@ntu.edu.tw
National Taiwan University
Taipei, Taiwan
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Zih-Yun, Chiou
dobelle.11@nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Ho-Hsuan, Chuang
stonechuang.ee07@nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Li-Ting, Su
liitiisu.cs10@nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Yong-Han, Lin
alberta.cs07@nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Yu-Rou, Lin
wn4doe.cs09@nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Yi-Chi, Lee
yichi170.cs08@nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Peng-Jui, Wang
james900425.cs08@nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Uei-Dar, Chen
udchen.cs08@nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Yung-Ju, Chang*
armuro@nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

ABSTRACT

People nowadays can use multiple devices to interact with notifications, whether via noticing, glancing, reading, or acting upon them. Prior research has focused on actual usage or on device preferences. However, users' ideal experience of cross-device notification-interaction might differ from their current practices (due to situational limitations) and/or across the four notification-interaction stages. We therefore conducted an experience-sampling method study with multi-device users to investigate these gaps and the influence of device context. Our results reveal that nearly half of the time, the non-phone devices the participants had ranked as their top preferences for notification-interaction were not actually used, due to the devices' context. Beyond device context, the participants' choices of devices for notification-interaction were heavily determined by 1) their preferences that particular notification-interaction stages to take place (or not) on particular devices; and 2) the device on which they had undertaken the former stage.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**.

KEYWORDS

Notifications; Multi-Device; Experience Sampling Method

*Corresponding author.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04...$15.00
https://doi.org/10.1145/3544548.3580731

ACM Reference Format:
Fang-Ching, Tseng, Zih-Yun, Chiou, Ho-Hsuan, Chuang, Li-Ting, Su, Yong-Han, Lin, Yu-Rou, Lin, Yi-Chi, Lee, Peng-Jui, Wang, Uei-Dar, Chen, and Yung-Ju, Chang. 2023. Multiple Device Users' Actual and Ideal Cross-Device Usage for Multi-Stage Notification-Interactions: An ESM Study Addressing the Usage Gap and Impacts of Device Context. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3580731

1 INTRODUCTION

Amid advances in mobile Internet and the increasing maturity of modern mobile operating systems, smartphones and a growing variety of applications (apps) available for use on these phones have become increasingly popular. In recent years, smartphones have been joined in the market by an array of other popular devices such as tablets and wearables like smartwatches and smart wristbands [6, 9, 46]. Thus, people can now receive information from apps via all of these devices' notification systems, and in theory, this enables them to flexibly choose between devices that they think suitable for specific tasks [29, 60]. For example, multi-device users can send emails on one device but receive the replies on another, and access messaging notifications via all devices but attend to and respond to them on whichever device they prefer.

However, recent research has found that different types of devices are not perceived as equally suitable or equally preferred for receiving all types of notifications [25, 41, 76] – probably because these different types serve different purposes [38], and thus carry different types of information that demand varying amounts of attention, levels of engagement, and subsequent actions. For example, addressing some notifications – e.g., weather reports – may only demand that device users glance at them [38], whereas a news notification is likely to demand more user attention in the form of reading, and a messaging notification may prompt users not only to read a message but also to provide a response, either via typing or speaking. The fact that devices have screens of different sizes, affordances for different kinds of input and output, and software for supporting different types of interaction with notifications [24, 48] is another likely reason that different device types are perceived as suitable or unsuitable for performing particular kinds of interaction with notifications [25, 76]. However, prior research on multi- and cross-device user behavior has not taken sufficient account of how different phases of notification-interaction [67] (N-I; e.g., merely glancing at vs. reading) might affect users' preferences about particular devices for N-I purposes.

Moreover, people's *actual* usage of devices for N-I may differ from their *ideal* usage of the same device due to various factors, including but not limited to the context of the device, e.g., whether it is accessible to them visually [15] or physically reachable [76] at that moment. Some of these contexts, however, may be influential (or not) only to specific interactions on specific devices, due to the aforementioned differences in devices' attributes. For example, reading a message on a computer demands that the computer be visible, whereas feeling vibration from a wearable device [9, 26, 56] and hearing an alert from a phone [33] do not demand any such minimum level of visibility.

As a result, when a user's ideal device for N-I is not in its suitable context for performing a particular stage of N-I at a particular moment – e.g., the user is not able to see the device's screen when wanting to glance at the notification, or not able to reply because the device is not within easy reach – it is reasonable to expect that the user will defer interaction until s/he can access the ideal device, or use an alternative device. Nevertheless, the extent to which devices' contexts are related to their owners' ideal usage of them for performing a particular N-I stage, and the extent to which people's actual usage of devices deviates from such ideal usage, remains unclear. We regard filling these gaps as essential to creating an effective multi-device notification ecosystem.

To do so, we have adapted a four-stage typology of the N-I process from Turner et al. [67], which includes four N-I stages – Notice, Glance, Read, and Act – and ask the following research questions:

- RQ1: What are multi-device users' ideal and actual usage of devices for proceeding to each of the four N-I stages, respectively?
- RQ2: At which of the four N-I stages, and to what extent, does users' actual usage of devices deviate from their ideal usage?
- RQ3: How does such ideal usage and any gaps between them relate to device context, respectively?

To answer these three research questions, we conducted an experience-sampling method (ESM) study with 31 multi-device users. The three main contributions of this paper are as follows.

- It shows how multi-device users' ideal usage of their devices along four N-I stages are associated with various device contexts.
- It reveals where gaps between their actual and ideal usage of their devices were most likely to occur within the four-stage N-I process
- It highlights that users' preferences, device contexts, and tendency to persist through the stages of the N-I process using the same device all play a vital role in their choice of devices for N-I.

2 RELATED WORK

2.1 Notification Management and Multi-Stages in Interaction Process

The prevalence of smartphone notifications has attracted considerable research attention. Generally, three lines of research have emerged. One line of research concerns the negative impacts of notifications on people, such as raising their negative emotions [52, 77, 79], causing interruption/disturbance to their task at hand [1, 33, 36, 65], resulting in inattention [33], and so on.

The second line of research concerns people's perceptions and various interactions with their notifications. Regarding perception, several studies have suggested smartphone users' preference of communication-related notification over other types [43, 53, 55, 59]; others suggested that users valued urgent, important, and attractive notifications than otherwise [38, 43, 55, 59, 74]. Regarding interaction with notifications, some researchers investigated smartphone users' specific notification-management practices, such as snoozing/deferring them [3, 49, 73], dismissing them [55], adjusting the alert modality of notifications [12, 33], and deciding whether to attend to them after speculating about their sources [11].

The third line of research concerns factors that influence people's receptivity to notifications. Factors that have been identified as influential include but not limited to: activity [22, 43, 51, 62], psychological status [44, 52, 54, 61], social environment [7, 64], notification content [22, 75], the sender of notification [35, 43].

In addition to exploring influential factors, prior research has also successfully predicted people's high-receptivity moments for receiving notifications, including at interruptible moments [27, 47, 50, 68], breakpoints [1, 27, 28, 47, 49], transitions [27], bored moments [54], sometimes referred to as opportune moments [21, 42, 50].

Moreover, other researchers looked at specific receptivity measures and suggested that they represent distinct interactions with notifications and thus should be treated differently. One commonly adopted distinction is between attentiveness (how often and/or quickly users can attend to notifications) and responsiveness (how often or quickly users can respond to notifications). For example, Chang et al. [12] found that their study participants' attentiveness to messaging notifications varied between ringer modes, but their responsiveness to the same notifications did not. Lee et al.'s [35] findings showed that attentiveness was positively correlated to sender closeness, but this correlation was absent for responsiveness. Wu et al. argued that the conceptual difference between attentiveness and responsiveness depends on whether users prefer to attend and respond to notifications at the same time [78]. Chang et al. [10], who investigated the perceived opportune moment for interacting with notifications, showed that participants' perceptions of the opportune moment for reading and acting upon notifications varied in their preliminary investigation. Turner et al. [66, 67] proposed a four-stage response process model, including steps of gaining the users' attention (react) before the three decisions of which the user decides whether to fixate on the notification (focus), whether the user views the content (read) and whether they act on the notification (act). They further proposed a three-stage model [66], consisting of stages of reachability, engage-ability, and lastly, receptivity.

Lastly, several studies specifically focused on studying/predicting attentiveness and responsiveness of notifications, respectively. Mashhadi et al. [40] indicated that a notification was 12 times more likely to be immediately attended to when they came with notification alerts. Dingler et al. [18] found that smartphone users are attentive to their notifications in 70% of their wake hours, and their inattentiveness sessions would last for merely five minutes in 75% of occurrences. Mehrotra et al. showed that the response time to notifications covering the foreground is twice as fast as those only appended to the notification drawer [43].

2.2 Multi-Device Experiences

Previous research showed that most multi-device users have a positive attitude toward multi-device interaction [46], and that they are likely to utilize several devices to complete their work for improving their overall productivity and performance [29]. The phenomenon of utilizing multiple devices for performing various tasks has attracted scholarly attention, with a focus on investigating multi-device usage and experiences. For example, Yuan et al. [80] introduced several multi-device usage patterns, including partitioning tasks, integrating multi-device usage, cloning tasks to other devices, expanding tasks to multiple devices, and migrating across devices. Jokela et al. [29], on the other hand, distinguished four patterns: sequential use, resource lending, related parallel use, and unrelated parallel use, and it was suggested that multi-device users' perceptions of task's characteristics played a vital role in their choices of devices [29].

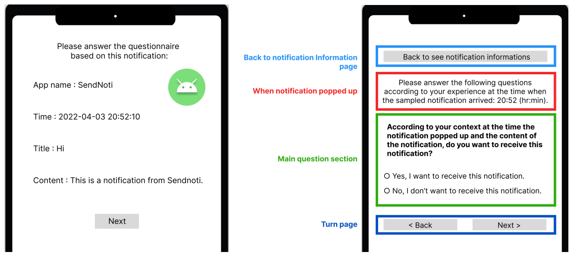
The perceived fit between tasks and devices, expectedly, has become one of the research focus in the multi-device literature. For example, Santosa et al. [60] in their study found that smartphones were usually used for managing small amounts of information and non-primary work, such as messaging, checking calendars, and listening to music [29], whereas other studies showed that computers were used mainly for complex tasks related to work or study [8, 29, 31, 45]. Smartwatches, on the other hand, were commonly regarded as a substitution for smartphones [6]; however Cecchinato et al. [9] suggested that smartphones were also preferred to receive proactive recommendations and to be used to quickly respond to notifications over the other devices. In contrast, research showed that tablets were primarily used by their users for consuming content and entertainment such as reading, playing games, and watching videos [17, 29, 45, 80]. However, occasionally they were also used to cooperate with larger devices as companion devices [29].

In addition, physical environment and social context have also been found to influence users' choices of devices [29]. For example, research has found that social unobtrusiveness and acceptability has made users choose smartphone over computers due to the latter's larger size that would capture more attention [29, 48]. Larger devices are also less preferred than smaller ones when the physical environment has specific space constraints such as in a kitchen [29, 31].

Several studies have also investigated multi-device users' preferences and experiences of receiving notifications, including the factors that would influence their choices of devices [25, 41, 76]. Specifically, Weber et al. [76] showed that the device on which their study participants preferred to receive notifications depended on their locations, the number of people nearby, and their proximity to the devices. For example, it was found that they preferred smartphones over other devices for receiving notifications because smartphones were usually the closest device to them. Likewise, Mehrotra et al. [41] suggested that users' physical activity, location, network connectivity, applications that posted notifications, and devices they used for previous notifications would influence their users' behaviors in handling notifications. Finally, prior research has also sought to develop intelligent systems that identify when and on which devices the system should deliver notifications [15, 41, 47].

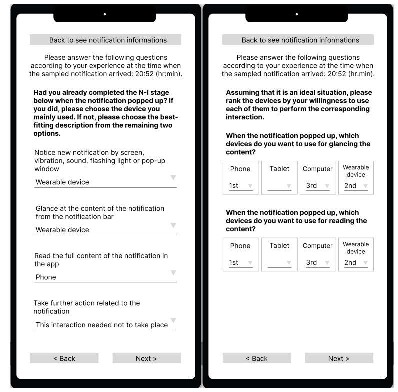
Nevertheless, although delivering notifications to multiple devices brings certain convenience to users, it sometimes also introduces more burden. For example, sending identical notifications across devices introduces continuous interruptions and duplicate notifications that can distress users [15]. Having more devices also means more configuration and coordination among devices, which is a trouble for people who do not often change notification settings [71]. In addition, not only that switching between different input and output modalities on multiple devices can lead to users' considerable overhead [24, 58], managing information across multiple devices sometimes can be also burdensome[17].

However, it is unclear how and when multi-device users' actual usage deviates from their ideal usage of them, and how these gaps relate to the device context of their devices. The current paper sheds light on these issue.



(a) (b)

Figure 1: The ESM questionnaire interface, with information about the sampled notification (a) and an example of an ESM questionnaire page (b)



(a) (b)

Figure 2: ESM questions about the participants' actual (a) and ideal (b) device usage

3 METHODOLOGY

3.1 ESM Study

To answer our research questions, we needed our research participants to accurately describe their daily experiences, interactions, and perceptions in their normal day-to-day environments [69]. To achieve this aim, we conducted an ESM study with 31 multi-device users. Further details are provided below.

3.1.1 ESM Research Instruments. We developed an Android research app, which 1) recorded notifications that arrived on its users' phones; 2) determined whether the arriving notification should be sampled and if so, delivered an ESM questionnaire pertaining to it (as described in section 3.1.2); and 3) logged phone-sensor data.

3.1.2 ESM Questionnaire. Each ESM questionnaire had several parts, as shown in Figure 1. On the first page, the respondent was shown information about the sampled notification. The ESM questionnaire *per se* began on the second page. On each page, the respondent saw a button that allowed him/her to return to p. 1 to review the information about the target notification if needed. Instructions below this button stated that the respondent should answer the ESM questions according to their experience *at the time when the sampled notification arrived*, the purpose being to include low-receptivity moments at which they could not or did not see their devices. In the ESM questionnaire, the participants were firstly asked if they wanted to receive the sampled notification regardless of what device it was received on. If the answer was negative, they skipped the rest of the questions in that ESM questionnaire.

Secondly, as shown in Figure 2, the respondent was asked about his/her actual usage of devices for performing each of the four N-I stages, i.e., *Notice* (noticing the alert from a device), *Glance* (skimming the notification to gain a rough idea of it), *Read* (reading the full content of the notification), and *Act* (acting upon the notification content). These stages were slightly modified from Turner et al.'s [67] four-stage response process, to make them more intuitive for our study participants to answer – an example being the change from React to Notice – and then tested for their understandability and straightforwardness in a pilot study. Specifically, the ESM respondents were asked to select the device on which they actually performed each relevant N-I stage. If they did not need to proceed to a particular N-I stage, or needed to but did not, they could choose the options *"This interaction did not need to take place"* and *"This interaction did not happen"*, respectively. It is notable that participants frequently went through multiple stages of the N-I process simultaneously or in a very short period of time, such as when they noticed, glanced at, and fully read a notification that arrived while they were using their phone. In such cases, they were expected to report using the same device for all of these actions.

Thirdly, the respondents provided context. Specifically, for each N-I stage, they first answered whether or not they wanted to proceed to that N-I stage *at the time they received the notification*, with answer options of *"Yes"*, *"No"*, and *"Not sure."* In addition, they reported the characteristics of the notification content, which included its importance [75], attractiveness [4], and task relevance [22]. Next, they reported the applicability of four dimensions of device context [15, 76] to each of their devices at the time, including *"Can see the device's screen"* (View), *"Can feel or hear the device"* (Sense), *"Can reach the device with my hands"* (Reach), and *"Can take out the device for use"* (Use). They also reported: whether each device had been used for anything within the preceding 10 minutes [41]; their location [20, 76]; their current activity [15, 19, 41, 43]; their level of engagement in that activity [35]; the number of other people around them [64]; their perceived privacy [9]; and their perceptions of the social norms prevailing in the exact time and place being discussed [7, 49, 64]. Our use of a 10-minute threshold in this case was inspired by Mehrotra et al.'s [41] finding that users' decisions about handling notifications are impacted by the device they used for handling the previous notification that arrived within the preceding 10 minutes. In explaining our device-use questions to the participants, we told them that as long as they were conscious of and could recall *any* action they performed on a device, even simply unlocking it or checking for notifications on a locked screen, they should consider themselves as using that device [57, 70].

Fourth, the ESM respondents reported their ideal usage of devices. For each of the three latter N-I stages that the participants reported wanting to proceed to, they ranked up to four *devices they deemed perfect* for that purpose. For the Notice stage, however, they were asked to choose all devices via which they wanted to notice the sampled notification, on the grounds that that stage is presumably triggered by alerts generated by multiple devices, and is not necessarily preceded by users' attendance to only one of them. Additionally, for each device they selected, they were asked for the way(s) in which they wanted to notice the notification on that device. The four options were *"screen turning-on or a pop-up"*, *"vibration"*, *"ringer"*, *"only show put in the notification drawer,"* and they could select all that applied.

3.1.3 ESM Mechanism. Upon installation of the research app, participants selected a time-window of at least 12 hours during which they were willing to receive ESM questionnaires each day. To diversify the time periods of the samples of participants' multi-device usage experience, the research app divided each day into three parts, in each of which participants would only see up to four ESM questionnaires. Our research app sampled notifications and decided whether to trigger an ESM questionnaire according to the following criteria: 1) the sampled notification was in the user-defined time-window; 2) the number of questionnaires the participant had finished that day was less than 12; 3) the participant had not submitted four ESM questionnaires in the ongoing segment of the day; 4) at least one hour had passed since the last questionnaire was submitted; and 5) the notification was not a duplicate of any previously sampled one. The research app also ignored system notifications and notifications of which the priority was set by the system as not higher than zero, since these notifications were arguably valuable to the user [59]. Our app also balanced the types of notification being sampled: i.e., when it recognized that a larger-than-average number of notifications from particular external apps had been received, it lowered the probability of those apps' notifications being sampled thereafter, until balance across external apps was restored. Whenever an arriving notification was sampled, the research app sent an ESM questionnaire immediately without any alert; this was to prevent an alert from drawing participants' attention to their phones, which could potentially increased their use of the phone and thus bias our results. All ESM questionnaires were dismissed if responses to them had not commenced within 30 minutes of their arrival, a time threshold adopted from prior research (e.g., [11]).

Due to the complexity of the questionnaire and the mechanism of the ESM, we tested our ESM with 14 pilot participants for six months, from December 2021 to May 2022. During this time, we monitored the time it took for participants to complete our ESM questionnaire and solicited their feedback on the difficulty and burden of answering its questions in different situations. We also tested whether the 30-minute threshold was adequate for participants to have interacted extensively with the sampled notifications without experiencing difficulty recalling their experiences. After iteratively adjusting the ESM questionnaire and mechanism, we were able to ensure that all pilot participants were able to complete it within three minutes on average, with a final completion time of 117 seconds (SD = 204, Mdn= 76), which was within the acceptable range for an ESM study, i.e. 2-3 minutes [5, 13, 16]). We then proceeded with the study after finalizing all adjustments.

3.2 Recruitment and Participants

We posted recruitment advertisements on several social-media platforms including Facebook, Instagram, and PTT (a terminal-based bulletin-board system based in Taiwan). We posted advertisements on social-media pages and PTT boards themed around wearable and other mobile technology. Each recruitment ad provided a link to a sign-up form, in which the respondents answered a set of questions, including the number of notifications they received daily, and the frequencies with which they used their phones, computers, tablets, and wearable devices. We selected participants based on the following criteria: 1) they had used at least three of the four aforementioned devices to receive and interact with notifications; 2) they received an average of more than 10 notifications per day; and 3) their frequency of phone use was at least once per two hours. This resulted in an initial pool of 33 participants, all of whom participated in the study for a full 14 days; however, two had their data removed because they told us that they had misunderstood the meaning of several ESM questions.

Therefore, the final cohort consisted of 31 participants, who ranged in age between 20 and 58 (M = 28.03; SD = 10.24). They included 19 students and 12 non-students, 19 females and 12 males. All owned smartphones; 30 also had laptop/desktop computers; 24, tablets; and 22, smart wristbands or smartwatches.

3.3 Study Procedure

Due to the COVID-19 pandemic, each participant was invited to an online pre-study meeting, in which the research team explained the study procedure and helped participants install the research app. All of them were invited to participate in optional online post-study interviews, and 30 accepted. In those interviews, the questions had three main themes: 1) how and why they switched devices across the four N-I stages; 2) how device contexts affected their device choices; and 3) the reasons behind the gaps between their actual and ideal usage, if any, in each N-I stage. Participants received compensation according to the number of ESM questionnaires they had finished, at the rate of NT$10 (at the time of submission, approximately US$0.32) for each questionnaire. If they participated in the post-study interview, they were given an additional NT$300 (about US$10). This study was approved by our university's Research Ethics Committee for Human Subject Protection.

3.4 Data Cleaning and Analysis

We received a total of 3,785 ESM responses from the original 33 participants. As mentioned earlier, two participants' responses (n=183) were removed, leaving 3,602 (min=55 , max=163, mean=116.19, median=115). We then removed 112 responses that participants reported as being incorrect, and/or within which the answers to two or more items contradict one another: e.g., that their device was not in any of the device contexts, but was in use for N-I at the same moment. After this cleaning process, the ESM responses from the remaining 31 participants numbered 3,490.

Participants' ESM responses in which they reported not wanting to receive the sampled notification at the sampled moment were also removed (n=1,743), on the grounds that they did not go on to reveal their actual and ideal device usage. This left a final dataset that consisted of 1,747 ESM responses. Among these responses, because the participants owned different sets of multiple devices, our analysis related to each kind of device only took account of those ESM responses from participants who owned that device. As a result, the numbers of ESM responses for each kind of device varied. Specifically, they were: phone, 1,747; computer, 1,601; tablet, 1,390; and wearable device, 1,184.

Given that our main target outcome was whether the participants chose a device or not (either actually or ideally), we mainly examined the effects of various factors on the odds of such an outcome's occurrence. Such factors included device type and N-I stage. To investigate the differences between levels of these categorical variables, we used mixed-effects logistic regression with the "lmerTest" [34] package in R software,[1] and included participant ID numbers as a random effect to account for individual differences among the participants. For each questionnaire, we labeled the actual device and the ideal device selected by participants in each N-I stage. For example, if a given participant selected their phone as their actual device and a computer as their ideal device in an ESM questionnaire, that questionnaire was separated into two data points, each containing a one-hot encoded vector indicating the device and the actual/ideal label (e.g., "actual", [1, 0, 0, 0], "ideal", [0, 0, 1, 0]). To examine the difference between ideal and actual usage of each device, we considered the actual/ideal label as a predictor of whether the device was selected (either as an actual device or the ideal device) or not. The equation used to achieve this was: $log(selected_i) = \beta * (L_i) + p_i$, where $selected_i$ is the odds of the specified device being selected as the top-1; β is the coefficient of the categorical variable of actual/ideal labels L_i; and p_i is the random intercept.

For analyzing relationships between device context and N-I stage, however, we used a different approach. This was because device contexts were not exclusive categories, and most of the time, more than two of them were jointly selected by the participants to describe their devices. To deal with this, we created contingency tables and used a chi-square test of independence to examine the association between device contexts and N-I stages.

For qualitative analysis, we transcribed our interview recordings and used affinity diagramming [39] to analyze the transcripts. We iteratively grouped and labeled affinity notes, discussing any that we were unsure about.From this bottom-up approach, several themes emerged, including: 1) potential causes of the reported gaps between actual and ideal usage, and 2) the participants' reasons for using or not using specific devices in particular contexts.

4 RESULTS

We first take a look at our participants' actual and ideal multi-device usage when proceeding to each of the notification-interaction (N-I) stages. We next reveal the gaps between actual and ideal usage on a stage-by-stage basis. Finally, we explore the participants' tendency to keep using the same device for proceeding to the next N-I stage.

4.1 Actual and Ideal Usage of Multiple Devices for Interacting with Notifications

4.1.1 Actual vs. Ideal Device Usage for Proceeding to Each Interaction Stage. This section looks at participants' actual and ideal usage of devices. For ideal device usage, in particular, because the ESM asked the participants to pick all ideal devices they wanted for the Notice N-I stage, in the analysis below, we mainly focus on the three latter N-I stages, for which the participants had to rank their ideal devices per stage, and on their single most-ideal device for proceeding to each intended N-I stage. Table 1 illustrates that phones were the most commonly used and preferred device for interacting with notifications across all N-I stages, while tablets were rarely used and not considered the ideal device by participants. Participants primarily used and preferred to use wearable devices to notice (actual: 9.3%, ideal: 35.5%) and glance at (8.6%; ideal: 9.0%) their notifications, and computers to read (6.3%; ideal: 7.1%) and act upon (7.1%; ideal: 8.0%) them. Overall, the trends in actual and ideal usage of devices were consistent. The rare selection of tablets as the preferred device may be due to individual participants' specific uses and preferences, as mentioned in the interviews. For example, several participants mentioned that they only used their tablets for specific purposes and others mentioned that they had not installed the relevant notification-emitting applications on their tablets. On the other hand, participants' ideal usage of devices was slightly higher than their actual usage. This was related to the fact that they were able to name their preferred devices regardless of whether they actually used them in the study. However, the differences between actual and ideal usage indicate that there were discrepancies between the devices participants preferred to use and the devices they actually used.

4.1.2 Ideal Device Usage in Different Device Contexts. Next, we divided participants' ideal usage of devices for each of the N-I stages according to which device contexts applied to the devices at the time. In Table 2, above, each cell contains the likelihood of the participants wanting to use a particular type of device for a given N-I stage when they perceived the device context to be appropriate to it, along with the number of instances. The five device contexts shown in the table rows are: View, Sense, Reach, Use, and lastly, None, defined as none of the four substantive categories named above being applicable to the device. The final row represents the overall likelihood of that device being seen as ideal for that N-I stage regardless of its device context.

The results showed that participants' ideal usage differed markedly across such contexts. First, they rarely assigned their top ranking to a device when none of the substantive device contexts applied to their situation. However, there were still such cases, suggesting gaps – wanting to use a device while they were inaccessible at the time. Because there were many instances of such inaccessibility, the overall likelihood of each device being chosen as the top ideal device was dragged down to quite a low level.

Phones were carried by the participants most of the time, and as a result, the overall likelihood of their being rated as the top ideal device in every N-I stage remained similar across device contexts. However, computers' and tablets' likelihoods of being chosen as the top ideal device were higher when participants reported they could see their screens than they could take them out for use. Specifically, tablets whose screens could be seen were named top ideal device in Glance 5.7% of the time, Read 6.8% of the time, and Act 9.4% of the time; and the parallel figures for computers whose screens could be seen were 16.4%, 25.1%, and 29.5%. All of these likelihoods were higher than their counterparts in the likelihoods of the Use context (Tablets: Glance (1.4%): $\chi^2 = 3.732$, $p = .053$, $df = 1$, Read (0.6%):

[1] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Table 1: Actual and Ideal usage of devices for proceeding to each notification-interaction stage

	Phone				Tablet				Computer				Wearable			
Device Usage	notice	glance	read	act	notice	glance	read	act	notice	glance	read	act	notice	glance	read	act
Actual	80.1% (1311)	82.0% (1278)	79.3% (737)	76.0% (595)	0.3% (4)	0.3% (4)	0.1% (1)	0.2% (1)	2.8% (42)	2.9% (42)	6.3% (53)	7.1% (52)	9.3% (103)	8.6% (91)	0.6% (4)	0.0% (0)
Ideal	96.6% (1582)	89.6% (1200)	92.1% (777)	91.7% (677)	8.3% (107)	0.7% (7)	0.6% (4)	0.9% (5)	13.5% (201)	4.5% (55)	7.1% (54)	8.0% (55)	35.5% (394)	9.0% (78)	1.5% (9)	0.2% (1)

Table 2: Ideal usage of devices for proceeding to each notification-interaction stage, by device context

	Phone				Tablet				Computer				Wearable			
Device Context	notice	glance	read	act	notice	glance	read	act	notice	glance	read	act	notice	glance	read	act
View	97.4% (1439)	91.0% (1118)	92.8% (723)	92.7% (621)	28.2% (33)	5.7% (5)	6.8% (4)	9.4% (5)	33.4% (128)	16.4% (52)	25.1% (48)	29.4% (48)	45.6% (236)	10.7% (44)	2.0% (6)	0.4% (1)
Sense	98.3% (1285)	91.6% (974)	93.1% (644)	93.0% (554)	28.7% (27)	4.6% (3)	0.0% (0)	0.0% (0)	31.5% (87)	14.6% (32)	24.1% (33)	26.1% (31)	48.6% (265)	15.9% (60)	2.0% (6)	0.4% (1)
Reach	96.8% (1438)	90.3% (1095)	92.3% (735)	91.8% (639)	19.2% (41)	1.9% (3)	0.0% (0)	0.0% (0)	31.1% (136)	12.2% (42)	21.3% (47)	24.9% (50)	44.8% (277)	12.3% (55)	1.3% (4)	0.3% (1)
Use	96.9% (1511)	90.3% (1149)	92.3% (769)	92.1% (669)	11.9% (47)	1.4% (4)	0.6% (1)	0.6% (1)	26.4% (144)	9.9% (42)	18.0% (49)	20.0% (41)	42.3% (280)	11.8% (56)	1.5% (5)	0.3% (1)
None	90.0% (9)	42.9% (3)	100.0% (5)	100.0% (5)	6.2% (54)	0.0% (0)	0.0% (0)	0.0% (0)	5.1% (46)	0.4% (3)	0.9% (4)	0.9% (4)	14.8% (56)	4.4% (15)	0.5% (1)	0.0% (0)
Overall	96.6% (1582)	89.6% (1200)	92.1% (777)	91.7% (677)	8.2% (107)	0.7% (7)	0.6% (4)	0.9% (5)	13.5% (201)	4.5% (55)	7.1% (54)	8.0% (55)	35.5% (394)	9.0% (78)	1.5% (9)	0.2% (1)

$\chi^2 = 5.025$, $p = .025$, $df = 1$, Act (0.7%): $\chi^2 = 7.979$, $p = .005$, $df = 1$; Computer: Glance (9.9%): $\chi^2 = 6.341$, $p = .012$, $df = 1$, Read (18.0%): $\chi^2 = 3.015$, $p = .083$, $df = 1$, Act (20.0%): $\chi^2 = 4.402$, $p = .036$, $df = 1$). Moreover, when tablets could be sensed as well as having their screens visible, they were chosen nearly 28% of the time as one of the devices for noticing notifications: considerably more often than they could be taken out for use (11.9%, view vs. use: $\chi^2 = 16.887$, $p < .001$, $df = 1$; sense vs. use: $\chi^2 = 15.360$, $p < .001$, $df = 1$). A different pattern was observed with wearable devices, which our participants were about equally likely to choose as one of the devices for noticing notifications irrespective of device context. To sum up, the participants' ideal usage of devices was influenced by device context, and such influence differed both across device types and across N-I stages. The gaps between actual and ideal usage will be explored in the next section.

4.2 Gaps between Ideal and Real Usage

4.2.1 Likelihood of a Top-ranked Device Actually Being Used. We closely compared the participants' actual and ideal usage of devices in each ESM response. Table 3 shows the likelihoods of each device being used to proceed to an N-I stage when that device was named the top one for that stage – or, in the case of Notice stage, was selected. Among the 268 self-reported ideal usage in which a non-phone device was named as the top ideal one for proceeding to one of the following N-I stages – Glance, Read, or Act, they were actually used by the participant for proceeding to that stage only 49.3% of the time (n = 132). Furthermore, among the 173 sampled notifications in which a non-phone device was ever named as a top ideal one, irrespective of N-I stage, only in 43.3% of these instances these devices was actually used (n = 75). These two results suggest clear gaps between actual and ideal usage for non-phone-devices. In contrast, among the 2,654 instances where phones were ranked as the top ideal device for a particular N-I stage, the participants did not use them only 14.2% of the time (n = 378), which ranged from 12%-18% of the time. In most such cases, this was because the intended stages did not happen at all, or not at that moment, and not because the participant had opted to use some other device. The only exception was participants noticing notification alerts on their wearable devices, which took place 5.4% of the time when they ranked their phone as their ideal option for this N-I stage.

For computers, the gap at the glance stage was very obvious. That is, at those times when participants ranked computers as their number-one ideal device for glancing at notifications, they only used them for that purpose in 50% of cases. Even among the times participants ranked computers as most ideal for reading and acting on the notifications, they still failed to use them for those two N-I stages in at least 18% of instances. Most of the usage that was ideally ascribed to computers at those two stages, but took place elsewhere, in fact occurred on phones. A common reason provided by participants for using their phones despite ideally wanting to use their computers was that the latter *"was not with me"* (P26). However, they also mentioned other reasons, such as that the computer's app for receiving that notification was inactive or *"not launched"* (P2). However, up to 11% of the time, these participants did not end up using any device to proceed to the Read and Act stages.

Wearable devices were also more prone than computers to gaps between actual and ideal usage. When they were ranked as most ideal for glancing at and reading notifications, our participants actually used them for those purposes 11% and 22% of the time, respectively. At times when participants ideally wanted to use wearables to glance at and read notifications, they did not use any device to proceed to Glance in 26% of cases, or to Read in 33% of cases. For these instances, participants reported that they

Table 3: Distribution of actual device usage vs. devices ranked as ideal, by notification-interaction stage

Top-ranked Device \ Actual Usage	N-I Stage	Phone	Tablet	Computer	Wearable	No Device Used
Phone	notice	82.0% (1297)	0.3% (4)	1.5% (24)	5.4% (85)	10.9% (172)
	glance	87.8% (1053)		0.3% (4)	2.6% (31)	9.3% (112)
	read	85.7% (666)		0.4% (3)	0.1% (1)	13.8% (107)
	act	82.3% (557)		0.2% (1)		17.6% (119)
Tablet	notice	79.4% (85)	3.7% (4)		1.9% (2)	15.0% (16)
	glance	100.0% (7)	0.0% (0)			
	read	75.0% (3)	25.0% (1)			
	act	80.0% (4)	20.0% (1)			
Computer	notice	61.2% (123)	0.5% (1)	19.9% (40)	6.5% (13)	11.9% (24)
	glance	45.5% (25)		49.1% (27)	1.8% (1)	3.6% (2)
	read	14.8% (8)		74.1% (40)		11.1% (6)
	act	12.7% (7)		81.8% (45)		5.5% (3)
Wearable	notice	63.2% (249)		2.3% (9)	24.4% (96)	10.2% (40)
	glance	52.6% (41)			21.8% (17)	25.6% (20)
	read	55.6% (5)			11.1% (1)	33.3% (3)
	act				0.0% (0)	100.0% (1)

Note. An empty cell indicates that no data points existed in our dataset for that combination

only wanted to use their wearable devices to attend to certain notifications but not other devices, some examples provided by the participants including: "sleep report" (P8), "weather" (P16), as P8 said, *"It's too troublesome to use the phone to see my sleep report. It's faster to use smartwatch."* Thus, when they did not use wearable devices for these N-I stages, they did not use their other devices to do so either.

Finally, in the relatively small number of instances where tablets were ranked as the number-one ideal device for interacting with notifications in the second, third and fourth N-I stages, the participants still rarely used them for this purpose; and in all such instances, they ended up using their phones instead. The context information participants provided in these ESM instances indicated that they were sometimes undertaking entertainment activities.

4.2.2 Differences in Device Context Between Realized and Unrealized Ideal Usage. Next, we investigated how device contexts differed between scenarios in which the ideal device was actually used vs. not. Figure 3 shows these differences. In it, a positive value indicates that a particular device context was more often associated with an ideal device being actually used than with it not being used; and a negative value means the opposite. A large positive value therefore indicates a relatively close association between the device context and the device being actually used as desired. Note that tablet is not presented in this section because the number of times tablets were selected as an ideal device was too small to further separate such occurrences by device context. As Figure 3 shows, the associations between device context and whether an ideal device was actually used varied both among devices and among device contexts. For instance, participants' perceived device context seemed to play a role, albeit a minor one, in whether they would use their phones as desired for glancing at notifications.

When computers were deemed the ideal device for reading and acting on notifications, they were much more often in states of being viewable, reachable and usable when they were actually used vs. when they were not actually used, but not in a state of sensible (Read: viewable: $\chi^2 = 6.600$, $p = .010$, $df = 1$; sensible: $\chi^2 = 1.440$, $p = .230$, $df = 1$; reachable: $\chi^2 = 11.431$, $p < .001$, $df = 1$, usable: $\chi^2 = 10.240$, $p = .001$, $df = 1$; Act: viewable: $\chi^2 = 4.632$, $p = .031$, $df = 1$; sensible: $\chi^2 = 1.601$, $p = .206$, $df = 1$; reachable: $\chi^2 = 8.945$, $p = .003$, $df = 1$, usable: $\chi^2 = 13.342$, $p < .001$, $df = 1$;). On the other hand, when computers were deemed the ideal device for glancing at notifications, they were also much more often in states of being reachable and usable when they were actually used vs. when they were not actually used, but not in the former two substantive device context (viewable: $\chi^2 = 0.560$, $p = .454$, $df = 1$; sensible: $\chi^2 = 0.907$, $p = .341$, $df = 1$; reachable: $\chi^2 = 13.581$, $p < .001$, $df = 1$, usable: $\chi^2 = 13.581$, $p < .001$, $df = 1$). This indicates that computers' viewability/audibility was less vital to whether participants could use them as desired for glancing at notifications than it was in the cases of reading and acting. Together, these results suggest that whether computers were reachable and usable at the moment was vital to whether they (when deemed the ideal device) would actually be used as desired to interact with notifications, irrespective of N-I stage. On the other hand, whether they were in a screen-viewable state was also vital to whether people would eventually use them to read and act upon notifications.

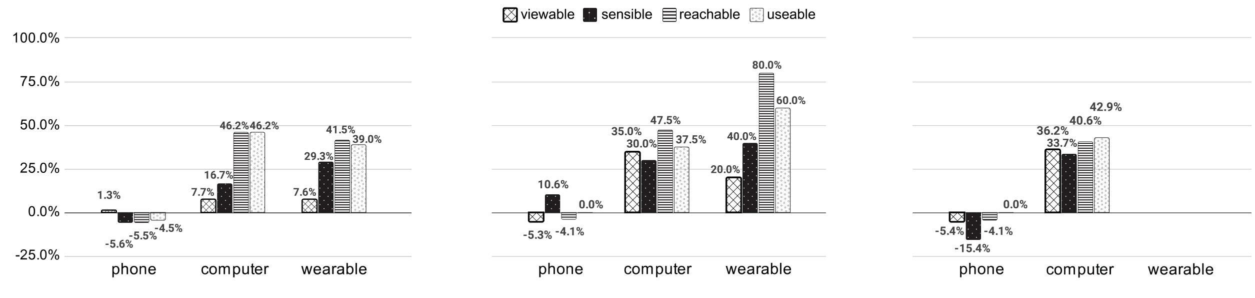


Figure 3: Comparison of the occurrence of top-ranked devices being not used vs. used, by device context. (a) glance stage, (b) read stage and (c) act stage.

Finally, when wearable devices were deemed ideal for glancing at notifications, they were much more often in states of being reachable and usable when they were actually used than they were not actually used (reachable: $\chi^2 = 8.071$, $p = .004$, $df = 1$; usable: $\chi^2 = 7.312$, $p = .007$, $df = 1$). This phenomenon appeared to also apply to these devices being deemed ideal for reading notifications; yet, these cases were rare in our dataset (N=6), and thus the difference between device contexts for this stage was not statistically significant. This result suggests that whether these devices could be sensed also mattered to whether they would actually be used for glancing at notifications.

All of these results above resonates with the results presented in the previous sections, suggesting that there were associations between device context and whether participants wanted to use, and would/could actually use their ideal devices for proceeding to particular N-I stage.

4.3 Switching between vs. Continuing to Use Devices

Finally, we found that the desire to continue N-I on the same device played a significant role in participants' choice of device. That is, when participants had used or wanted to use a particular device to proceed to a particular N-I stage, the likelihood of them wanting to continue using the same device for its following N-I stage was high. Figures 4(a) and 4(b) respectively show the likelihood of participants' desire to use the same device for proceeding to the next N-I stage when they 1) had used and 2) wanted to use that device for the previous stage. Here, we included all cases where participants listed the device as the top ideal one, regardless of whether they had listed any subsidiary ideal ones or not. This was because, in all these cases, the participants wanted to continue using the same device.

In Figure 4(a), very high likelihoods are shown for both phones and computers continuing to be used after being utilized for the previous N-I stage. Although in the case of tablets the number of samples was low (n = 5), it is noteworthy that in all of these instances, they wanted to use their tablet to proceed to the next N-I stage instead of switching to a different device. This result shows that tablets were the devices least likely to be chosen as ideal in every N-I stage; and suggests the very high impact of people's desire for continuity in the N-I process.

With regard to the ideal flow between N-I stages, this desire for continuing the N-I process on the same device became more obvious in its later stages. Participants wanted to continue using their tablets and computers 27.9% and 55.4% of the time, respectively, when they wanted to glance at and notice them on the same device. However, when they wanted to use them to glance at notifications and read them on these same two device types, the parallel figures were 81.8% and 92.2%; and the likelihoods then increased to 93.3% and 95.6% in the transition from Read to Act.

Wearable devices exhibited a distinctive pattern, possibly due to their limited capabilities. The likelihood of such a device being chosen as ideal after participants had glanced at it declined to less than 28%; and this decline continued as the N-I process unfolded. This result corresponds to many participants reported that their wearable device did not allow them to act on their notifications, or made doing so inconvenient. The decline was also observed in their ideal flow between devices: from 52.9%, 31.5%, down to 20.8%.

5 DISCUSSION

5.1 Device Context is Key to Understanding Gaps between Actual and Ideal Usage

Our findings align with existing research on the use of multiple devices, in that they confirm the importance of personal preferences for devices and the important role of users' perceptions of device characteristics in shaping those preferences. For example, participants in our study cited factors such as screen size, proximity, ease of input, availability of software, and connectivity as reasons for their device preferences[2, 14, 15, 41, 63, 71, 72, 76]. However, our results extend the existing literature by showing that these preferences largely reflected users' perceptions of the suitability of a particular device for a given N-I stage, based on their perceptions of the characteristics of their devices. For example, although participants mostly preferred to use their phones and rarely used their tablets across all N-I stages, they used computers mainly for reading full notification content and acting upon it; and used wearable devices mainly in the Notice and Glance stages.

Additionally, while prior studies have mainly focused on high-level contexts such as activity [9, 41], location [29, 31, 76], social

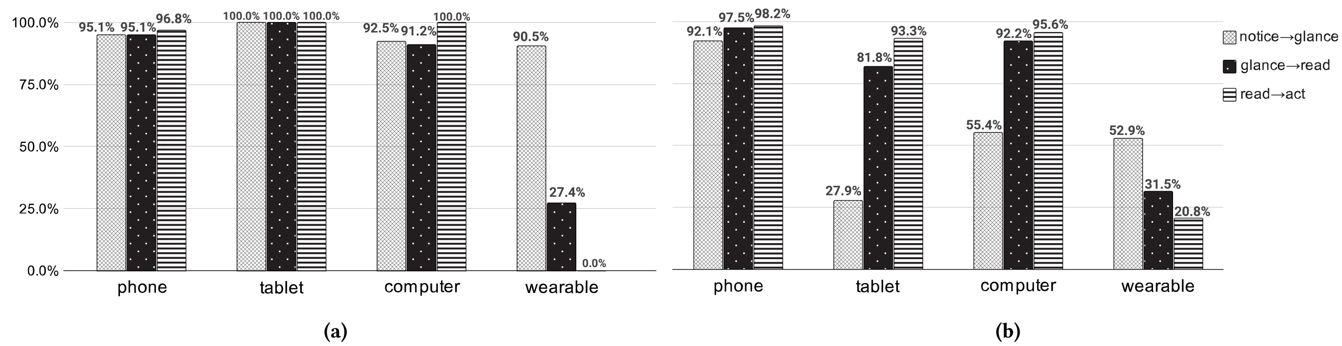


Figure 4: The likelihood of a device being chosen as the number-one ideal device for proceeding to the next N-I stage (a) after the participant had used that device for its preceding stage, and (b) when participants had named that device as ideal for its preceding stage

context [9, 29, 31, 48], and privacy concerns [9, 29, 48], with some research also examining the role of proximity in device choice [76], our findings suggest that, in addition to proximity, device context plays a significant role. Specifically, we found that the gaps between participants' actual and ideal usage of devices were often linked to their perceptions of certain device contexts as ideal for certain N-I stages. These gaps were particularly noticeable for non-phone devices, which our results showed were used for their intended N-I stages in less than half of the cases. Other than lack of proximity, out participants' frequently cited reasons for not using such devices included them being in a bag, their screens not being visible, and their not being turned on. The participants also reported that they often had to use an alternative device when they did not (or perhaps could not) use their top-ranked device for their intended purpose – in some cases, because they did not want to expend the extra effort it would have taken to make the device suitable for their intended context (e.g., making it "visibly available"). In other cases, however, participants reported that they did not use an alternative device simply because they preferred to use their ideal device.

Our results also indicated that, from the user's perspective, certain device contexts might be more or less important than others when it comes to deciding whether a given device should be used for a specific N-I stage. For example, phones' actual and ideal usage were similar regardless of our four main device contexts, whereas computers were less likely to be considered the most ideal device for reading and acting on notifications when participants reported that they could not see their screens; in contrast, wearable devices were more frequently named as the ideal means of noticing notification alerts when participants could sense them, compared to when they could see their screens. Our findings that a device being in a certain state affected whether it was considered ideal for use in specific N-I stages suggest that a future multi-device notification ecosystem should take into account not only the use of multiple devices, but also the impact of specific device contexts on different stages of the N-I process. In Section 5.3, we provide more discussion of this study's implications for the design of future notification ecosystems.

5.2 Switching to a Different Device vs. Continuing to Use the Same One

Lastly, our results show that, in addition to device context and personal preferences, participants' ideal choices of devices for proceeding to particular N-I stages was heavily influenced by which device they had used to undertake (or had wanted to undertake) the previous N-I stage. This result is somewhat conceptually similar to Mehrotra et al.'s [41] result that the devices users used for previous notifications would influence their choice of devices for the next, in the sense that it describes users' tendency of using the same device. But it differs in the sense that it concerns the continuity of different stages of interactions with the same notification. That is, while many previous studies on multi-device usage suggest that users switch devices to improve their performance and workflow [17, 30, 60, 80], we found that in the context of N-I, people displayed a salient tendency to proceed from their present N-I stage to the next on the same device. In the case of non-phone devices, and especially rarely used ones such as tablets, this likelihood was much higher than the participants' overall actual usage.

We also noted that participants occasionally considered the possibility of device-switching for subsequent N-I stages when reporting their ideal flow among devices, possibly due to consideration of the devices' suitability for each stage; but in reality, they mostly just continued to the next N-I stage on the same device. On the surface, such high actual device continuity across N-I stages might be attributed to participants being in situations where that particular device was ideal for all N-I stages from their perspective. However, it seems just as likely that participants wanted to reduce the attentional and/or time cost of switching devices for interacting with the same notifications, following the principle of least effort [81].However, we are unable to confidently conclude from our dataset whether or not users' desire to proceed smoothly through the N-I stages was more influential on their ideal and actual choices of devices for those stages than their preference of devices. It is also likely that the perceived overhead involved in switching and managing configurations and information across multiple devices [24, 58] play a vital role. Future research should therefore include deeper investigation into the tensions among them.

Finally, it is noteworthy that participants' self-reported ideal flow among devices for proceeding to the four N-I stages to a large extent resonates with previous studies on the distinction between attentiveness and responsiveness [12, 35, 78]. That is, on average, our participants reported ideally wanting to use the same device for Read and Act, but not minding whether the same device was used for Glance and Notice – indicating that the first two N-I stages were the most dissimilar and the last two N-I stages were the most similar, in terms of the device suitable for performing them.

5.3 Design Implications

Most of the time, our participants considered their phones to be the ideal devices for all N-I stages, indicating that phones are often suitable and sufficient for completing all stages. However, the fact that many of them ranked non-phone devices as their ideal choices for certain N-I stages suggests that a multi-device notification ecosystem (MDNE) could be beneficial. Our identification of gaps also highlights the potential benefits of an MDNE, insofar as it could bridge them to better support people's ideal device usage and reduce their reliance on any one particular device. This, in turn, could improve users' workflow across devices [60, 80]. Our design proposals for an MDNE focus on enabling users to effectively perform the N-I stages across multiple devices.

One key recommendation is that users should be provided with the flexibility to select a specific device to use for a given N-I stage for a given set of notifications. Currently, this is not viable because notifications are created and stored locally on different devices and are not synchronized across devices such as phones, tablets, and computers. Based on our finding that users sometimes considered non-phone devices to be more suitable for certain N-I stages (e.g., preferring to use computers for the last two stages), we recommend that future MDNEs enable notifications to be synchronized across devices, allowing users to have this flexibility. Such synchronization could also potentially resolve the issue of duplicate notifications across devices [15], by removing notifications dismissed on one device from all the users' other synchronized devices. Currently, certain communication-related applications offer their own synchronization features, but notification synchronization allows users to have the same flexibility for all types of notifications. This means that users can manage and access their notifications in the same way across different applications and devices. For this synchronization to effectively support users' performance of the N-I stages, an MDNE provider should encourage application developers to make their apps available across devices. This is because, as mentioned by some participants, software availability was a factor that hindered them from using particular devices.

It is noteworthy that, however, for cross-device N-I stage synchronization to be successful, it is essential for MDNEs to avoid automatically dismissing notifications after they are read, as it would result in already read notifications being removed on all other devices, making interaction with notifications across devices impossible. Therefore, we suggest that MDNEs use a different method to indicate notification status, such as using colors to differentiate between read and unread notifications, similar to an email inbox. Additionally, MDNEs should enable users to mark notifications as "to-do" or "pinned" items to be acted upon at a later time across multiple devices. On the other hand, once notifications are not automatically dismissed, MDNEs should offer efficient dismissal options, such as batch dismissing all read or unmarked notifications. Furthermore, if users want to hide already read notifications, the MDNE can provide options such as moving them to a less visible location.

Our second high-level recommendation is that MDNEs should learn from users' device choices, device contexts, and N-I process flow across multiple devices over time. If effective learning of such patterns is to occur, it would be beneficial to grant users the aforementioned flexibility to select a specific device to use for a given N-I stage, as otherwise their choice of device might not be their ideal ones. Before such a system has learned a user's individual pattern, it can use our results as a basis for making default delivery choices, such as not displaying notifications on a computer when its screen is not visible to the user, or not generating alerts on a wearable when it is not being worn by the user. Once the MDNE has learned such patterns, on the other hand, we recommend that it deliver and present notifications based on them. If the user's ideal device for a given notification is not in the preferred context for interacting with it, the MDNE should defer that notification until the ideal device is in that context. For example, it could defer notifications that users prefer to receive on wearables until the wearables are being worn, or defer notifications that require large screens for content display until the computer screen is on and visible to the user. Such deferral aligns well with users' tendency to continue interacting with a given notification on the same device they began interacting with it on, as it would allow them to perform all N-I stages in rapid succession on the device they prefer. However, as some notifications may be time-sensitive, users should be given the option to activate or deactivate this deferral feature and customize its settings. In any case, additional research on the deferral of notifications will be needed if we are to understand users' perceptions of the tradeoff between device suitability and the need to deal with notifications in a timely way.

5.4 Study Limitations

The current study is subject to several limitations. First, despite having a similar sample size to that of many other ESM studies (e.g. [7, 11, 32, 35, 37, 44]), it did not have enough participants for us to analyze the interwoven relationships of device contexts, device preferences, the participants' tendency not to engage in device-switching across N-I stages, notification types, sensor data, and other contextual factors such as activity, location, and perceived social environment. In addition, we also did not gather information on various other factors that might have affected their responses in the moment, such as alert modalities, notification settings, device specifications (e.g. screen size, keyboard style), delays in notification presentation, and whether they were actively using a device at a given moment. However, because many of them were not necessarily observable by us or the participants, and because considering all of them would make our ESM and analysis even more complex, we maintained our existing focus on device types, device contexts, and N-I stages. Given this absence, the present paper cannot clarify the interrelationships of those factors, or indeed their respective bilateral relationships with the factors we did study, as that would have required a much larger sample size for each such condition to ensure that we had enough data points for pairwise comparisons among them. That being said, the factors we did focus on have hitherto received relatively little attention, despite – in our view – being essential for the development of an MDNE. We therefore encourage future researchers to extend this line of inquiry to include the impacts and interrelationships of all the factors mentioned above, perhaps among others.

Second, we chose to let participants select multiple devices for generating notification alerts instead of – as in the other N-I stages – only their top devices, on the grounds that the latter stages required the user to initiate attention-switching to a single device, whereas alerts are initiated by a system. This choice, though reasonable in itself, resulted in a fundamental difference between, on the one hand, the percentages of each device being chosen for the Notice stage, and on the other, the equivalent (but inevitably lower) percentages for the other three N-I stages. Thus, direct comparison of statistics between Notice and the other stages should be conducted with caution; and we encourage future researchers to let their participants also rank their top devices in the Notice stage, to facilitate direct comparison between it and the other stages.

Third, given the relatively low prevalence and adoption of other types of devices for receiving notifications in the market, we could only focus our inquiry on four device types. However, not all of our participants owned all four focal kinds of device, though all had smartphones. This could have made device prevalence another factor that influenced their actual usage, and even possibly their preferences, as a result of implicit association [23]. Although our findings regarding device preferences for performing specific interactions with notifications resonate with those of previous studies [41, 76], it is possible that the uneven distribution of device types among our participants could have impacted the overall patterns of preference that we observed. Moreover, our study was conducted in the early post-COVID pandemic period (i.e., in May 2022) and the participants were mostly young Android users in Taiwan. Therefore, the participants' cross-device experiences and preferences could differ from those of iOS users, older people, and people living in other parts of the world in different pandemic situations.

Fourth, our dataset could have been biased towards moments when participants were more likely to respond to ESM questionnaires, and therefore more receptive to notifications. Additionally, the fact that participants were only able to complete questionnaires on their phones, rather than on other devices, may have biased our actual-usage data towards the usage of phones.

6 CONCLUSION

In this paper, we have presented the results of an ESM study that investigated smartphone users' multi-device interaction with notifications. Unlike other multi-device research, it revealed gaps between these users' actual and ideal usage of multiple devices for N-I across its multiple stages, including noticing the alert, glancing at the notification, reading the full content, and acting upon it. In addition to revealing these gaps, we delineated the influence of various device contexts on device choices, and how the aforementioned gaps might be related to such contexts, which were further shown to be device-type specific. We also identified a strong tendency for people to continue using the same device that they had used to complete a given N-I stage to proceed to the next such stage. Lastly, along with in-depth discussion of these findings, we highlighted design implications for a future multi-device notification ecosystem.

ACKNOWLEDGMENTS

We greatly thank all of our study participants. This project is supported by National Science Technology Council, Taiwan (110-2222-E-A49-008-MY3), as well as the Higher Education Sprout Project of National Yang Ming Chiao Tung University and Ministry of Education (MOE), Taiwan.

REFERENCES

[1] Piotr D. Adamczyk and Brian P. Bailey. 2004. If not now, when? the effects of interruption at different moments within task execution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. Association for Computing Machinery, New York, NY, USA, 271–278. https://doi.org/10.1145/985692.985727

[2] Daniel L. Ashbrook, James R. Clawson, Kent Lyons, Thad E. Starner, and Nirmal Patel. 2008. Quickdraw: The Impact of Mobility and on-Body Placement on Device Access Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 219–222. https://doi.org/10.1145/1357054.1357092

[3] Jonas Auda, Dominik Weber, Alexandra Voit, and Stefan Schneegass. 2018. Understanding User Preferences towards Rule-based Notification Deferral. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188688

[4] Jiang Bian, Anlei Dong, Xiaofeng He, Srihari Reddy, and Yi Chang. 2013. User Action Interpretation for Online Content Optimization. *IEEE Transactions on Knowledge and Data Engineering* 25, 9 (Sept. 2013), 2161–2174. https://doi.org/10.1109/TKDE.2012.130 Conference Name: IEEE Transactions on Knowledge and Data Engineering.

[5] Leslie H Brown, Timothy Strauman, Neus Barrantes-Vidal, Paul J Silvia, and Thomas R Kwapil. 2011. An experience-sampling study of depressive symptoms and their social context. *The Journal of nervous and mental disease* 199, 6 (2011), 403–409.

[6] Michael Bub, Evgeny Volynsky, and Wolfgang Wörndl. 2018. User Acceptance of Proactive Recommendations on Smartphone and Smartwatch. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization (UMAP '18)*. Association for Computing Machinery, New York, NY, USA, 143–150. https://doi.org/10.1145/3213586.3226210

[7] Seyma Kucukozer Cavdar, Tugba Taskaya-Temizel, Mirco Musolesi, and Peter Tino. 2020. A Multi-perspective Analysis of Social Context and Personal Factors in Office Settings for the Design of an Effective Mobile Notification System. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 15:1–15:38. https://doi.org/10.1145/3381000

[8] M. Cecchinato, A. Sellen, M. Shokouhi, and G. Smyth. 2016. Finding email in a multi-account, multi-device world. http://dx.doi.org/10.1145/2858036.2858473 Conference Name: CHI 2016 Conference on Human Factors in Computing Systems Meeting Name: CHI 2016 Conference on Human Factors in Computing Systems Pages: 1200-1210 Place: New York, NY, USA Publisher: ACM publishing.

[9] Marta E. Cecchinato, Anna L. Cox, and Jon Bird. 2017. Always On(line)? User Experience of Smartwatches and their Role within Multi-Device Ecologies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3557–3568. https://doi.org/10.1145/3025453.3025538

[10] Chung Chiao Chang, Meng-Hsin Wu, Yu-Jen Lee, XiJing Chang, and Yung-Ju Chang. 2021. Opportune Moments for the Multi-Stage Notification Responding Process: A Preliminary Investigation. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*. 9–10.

[11] Yung-Ju Chang, Yi-Ju Chung, and Yi-Hao Shih. 2019. I Think It's Her: Investigating Smartphone Users' Speculation about Phone Notifications and Its Influence on Attendance. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3338286.3340125

[12] Yung-Ju Chang and John C. Tang. 2015. Investigating Mobile Users' Ringer Mode Usage and Attentiveness and Responsiveness to Communication. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile*

Devices and Services (MobileHCI '15). Association for Computing Machinery, New York, NY, USA, 6–15. https://doi.org/10.1145/2785830.2785852

[13] Sunny Consolvo and Miriam Walker. 2003. Using the experience sampling method to evaluate ubicomp applications. *IEEE pervasive computing* 2, 2 (2003), 24–31.

[14] Fulvio Corno, Luigi De Russis, and Teodoro Montanaro. 2015. A Context and User Aware Smart Notification System. https://doi.org/10.1109/WF-IoT.2015.7389130

[15] Fulvio Corno, Luigi De Russis, and Alberto Roffarello. 2018. AwareNotifications: Multi-Device Semantic Notification Handling with User-Defined Preferences. *Journal of Ambient Intelligence and Smart Environments* 10 (Aug. 2018). https://doi.org/10.3233/AIS-180492

[16] Mihaly Csikszentmihalyi and Reed Larson. 2014. Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology*. Springer, 35–54.

[17] David Dearman and Jeffery S. Pierce. 2008. It's on my other computer! computing with multiple devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. Association for Computing Machinery, New York, NY, USA, 767–776. https://doi.org/10.1145/1357054.1357177

[18] Tilman Dingler and Martin Pielot. 2015. I'll be there for you: Quantifying Attentiveness towards Mobile Messaging. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/2785830.2785840

[19] Jon D. Elhai, Dmitri Rozgonjuk, Ahmad M. Alghraibeh, and Haibo Yang. 2021. Disrupted Daily Activities From Interruptive Smartphone Notifications: Relations With Depression and Anxiety Severity and the Mediating Role of Boredom Proneness. *Social Science Computer Review* 39, 1 (Feb. 2021), 20–37. https://doi.org/10.1177/0894439319858008 Publisher: SAGE Publications Inc.

[20] Anja Exler, Marcel Braith, Kristina Mincheva, Andrea Schankin, and Michael Beigl. 2018. Smartphone-Based Estimation of a User Being in Company or Alone Based on Place, Time, and Activity. In *Mobile Computing, Applications, and Services (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering)*, Kazuya Murao, Ren Ohmura, Sozo Inoue, and Yusuke Gotoh (Eds.). Springer International Publishing, Cham, 74–89. https://doi.org/10.1007/978-3-319-90740-6_5

[21] Joel E. Fischer, Chris Greenhalgh, and Steve Benford. 2011. Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 181–190. https://doi.org/10.1145/2037373.2037402

[22] Joel E. Fischer, Nick Yee, Victoria Bellotti, Nathan Good, Steve Benford, and Chris Greenhalgh. 2010. Effects of content and time of delivery on receptivity to mobile interruptions. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (MobileHCI '10)*. Association for Computing Machinery, New York, NY, USA, 103–112. https://doi.org/10.1145/1851600.1851620

[23] Anthony G. Greenwald and Mahzarin R. Banaji. 1995. Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review* 102 (1995), 4–27. https://doi.org/10.1037/0033-295X.102.1.4 Place: US Publisher: American Psychological Association.

[24] Jens Grubert, Matthias Kranz, and Aaron Quigley. 2016. Challenges in Mobile Multi-Device Ecosystems. *mUX: The Journal of Mobile User Experience* 5, 1 (Dec. 2016), 5. https://doi.org/10.1186/s13678-016-0007-y arXiv:1605.07760 [cs].

[25] Marcelo Guzman, Gustavo López, and Luis A. Guerrero. 2018. Human Factors Affecting the Development of Smart Device-Based Notifications. In *Advances in Human Factors and Systems Interaction (Advances in Intelligent Systems and Computing)*, Isabel L. Nunes (Ed.). Springer International Publishing, Cham, 251–259. https://doi.org/10.1007/978-3-319-60366-7_24

[26] Daniel Harrison and Marta E. Cecchinato. 2015. "Give me five minutes!": feeling time slip by. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 45–48. https://doi.org/10.1145/2800835.2800858

[27] Joyce Ho and Stephen S. Intille. 2005. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. Association for Computing Machinery, New York, NY, USA, 909–918. https://doi.org/10.1145/1054972.1055100

[28] Shamsi T. Iqbal and Brian P. Bailey. 2007. Understanding and developing models for detecting and differentiating breakpoints during interactive tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 697–706. https://doi.org/10.1145/1240624.1240732

[29] Tero Jokela, Jarno Ojala, and Thomas Olsson. 2015. A Diary Study on Combining Multiple Information Devices in Everyday Activities and Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3903–3912. https://doi.org/10.1145/2702123.2702211

[30] Amy K. Karlson, Shamsi T. Iqbal, Brian Meyers, Gonzalo Ramos, Kathy Lee, and John C. Tang. 2010. Mobile taskflow in context: a screenshot study of smartphone usage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 2009–2018. https://doi.org/10.1145/1753326.1753631

[31] Fahim Kawsar and A.J. Bernheim Brush. 2013. Home computing unplugged: why, where and when people use different connected devices at home. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing (UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 627–636. https://doi.org/10.1145/2493432.2493494

[32] Inyeop Kim, Hwarang Goh, Youngtae Noh, and Uichin Lee. 2021. Poster: Exploring User Contexts and Needs for Context-aware Smartphone Distraction Management. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp '21)*. Association for Computing Machinery, New York, NY, USA, 41–43. https://doi.org/10.1145/3460418.3479307

[33] Kostadin Kushlev, Jason Proulx, and Elizabeth W. Dunn. 2016. "Silence Your Phones": Smartphone Notifications Increase Inattention and Hyperactivity Symptoms. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1011–1020. https://doi.org/10.1145/2858036.2858359

[34] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82, 13 (2017), 1–26. https://doi.org/10.18637/jss.v082.i13

[35] Hao-Ping Lee, Kuan-Yin Chen, Chih-Heng Lin, Chia-Yu Chen, Yu-Lin Chung, Yung-Ju Chang, and Chien-Ru Sun. 2019. Does Who Matter? Studying the Impact of Relationship Characteristics on Receptivity to Mobile IM Messages. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300756

[36] Luis Leiva, Matthias Böhmer, Sven Gehring, and Antonio Krüger. 2012. Back to the app: the costs of mobile application interruptions. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI '12)*. Association for Computing Machinery, New York, NY, USA, 291–294. https://doi.org/10.1145/2371574.2371617

[37] Tianshi Li, Julia Katherine Haines, Miguel Flores Ruiz de Eguino, Jason I. Hong, and Jeffrey Nichols. 2021. Alert Now or Never: Understanding and Predicting Notification Preferences of Smartphone Users. *ACM Transactions on Computer-Human Interaction* (2021). https://doi.org/10.1145/3478868 Just Accepted.

[38] Tzu-Chieh Lin, Yu-Shao Su, Emily Helen Yang, Yun Han Chen, Hao-Ping Lee, and Yung-Ju Chang. 2021. "Put it on the Top, I'll Read it Later": Investigating Users' Desired Display Order for Smartphone Notifications. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445384

[39] Andrés Lucero. 2015. Using Affinity Diagrams to Evaluate Interactive Prototypes. In *Human-Computer Interaction – INTERACT 2015 (Lecture Notes in Computer Science)*, Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler (Eds.). Springer International Publishing, Cham, 231–248. https://doi.org/10.1007/978-3-319-22668-2_19

[40] Afra Mashhadi, Akhil Mathur, and Fahim Kawsar. 2014. The myth of subtle notifications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 111–114. https://doi.org/10.1145/2638728.2638759

[41] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. 2019. NotifyMeHere: Intelligent Notification Delivery in Multi-Device Environments. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (CHIIR '19)*. Association for Computing Machinery, New York, NY, USA, 103–111. https://doi.org/10.1145/3295750.3298932

[42] Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. 2015. Designing content-driven intelligent notification mechanisms for mobile applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 813–824. https://doi.org/10.1145/2750858.2807544

[43] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1021–1032. https://doi.org/10.1145/2858036.2858566

[44] Abhinav Mehrotra, Fani Tsapeli, Robert Hendley, and Mirco Musolesi. 2017. MyTraces: Investigating Correlation and Causation between Users' Emotional States and Mobile Phone Interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 83:1–83:21. https://doi.org/10.1145/3130948

[45] Alberto Monge Roffarello and Luigi De Russis. 2021. Coping with Digital Wellbeing in a Multi-Device World. In *Proceedings of the 2021 CHI Conference on Human*

Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445076

[46] Ngoc Thi Nguyen, Agustin Zuniga, Huber Flores, Hyowon Lee, Simon Tangi Perrault, and Petteri Nurmi. 2021. Intelligent Shifting Cues: Increasing the Awareness of Multi-Device Interaction Opportunities. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21)*. Association for Computing Machinery, New York, NY, USA, 213–223. https://doi.org/10.1145/3450613.3456839

[47] Tadashi Okoshi, Julian Ramos, Hiroki Nozaki, Jin Nakazawa, Anind K. Dey, and Hideyuki Tokuda. 2015. Reducing users' perceived mental effort due to interruptive notifications in multi-device mobile environments. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 475–486. https://doi.org/10.1145/2750858.2807517

[48] Antti Oulasvirta and Lauri Sumari. 2007. Mobile kits and laptop trays: managing multiple devices in mobile information work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1127–1136. https://doi.org/10.1145/1240624.1240795

[49] Chunjong Park, Junsung Lim, Juho Kim, Sung-Ju Lee, and Dongman Lee. 2017. Don't Bother Me. I'm Socializing! A Breakpoint-Based Smartphone Notification System. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 541–554. https://doi.org/10.1145/2998181.2998189

[50] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: designing intelligent prompting mechanisms for pervasive applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 897–908. https://doi.org/10.1145/2632048.2632062

[51] Veljko Pejovic, Mirco Musolesi, and Abhinav Mehrotra. 2015. Investigating The Role of Task Engagement in Mobile Interruptibility. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 1100–1105. https://doi.org/10.1145/2786567.2794336

[52] Martin Pielot, Karen Church, and Rodrigo de Oliveira. 2014. An in-situ study of mobile phone notifications. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 233–242. https://doi.org/10.1145/2628363.2628364

[53] 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 (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3319–3328. https://doi.org/10.1145/2556288.2556973

[54] 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 (UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 825–836. https://doi.org/10.1145/2750858.2804252

[55] Martin Pielot, Amalia Vradi, and Souneil Park. 2018. Dismissed! a detailed exploration of how mobile phone users handle push notifications. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3229434.3229445

[56] Henning Pohl, Justyna Medrek, and Michael Rohs. 2016. ScatterWatch: Subtle Notifications via Indirect Illumination Scattered in the Skin. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 7–16. https://doi.org/10.1145/2935334.2935351

[57] Aditya Ponnada, Jixin Li, Shirlene Wang, Wei-Lin Wang, Bridgette Do, Genevieve F. Dunton, and Stephen S. Intille. 2022. Contextual Biases in Microinteraction Ecological Momentary Assessment (MEMA) Non-Response. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 1, Article 26 (mar 2022), 24 pages. https://doi.org/10.1145/3517259

[58] Umar Rashid, Miguel A. Nacenta, and Aaron Quigley. 2012. The cost of display switching: a comparison of mobile, large display and hybrid UI configurations. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '12)*. Association for Computing Machinery, New York, NY, USA, 99–106. https://doi.org/10.1145/2254556.2254577

[59] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-scale assessment of mobile notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3055–3064. https://doi.org/10.1145/2556288.2557189

[60] Stephanie Santosa and Daniel Wigdor. 2013. A field study of multi-device workflows in distributed workspaces. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing (UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 63–72. https://doi.org/10.1145/2493432.2493476

[61] Florian Schulze and Georg Groh. 2014. Studying how character of conversation affects personal receptivity to mobile notifications. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 1729–1734. https://doi.org/10.1145/2559206.2581320

[62] Florian Schulze and Georg Groh. 2016. Conversational context helps improve mobile notification management. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 518–528. https://doi.org/10.1145/2935334.2935347

[63] Alireza Sahami Shirazi and Niels Henze. 2015. Assessment of Notifications on Smartwatches. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 1111–1116. https://doi.org/10.1145/2786567.2794338

[64] Eran Toch, Hadas Chassidim, and Tali Hatuka. 2020. Can you Turn it Off? The Spatial and Social Context of Mobile Disturbance. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 91:1–91:18. https://doi.org/10.1145/3415162

[65] J. Gregory Trafton, Erik M Altmann, Derek P Brock, and Farilee E Mintz. 2003. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies* 58, 5 (May 2003), 583–603. https://doi.org/10.1016/S1071-5819(03)00023-5

[66] Liam D. Turner. 2017. *Decomposing responses to mobile notifications.* phd. Cardiff University. https://orca.cardiff.ac.uk/id/eprint/102135/

[67] Liam D. Turner, Stuart M. Allen, and Roger M. Whitaker. 2015. Push or Delay? Decomposing Smartphone Notification Response Behaviour. In *Human Behavior Understanding (Lecture Notes in Computer Science)*, Albert Ali Salah, Ben J.A. Kröse, and Diane J. Cook (Eds.). Springer International Publishing, Cham, 69–83. https://doi.org/10.1007/978-3-319-24195-1_6

[68] Liam D. Turner, Stuart M. Allen, and Roger M. Whitaker. 2017. Reachable but not receptive: Enhancing smartphone interruptibility prediction by modelling the extent of user engagement with notifications. *Pervasive and Mobile Computing* 40 (Sept. 2017), 480–494. https://doi.org/10.1016/j.pmcj.2017.01.011

[69] Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *ACM Comput. Surv.* 50, 6, Article 93 (dec 2017), 40 pages. https://doi.org/10.1145/3123988

[70] Niels van Berkel, Jorge Goncalves, Simo Hosio, Zhanna Sarsenbayeva, Eduardo Velloso, and Vassilis Kostakos. 2020. Overcoming compliance bias in self-report studies: A cross-study analysis. *International Journal of Human-Computer Studies* 134 (2020), 1–12. https://doi.org/10.1016/j.ijhcs.2019.10.003

[71] Alexandra Voit, Dominik Weber, and Niels Henze. 2018. Qualitative Investigation of Multi-Device Notifications. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 1263–1270. https://doi.org/10.1145/3267305.3274117

[72] Dominik Weber. 2017. Towards smart notification management in multi-device environments. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3098279.3119921

[73] Dominik Weber, Alexandra Voit, Jonas Auda, Stefan Schneegass, and Niels Henze. 2018. Snooze! investigating the user-defined deferral of mobile notifications. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3229434.3229436

[74] Dominik Weber, Alexandra Voit, and Niels Henze. 2019. Clear All: A Large-Scale Observational Study on Mobile Notification Drawers. In *Proceedings of Mensch und Computer 2019 (MuC'19)*. Association for Computing Machinery, New York, NY, USA, 361–372. https://doi.org/10.1145/3340764.3340765

[75] Dominik Weber, Alexandra Voit, Gisela Kollotzek, and Niels Henze. 2019. Annotif: A System for Annotating Mobile Notifcations in User Studies. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia (MUM '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3365610.3365611

[76] Dominik Weber, Alexandra Voit, Philipp Kratzer, and Niels Henze. 2016. In-situ investigation of notifications in multi-device environments. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 1259–1264. https://doi.org/10.1145/2971648.2971732

[77] Tilo Westermann, Sebastian Möller, and Ina Wechsung. 2015. Assessing the Relationship between Technical Affinity, Stress and Notifications on Smartphones. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 652–659. https://doi.org/10.1145/2786567.2793684

[78] Ting-Wei Wu, Yu-Ling Chien, Hao-Ping Lee, and Yung-Ju Chang. 2021. IM Receptivity and Presentation-type Preferences among Users of a Mobile App with Automated Receptivity-status Adjustment. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445209

[79] SungHyuk Yoon, Sang-su Lee, Jae-myung Lee, and KunPyo Lee. 2014. Understanding notification stress of smartphone messenger app. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 1735–1740. https://doi.org/10.1145/2559206.2581167

[80] Ye Yuan, Nathalie Riche, Nicolai Marquardt, Molly Jane Nicholas, Teddy Seyed, Hugo Romat, Bongshin Lee, Michel Pahud, Jonathan Goldstein, Rojin Vishkaie, Christian Holz, and Ken Hinckley. 2022. Understanding Multi-Device Usage Patterns: Physical Device Configurations and Fragmented Workflows. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–22. https://doi.org/10.1145/3491102.3517702

[81] George Kingsley Zipf. 1949. *Human behavior and the principle of least effort*. Addison-Wesley Press, Oxford, England. Pages: xi, 573.