“I Got Some Free Time”: Investigating Task-execution and Task-effort Metrics in Mobile Crowdsourcing Tasks

Chia-En Chiang
National Chiao Tung University
Hsinchu, Taiwan
vivian123.cs02@g2.nctu.edu.tw

Yu-Chun Chen
National Taiwan University
Taipei, Taiwan
lesleychen@ntu.edu.tw

Fang-Yu Lin
National Chiao Tung University
Hsinchu, Taiwan
emilylin3130.aa07@nctu.edu.tw

Felicia Feng
International Bilingual School at Hsinchu-Science-Park
Hsinchu, Taiwan
feliciafeng.ff@gmail.com

Hao-An Wu
National Chiao Tung University
Hsinchu, Taiwan
haoan.cs09g@nctu.edu.tw

Hao-Ping Lee
National Chiao Tung University
Hsinchu, Taiwan
dimension4.cs03@nctu.edu.tw

Chang-Hsuan Yang
National Chiao Tung University
Hsinchu, Taiwan
j09190926.06g@g2.nctu.edu.tw

Yung-Ju Chang
National Chiao Tung University
Hsinchu, Taiwan
armuro@cs.nctu.edu.tw

ABSTRACT

Using a mixed-methods approach over six weeks, we studied 30 smartphone users' task choices, task execution and effort devoted to two commercial mobile crowdsourcing platforms in the wild. We focused on the influence of activity contexts, characterized by breakpoint situations and activity attributes. In line with their stated preferences, the participants were more likely to proactively perform mobile crowdsourcing tasks during transitions between activities than during an ongoing activity and during long breaks, respectively. Their task choices were influenced by various activity attributes, and more impacted by their current and preceding activities than their upcoming ones. Two of our three target outcomes, task execution and task choice, were also influenced by individuals' stress and energy levels. Our qualitative data provide further insights into participants' decisions about which crowdsourcing tasks to perform and when; and our results' implications for the design of future mobile crowdsourcing task-prompting mechanisms are also discussed.

CCS CONCEPTS

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

KEYWORDS

Mobile crowdsourcing; interruption; notification; ESM, mixed-effect logistic regression; qualitative analysis

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05...$15.00
https://doi.org/10.1145/3411764.3445477

ACM Reference Format:
Chia-En Chiang, Yu-Chun Chen, Fang-Yu Lin, Felicia Feng, Hao-An Wu, Hao-Ping Lee, Chang-Hsuan Yang, and Yung-Ju Chang. 2021. “I Got Some Free Time”: Investigating Task-execution and Task-effort Metrics in Mobile Crowdsourcing Tasks. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3411764.3445477

1 INTRODUCTION

To improve the services they provide to their users, an increasing number of entities are collecting data via crowdsourcing. In recent years, the performance of crowdsourcing tasks on smartphones have become more prevalent, with examples including Waze[1], Local Guide[2], and Google Crowdsource[3]. This not only allows people to perform mobile crowdsourcing tasks nearly anywhere and at any time, but also allows entities seeking specific types of crowdsourced information to prompt mobile users to contribute them [11, 20]. However, whether such users will actually contribute data at any given moment appears to depend largely on their receptivity[47].

Finding opportune moments at which users are receptive to mobile crowdsourcing tasks, however, remains unclear. That is, prior research attempts have sought to identify moments at which users are receptive for responding to messages [26], reading news articles [42], reading advertisements [55], responding to interventions [25, 44, 47], responding to questionnaires [24, 41], and so on. By and large, such moments have been found to relate to users' prevailing contexts, including their personal condition (e.g., emotional state [24]), the characteristics of whatever activity they are involved in [24, 28], their level of engagement with that activity [28, 39], and

[1] Waze is a global positioning system-based navigation app, owned by Google, and which provides turn-by-turn navigation information and user-submitted travel times and route details, and downloads location-dependent information over mobile-telephone networks.
[2] Local Guides was launched by Google to garner user contributions to Google Maps, and provides its contributors with various perks and benefits for this work.
[3] Crowdsource is a crowdsourcing platform developed by Google, intended to improve a host of Google services via user-facing training of various algorithms.

their social context [17]. Others suggested that receptivity is also dependent upon notification content [10, 16]. However, it remains unclear whether the moments users consider to be opportune for dealing with the range of notification types mentioned above would also be deemed opportune for responding to commercial mobile crowdsourcing tasks – especially because such tasks are, for the most part, less personally relevant. Commercial mobile crowdsourcing services also typically offer a menu of dissimilar tasks at any given time, but how and why their users' task choices vary across contexts and their personal conditions also remains under-explored. A better understanding of these dynamics is vitally important to mobile crowdsourcing services aiming to increase their users' contributions via task-invitation prompts.

Inspired by several prior studies that reported smartphone users are more receptive to incoming tasks and interruptions during breakpoints [1, 32, 33, 36], the present paper focuses primarily on mobile crowdsourcing contributors' activity contexts and personal conditions. Extending Mehrotra et al.'s [31] study, such activity contexts will be examined in two main dimensions: breakpoint situations and activity attributes, where the former can be further subdivided into 1) within an activity, 2) between two different activities, 3) preceded by but not succeeded by an activity, 4) not preceded by but succeeded by an activity, and 5) neither preceded by nor succeeded by an activity. Personal conditions include stress level and energy level. Our two primary research questions are:

- RQ1. How is people's performance of mobile crowdsourcing tasks influenced by their two personal conditions and various activity contexts, respectively ?
- RQ2. How are people's mobile crowdsourcing task choices influenced by their two personal conditions and various activity contexts, respectively?

To answer these questions, we conducted a six-week mixed-methods study with 30 smartphone users. The quantitative phase of this research included experience sampling method (ESM) questionnaires and screen-recording, and its qualitative phase comprised semi-structured interviews. This paper's four main contributions to the literature on mobile crowdsourcing are as follows.

1. It shows that participants were more likely to proactively perform tasks during transitions than during an ongoing activity and during long break time, respectively, and that they also universally deemed activity transitions to be opportune for performing mobile crowdsourcing tasks, as a way to "kill time" over during a long break.
2. It establishes that different task choices were influenced by differing sets of activity attributes, as well as by individuals' stress and energy levels; and were generally more influenced by a person's previous and ongoing activities than by their next planned ones.
3. It provides qualitative insights into the factors that influenced the participants' decisions about when to perform crowdsourcing tasks, and which ones to perform.
4. It highlights important design implications for future mobile crowdsourcing task-prompting mechanisms.

2 RELATED WORK

2.1 Mobile Crowdsourcing

Crowdsourcing is a powerful way to collect data from a large group of people, and has been used to harvest a variety of types of information and to solve a variety of problems. Mobile crowdsourcing is characterized by its collection of data in mobile environments, commonly linked to specific locations: an aspect sometimes referred to as situated crowdsourcing [12], environmental information. Researchers have leveraged this approach to collect a variety of information in the wild, including about local events [2, 52], transportation [4, 54], noise [27], accessible facilities and accessibility problems [21, 45, 48], status of points of interest [7], animal populations [49], and emergencies [22, 46].

In recent years, researchers devised various approaches to increasing contributions from the crowd. Kandappu et al. [18] developed TASKer, which increases the crowd's productivity by bundling tasks and offering multiple pricing models. Garg et al. [11] proposed a framework called 4X, which combines four collection methods to increase contribution opportunities and thus data quantity. Kim et al.'s [20] Hit-or-Wait allocates crowdsourcing tasks and notifies contributors based on their mobility patterns; and Kandappu et al. [18] also found that recommending tasks according to predicted movement patterns led to higher completion rates. Pestana et al. [40] built a model for efficient task distribution, which succeeded in increasing contributions. Recently, Kandappu et al. [18] proposed delivering tasks based on contributors' locations and time spent at those locations to improve task-acceptance rates and to balance the quality of data received from multiple locations. Finally, Crowd-PickUp [12] combines mobile and situated crowdsourcing to ensure that a sufficient number of contributors are recruited, and that a wide breadth of tasks is deployed.

In addition to reporting location-specific information, researchers have explored using mobile crowdsourcing to collect annotated data for the purpose of improving machine learning models. Chang et al. [5] studied three labeling methods for collecting labeled transportation, activity data, and compared their resulting data quality and quantity. In 2016, Google released a commercial smartphone app called Crowdsource, which offers a number of common crowdsourcing tasks geared to this purpose, including handwriting recognition, image tagging, and sentence translation. The Crowdsource app does not provide their users with any monetary incentives [8, 9], so their motivations vary widely, the top three being that they feel app use gives them recognition, provides a game-like experience [8, 9], and satisfies their curiosity. Similar results were also reported in Chi et al.'s [9] study of image-capturing tasks , to which people contributed mostly because they believed they were helping technologies to improve; the innate entertainment value of the tasks was cited as their next most important reason.

The recent marriage of crowdsourcing tasks with smartphones offers opportunities for prompting contributors to perform tasks on their phones in diverse contexts. Users of Crowdsourcing Go [14] have reported perceptions that busyness, fatigue, and the presence of companions all affect their completion rates, task prices, and task qualities. However, the kinds of contexts that may be associated with greater and lesser willingness to perform crowdsourcing tasks when receiving task requests on the phone remains under-researched; for example, it remains unclear how users’ mobile crowdsourcing task execution is impacted by the activity context, such as presence of the preceding, current, and/or subsequent activity, respectively, as well as personal conditions, such as individuals’ energy and stress level.

The present study applies receptivity research to mobile crowdsourcing, seeking to understand the kinds of activity context and personal conditions that render individuals more likely to perform tasks on commercial mobile-crowdsourcing platforms.

2.2 Interruption and Receptivity Research

Mobile crowdsourcing task requests that are delivered to individuals’ smartphones can be considered a kind of interruption for smartphone users. Such interruptions can arrive at times users feel are inappropriate, and thus have a range of negative impacts [15]. The negative impact of interruptions has been revealed in many studies [3, 24, 26, 28, 29, 31, 39, 50], and users have used strategies to control phone alerts to manage whether they want to be aware of such interruptions (e.g. [6].) Analysis of the impact of interruptions has mostly focused on two aspects: the characteristics of the interruption itself, and the contexts in which people are interrupted. In the case of the former, prior research has found that individuals’ interest in interruptions’ content and their perceptions of the relevance of that content positively influence their receptivity to processing those interruptions [10, 16]. In the case of the latter, the contextual factors that have been identified in desktop settings [3, 29, 50] and mobile ones [24, 26, 31] have included the characteristics of users’ activity at the time of the interruption; their levels of engagement with that activity; their psychological condition; and their social context. Moreover, Mehrotra et al. [31] have suggested that both response time to a notification and perceptions of its disruptiveness can be influenced by the complexity of the task in which the receiving party is engaged. Mark et al. [28] studied how receivers’ original attentional states influenced their susceptibility to interruptions, and reported people are prone to particular types of interaction with certain attentional states. Lee et al. [26] found that activity engagement was negatively correlated with receptivity to messaging notifications. Similarly, Pejovic et al.’s [39] found that task engagement played a significant role in determining interruptibility. And Kushlev et al.’s [24] analysis of whether people’s feelings affected their choice of what kinds of content to engage with suggested that they are less attentive to distractions when they feel energetic.

Another large body of research aims to identify and predict opportune moments for receiving interruptions, at which users are presumed to be receptive to the interruptions. For example, many researchers have leveraged information extracted from people’s phones to predict when receptivity to notifications will be high (e.g., [41, 43, 51]), sometimes supplemented by information on the sender-recipient relationship [30] or activity, location, and emotional context [38]. Pielot et al. [41] built a model aimed at predicting opportune moments for engagement with different types of content, and found it to be effective.

Breakpoints, in particular, have been suggested as ideal times for notification delivery. Iqbal et al. [15] scheduled notifications at breakpoints among one subgroup of participants, and found that these recipients’ reaction times and frustration levels were both lower than those of recipients in a baseline condition. Adamczyk et al. [1] used coarse and fine units to describe different kinds of within-task breakpoints and to detect such moments for delivering interruptions; this approach also resulted in lower frustration/annoyance. Park et al. [37] identified four distinct types of breakpoints – long silences; a participant leaving a table; others in a group using their own smartphones; and a participant being left alone – and found that sending notifications only during these breakpoints resulted in a (54.1%) decrease in perceptions that notifications were interruptive. Robertson et al. [44] developed a system for delivering prompts during transitions between activities, and showed that this approach was not only more favored by users than during non-transition, but that it made initial prompts more effective. Similarly, albeit not using breakpoints, Isaacs et al. [17] proposed the concept of microwaiting to describe brief times of idleness during an activity, and suggested that people frequently requested content during such moments. Pielot et al. [42], on the other hand, predicted moments when smartphone users were bored, and reported that they were more likely to engage with recommended content at such times.

Finally, Okoshi et al. [32] proposed Attelia, a middleware that could detect breakpoints of users’ activity on smartphone and showed 80%-90% accuracy in detection. Later, Okoshi et al. [33] showed notification delivery based on breakpoint detection of Attelia resulted in 46% lower cognitive load compared with randomly-timed notifications. Okoshi et al. [34] also introduced physical activity-based breakpoint detection, together with UI Event-based breakpoint (i.e., users’ activity on smartphones) and showed 71.8% greater reduction of users’ perception of workload. One step further, Okoshi et al. [35] deployed Attelia on large-scale mobile application to investigate user interruptibility and engagement based on breakpoint detection. It further showed that, delivering notifications during these breakpoints decreased its users’ perceived workloads and increased their click rates in advertising for a real-world product [36]. Yet, while this series of studies provide rich insights into characteristics of low-level and fine-grained breakpoints such as physical activities and UI breakpoints, the impact of high-level activity context on mobile crowdsourcing task performance remains unexplored. Beyond these studies in interruption and receptivity research, which were focused on personal-relevant interruptions, we attempt to clarify opportune moments for mobile crowdsourcing tasks and show that activity context and personal condition impact both task execution and task choices.

The three subsections below describe our research constructs in more detail.

3 METHODOLOGY

3.1 Activity Context and Personal Conditions

The first of the two dimensions of activity context, breakpoints, are here characterized according to their temporal relationships with previous, subsequent, and concurrent activities. Such a typology, which was inspired by Okoshi et al. [32, 34, 36], and Mehrotra et al. [31] results in five types of breakpoint situations: 1) within an activity (WITHIN); 2) between two different activities (BETWEEN); 3) preceded by but not succeeded by an activity (PRECEDED); 4)

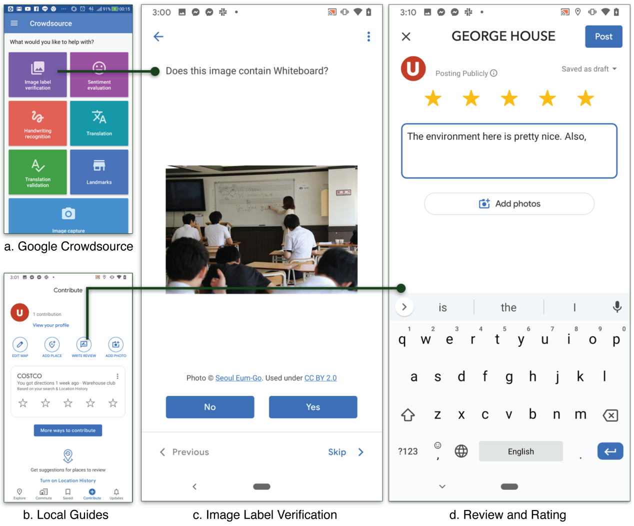

Figure 1: The mobile crowdsourcing tasks, (a) Google Crowdsource interface, (b) Local Guides interface, (c) Image Label Verification on Google Crowdsource, and (d) Review and Rating on Local Guides.

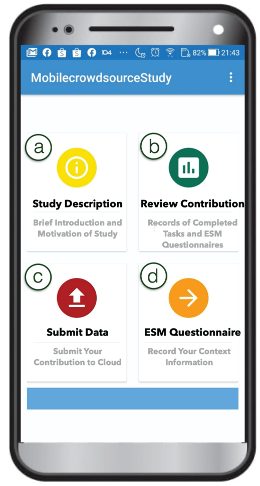

Figure 2: The research app interface, showing its (a) description of the study, (b) its "Review contribution" area, (c) its "Submit data" area, (d) and its ESM questionnaire area.

not preceded by but succeeded by an activity (SUCCEEDED); and 5) neither preceded by nor succeeded by an activity (NONE). The second, third, and fourth of these activity-context subdimensions are referred to as transitions, because they are either preceded by or succeeded by another activity. The second activity-context dimension, activity attributes, was inspired by prior findings that the attributes of their activities may influence people's receptivity to interruption [39], each preceding, succeeding, and current activity was characterized by four attributes: 1) perceived complexity, 2) perceived physical effort required, 3) perceived mental effort required, and 4) perceived attention required. Finally, for personal conditions, we consider stress level and energy level, inspired by [24, 47].

3.2 Target Outcomes of Tasks

3.2.1 *Task Characteristics.*

Inspired by prior findings that characteristics of the interrupting task matter to individuals' receptivity to that task [24], we were also interested in the characteristics of the tasks users chose to perform, and how those task characteristics varied across a range of their other activities. We therefore characterized crowdsourcing tasks with the same four attributes: 1) perceived complexity, 2) perceived physical effort required, 3) perceived mental effort required, and 4) perceived attention required.

3.2.2 *Task-execution Outcomes and Effort Devoted.*

We conceive of every crowdsourcing task as having one of three mutually exclusive outcomes, ranked in order of desirability from the point of view of a mobile crowdsourcing service, as follows: 3 (most desirable), proactive execution, i.e., the task was self-initiated; 2: reactive execution, i.e. the task was completed after the issuance of one or more reminders; and 1 (least desirable), non-execution. In our analysis, we treat these three possible task-execution outcomes as an ordinal variable with three levels. The main goal of that phase of our analysis is to find out when proactive execution is most likely, and non-execution least likely, to occur.

Additionally, for each task session – irrespective of whether it was initiated proactively or reactively – we collected two effort devoted measures: 1) length, i.e., the total amount of time users spent in that task session; and 2) number of actions, i.e., the quantity of discrete events within that session, such as 'items responded to', 'photos uploaded'. Task sessions were distinguished from other participant activities by means of ESM responses, as will be described more fully below.

3.3 Mobile Crowdsourcing Apps

In this study, we focused on the use of two existing commercial mobile crowdsourcing services, both with more than 1 million users around the world. The first was Local Guides[4] (Fig. 1(b)), a crowdsourcing platform built into Google Maps[5] that encourages people to contribute location information. The categories of mobile crowdsourcing tasks on Local Guides include Reviews, Ratings, Photos, and Q&A. The second was Google Crowdsource[6] (Fig. 1 (a)): a standalone platform that offers a variety of crowdsourcing tasks such as Image Label Verification, answering Yes/No questions about labeled images, and Handwriting Recognition, i.e., submitting one's transcription of handwritten letters. A complete list of the available tasks can be found in Google Maps and Crowdsource .

3.4 Research Instrument

We developed an Android research app (see Fig. 2), which 1) captured the moments at which its users open either of the two focal crowdsourcing apps, 2) screen-recorded their task execution, and

[4] https://maps.google.com/localguides
[5] https://play.google.com/store/apps/details?id=com.google.android.apps.maps
[6] https://play.google.com/store/apps/details?id=com.google.android.apps.village.boond

3) delivered ESM questionnaires. Specifically, it utilized the Accessibility Service in Android to record the screen content and users' interactions so that we could later identify whether they entered either of the two focal crowdsourcing apps. The Accessibility Service detected possible pages in the target apps via specific keywords (e.g., "contribution", "helpful", "missing", "translation") that were indicative of mobile crowdsourcing tasks being conducted. Due to the Accessibility Service's limited capacity to acquire all the elements contained in the screens of the two crowdsourcing apps, our research app adopted screen recording to capture the process of crowdsourcing task execution. A screen-recording prompt was triggered whenever an app-entering event was detected, inviting the user to begin recording. Our participants could always opt out of recording if they felt uncomfortable. They could also end the recording manually at any time, or automatically, by closing whichever app was being recorded.

After a participant finished any crowdsourcing task, the research app prompted him or her to answer an ESM questionnaire. Its interface, as shown in Fig. 2, allowed its users to review their task contributions, refer to a description of the present research project, submit data, and fill out ESM questionnaires. The data-upload process (Fig. 2 (c)) was initiated by the user when s/he thought connectivity was suitable for uploading large data files such as videos.

3.4.1 Task Reminders.

To compare task execution in different activity contexts, we deemed it important to remind our participants to perform mobile crowdsourcing tasks via mobile prompts across contexts throughout a day. Thus, our research app sent multiple task-reminder prompts on the participant's phone's notification system randomly between 8 am to 10 pm, at intervals of between 90 minutes and 2 hours. To mitigate users' sense of such reminders as interruptive, our app did not send a prompt if a reminder from either of the focal crowdsourcing apps had been sent within the previous hour. In addition, whenever the user self-initiated a crowdsourcing task, the system canceled the next reminder that it would otherwise have sent. Reminders could be dismissed, either by the participant, or by the research app itself if no user response had been received within 30 minutes. Eventually, our participants received on average 2.84 prompts per day

3.4.2 Experience Sampling Method.

ESM is widely used to collect smartphone users' *in situ* experiences via questionnaires. In this study, we sent ESM questionnaires to a participant in the following two circumstances: 1) when mobile-crowdsourcing tasks were detected on his/her phone, and s/he then left Google Crowdsource or Local Guides; and 2) when mobile-crowdsourcing tasks were issued to that participant directly by one of the focal crowdsourcing apps, or when a reminder prompt was sent by our research app. Participants could respond to ESM prompts either via the notifications that had been sent to them, or by opening the research app. To minimize self-reporting inaccuracies caused by recall errors, all ESM questionnaires were dismissed if responses to them had not commenced within 15 minutes of their arrival.

Each ESM questionnaire[7] comprised two dimensions: context, and crowdsourcing tasks. It began with two questions, one about whether a mobile crowdsourcing task had just been completed, and the other about what type of breakpoint situation the user would categorize themselves as being in when completing the task or when receiving the prompt (depending on how the ESM had been triggered). The answers to these two questions determined the follow-up questions they would answer. If someone reported having completed at least one crowdsourcing task (even if that task were an element of a longer task session), s/he received follow-up questions about 1) whether that task had been self-initiated, vs. prompted by a reminder to perform it; and 2) the summarized attributes of the task(s) they had just completed. This structure allowed us to measure the participants' task-execution outcomes.

Follow-up questions aimed at capturing activity-context information varied, depending on the types of breakpoint situations users identified. These questions included, for each relevant activity (preceding, current, and succeeding), its four attributes and elapsed time. For example, if a participant said that they were engaged in an activity at the time the follow-up questions were being asked, s/he was asked about the four attributes of that activity and how long it had been since it started. If s/he had just finished an activity and perceived that another one was about to begin, however, the follow-up questions asked for the four attributes of both the preceding and the succeeding activity; how much time had passed since the preceding activity ended; and how much time was likely to pass before the next activity began. In addition to these breakpoint situation and task-dependent questions, each ESM asked the participants for contextual and personal condition information, including their locations, energy levels, stress levels, and interactions with any companions.

3.5 Recruitment and Participants

We recruited participants via advertisements posted on social-media pages, including one tailored to recruiting research participants in our country. The 30 participants recruited in this way had used Local Guides and/or Google Crowdsource before; and had left comments on online crowdsourcing platforms. They included 13 students and 17 non-students, and 10 females and 20 males, aged between 20 and 64 (M = 29.2, SD = 10.3). The experiments took place in both urban and rural places in Taiwan from April to July 2020. Only one participant had experience working on paid crowdsourcing platforms. All participated in our study for at least six weeks, with some volunteering to extend the period of their involvement because they wanted to complete more tasks.

3.6 Study Procedure

Due to the COVID-19 pandemic situation, we allowed participants to choose whether to attend a physical or a remote pre-study meeting, during which the researchers explained the study procedure and helped them install the app on their phones. Eighteen participants had chosen to be briefed about the study remotely. Every two weeks, the participants were invited to a short, optional, semi-structured interview aimed at helping the researchers gain additional, richer information about their experience of executing

[7] The ESM questionnaire in this study was attached in the supplemental materials.

mobile crowdsourcing tasks. The participants were encouraged to complete at least three ESM questionnaires per day and 10 mobile crowdsourcing tasks per week. Despite the two focal crowdsourcing apps not providing monetary incentives, we assume that the gamification mechanisms and reputation systems of the two platforms could also motivate our participants to contribute, which has also been reported in [8]. Nevertheless, acting on an assumption that future crowdsourcing platforms will also offer monetary incentives, along the lines of Amazon Mechanical Turk[8], we told our participants they would receive NT$3 (approximately US$0.10) for every crowdsourcing task that they completed. For every ten mobile crowdsourcing tasks they completed, they obtained additional NT$25 (i.e., a total of 20 fulfilled tasks led to a bonus of NT$50; 30, to a bonus of NT$75, and so forth). They were also offered NT$300 for every interview session they participated in, to a maximum of NT$900 if they participated in all three such sessions.

3.7 Data Cleaning and Analysis

A total of 7,699 ESM questionnaires were issued, meaning that each participant received an average of 6.11 ESM prompts per day. We received a total of 4,087 ESM responses (a response rate of 53%), and 3,360 screen-recording clips. Five members of the research team coded the recordings separately. For each task session, the codes they assigned covered task types, task lengths, and the number of actions involved in each task. These five coders first selected a subset of 10.2% of the 3,360 recordings (n = 346), and calculated the inter-rater reliability of their video coding, using ICC for both number of actions and task length [23]. The single-rater agreement ICC and mean agreement between raters of the former were .8 and .95, and of the latter were .99 and 1.0, suggesting that both measures had high reliability. After this high level of agreement was attained, the five members coded the rest of the video clips independently; and the full research team then verified all ESM responses by matching them to their corresponding clips. This process led to 879 video clips being removed, either because a corresponding questionnaire was not found, or because the video did not show any task being completed. Next, the research team removed all responses that contained mismatches between what the video clips showed and the ESM responses the relevant participants had given. This led 484 responses to be removed from further consideration. The final dataset that we analyzed therefore consisted of 3,603 ESM responses. Among these, 2,481 (68.8%) indicated that a given participant had performed at least one mobile-crowdsourcing task. Slightly over half (n = 1,910, 56.8%) of these 3,603 ESM responses were categorized as WITHIN; 821 (22.7%) as BETWEEN; 302 (8.3%) as PRECEDED; 232 (6.4%) as SUCCEEDED; and 338 (9.4%) as NONE. The final dataset of recordings reflected 2,889 completed task sessions, including 90 hours, 44 minutes and 43 seconds of task execution; and 101,502 discrete task actions. More than one type of task was involved in 11.5% of task sessions (n = 333), suggesting that in most of the task sessions participants preferred repeatedly performing the same type of tasks. The average task session lasted

[8] It should be noted here that the number of actions was more vague than length of time because, from the screen recording, it was sometimes difficult to tell clearly if a transition had been triggered by action or by the app itself. Thus, the ICC of number of clicks was relatively lower than that of length of time.

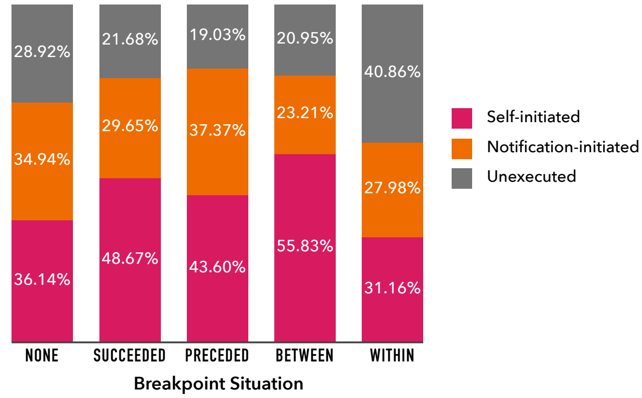

Figure 3: Percentages of the three focal execution outcomes across the five types of breakpoint situations.

slightly under two minutes (113.1 seconds, SD = 106.65), and its average number of actions was 35.13 (SD = 56.63).

We used mixed-regression models for most of our quantitative analyses. A random effect that accounted for individual differences between the participants was included, because each participant contributed a number of sample points. Linear regression was used to examine the influence of a variable on a numerical outcome, such as effort devoted, i.e., time spent and number of actions. Logistic regression was used to examine effects on the odds of a given outcome occurring. Ordinal logistic regression, in particular, was used to examine effects on task-execution outcomes. In addition to the targeted main effect, fixed effects of energy level, stress level, and companion-interaction were included; and we used a chi-square test of independence to examine the association between two variables.

For qualitative analysis, we used a bottom-up process of affinity diagramming [13, 56], which involved iterative grouping and labeling. The elements of the affinity diagram consisted of notes transcribed from audio recordings of the interviews. In each grouping and labeling session, the researchers discussed notes and walked through to the affinity wall together and re-grouped whenever necessary. The process resulted in ten high-level themes. Overall, our findings fall into two main categories: 1) factors affecting whether or not participants performed the tasks. 2) factors affecting how participants chose the tasks. Factors included preceding activities, succeeding activities, current activities, activity transition, mood, stress, energy, and reminder prompt.

In the following sections, we present the highlights of our findings.

4 RESULTS: TASK-EXECUTION OUTCOMES

4.1 More Proactive Execution at Transitions

Our participants were more likely to proactively initiate mobile crowdsourcing tasks at transitions (BETWEEN: 55.83%; PRECEDED: 43.6%; SUCCEEDED: 48.67%) than when they were engaged in an activity (WITHIN, 31.16%) or experiencing a lengthy period of idle time (NONE, 36.14%). Their likelihoods of not performing any task were also lower during transitions (BETWEEN: 20.95%; PRECEDED: 19.03%; SUCCEEDED: 21.68%) than when they were engaged in an activity (WITHIN, 40.86%) or inactive for a lengthy period (NONE, 28.92%), as Fig. 3 shows. The differences between transitions and NONE (Z = -2.22, p < .001), and between transitions and WITHIN (Z = -10.78, p = .03) were both statistically significant. While we expected to see less self-initiation of tasks when participants were in the middle of some other activity, it was surprising to see that self-initiation was also low when they had no recent, ongoing, or planned activities at all.

It is also noteworthy that the percentage of self-initiated tasks in both SUCCEEDED (48.67%) and BETWEEN (55.83%), were markedly higher than in PRECEDED (43.6%) and NONE (36.14%), i.e., the two activity categories without any activity after them. We used a chi-square test of independence to examine the proportions of proactiveness among breakpoint situations, the results of which indicated that the association was highly significant ($\chi^2_{(4,\ N=3,438)}$ = 152.68, p < .001). This implies that when they had no activity planned for afterwards, our participants were also less likely to proactively initiate crowdsourcing tasks than they were if there was an upcoming activity.

4.2 Relations of Proactive Task Execution to Stress, and of Effort Devoted to Energy

We found a strong negative correlation between the participants' stress levels and the likelihood of their proactively initiating tasks (Z = -5.98, p < .001). We did not observe any significant correlation between task initiation and energy levels (Z = .63, p = .527). However, once a participant had started a task session, whether self-initiated or not, there were strong positive correlation between energy level, and the time s/he devoted to tasks within that session and the number of actions taken in it (Length: t(2,316.35) = 2.87, p < .001; Number of Actions: t(2,323.69) = 4.28, p < .001). Stress level was only found to have a marginally significant negative correlation number of actions, and no significant relation with time spent (Length: t(2,323.79) = -1.44, p = .149; Number of Actions: t(2,327.94) = -1.87, p = .0613).

4.3 Relations of Breakpoint Situation and Proactiveness to Task Effort Devoted

Our participants' number of actions per session tended to be higher when the session was self-initiated (Length: M = 114.92 sec, SD = 119.08 sec; Number of Actions: M = 57.55, SD = 62.89) than when it had been the subject of a reminder (Length: M = 79.22 sec, SD = 82.12 sec; Number of Actions: M = 34.61, SD = 43.17). However, proactiveness' apparent effect on number of actions (t(2,306.98) = 2.36, p = .0183) was not paralleled by any similar effect on time spent per session (t(2,327.84) = .53, p = .5988).

4.4 Activity Attributes on Task-execution

We found a strong negative main effect of the attention required by a person's ongoing activity (Z = -4.54, p < .001) on his/her task-execution outcome. We did not find any effects of the other three activity attributes. This result was supported by our qualitative finding that, if they were concentrating on something else, participants

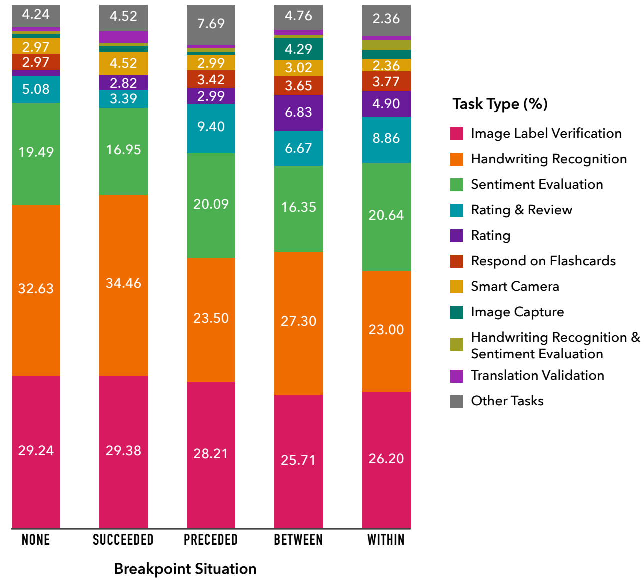

Figure 4: Percentages of the task choice outcomes across the five types of breakpoint situations. Note that the numbers in the table do not share the label, e.g., 29.24 means 29.24%.

felt they were less likely to perform mobile crowdsourcing tasks. We also found a marginal negative effect of the perceived mental effort required for participants' succeeding activities on their task-execution outcomes (Z = -1.92, p = .055). This, too, was supported by our qualitative data: with interviewees reporting lower willingness to perform crowdsourcing tasks when they felt a need to mentally prepare for their upcoming activities.

5 RESULTS: TASK CHOICE

5.1 Top Choices of Crowdsourcing Tasks

Based on the coded video clips, we were able to confirm that the 30 participants performed 17 distinct types of crowdsourcing tasks. Fig. 4 presents their task choices across all task sessions, including those containing a single task type and those containing two or more different types. The top five choices, which made up more than 84% of all task sessions, were: Image Label Verification (26.8%), Handwriting Recognition (26%), Sentiment Evaluation (19%), Rating and Review (7.5%), and Rating (4.7%). A chi-square test of independence showed that task choices were not independent from breakpoint situations ($\chi^2_{(20,\ N=2,338)}$ = 48.09, p < .001).

5.2 Attributes of Mobile Crowdsourcing Tasks

Fig. 5 presents the average ratings assigned to each task attribute for the top 10 crowdsourcing tasks participants performed (with tasks performed fewer than 20 times not shown in the graph). On the whole, the participants perceived that all types of mobile crowdsourcing tasks required little physical effort and were not complex. They also tended to perceive that translation-related tasks required high attention and mental effort. The discrepancies in the ratings they assigned to the four focal task attributes suggest that they

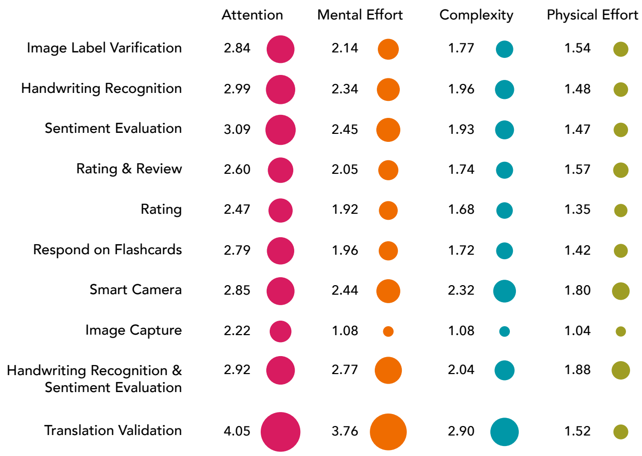

Figure 5: User-reported task attributes for top 10 task choices during the study, including (a) average attention required, (b) average mental effort required, (c) average task complexity, and (d) average physical effort, sorted by task type.

thought these attributes differed from one another conceptually. Particularly in the cases of attention and mental effort, the most noticeable discrepancies occurred in connection with the tasks Image Label Verification and Sentiment Evaluation: i.e., the participants considered these task types to require low to moderate mental effort, but relatively high attention. The smallest difference among the four attributes was observed in connection with the Smart Camera task type.

5.3 Breakpoint Situations on Task Choices

We found that the tasks participants performed in NONE involved more physical effort than those they chose to do when in other breakpoint situations (NONE vs. SUCCEED: $t(2,308) = -2.04$, $p = .04$; NONE vs. PRECEDED: $t(2,313) = -2.68$, $p = .007$; NONE vs. BETWEEN: $t(2,323) = -3.39$, $p < .001$; NONE vs. WITHIN: $t(2,321) = -2.17$, $p = .03$). This seems to suggest that the more distant in time the participants perceived themselves as being from a non-crowdsourcing-related activity (with NONE being the farthest, WITHIN the closest, and BETWEEN the second closest), the higher the physical effort of the tasks they were likely to perform.

5.4 Activity Attributes on Task Choices

The attributes of the tasks our participants selected were closely correlated to the attributes of their proximate activities. Fig. 6 presents the regression results of all models of our four focal task attributes across three conditions. These conditions were created by combining our five breakpoint situations as follows, i.e., Condition 1: WITHIN, Condition 2 (preceded by activities): PRECEDED + BETWEEN, and Condition 3 (succeeded by activities): SUCCEEDED + BETWEEN. Among these results, three trends are especially worth highlighting. First, each task attribute was correlated with a different set of activity attributes; and very interestingly, perceived attention required and physical effort required did not share any of the same effects of activity attributes across any of the three above-mentioned conditions, suggesting that people's choice of tasks that were physically vs. attention oriented were driven by distinct sets of activity attributes. Second, whereas people's task-execution outcomes were most influenced by the attributes of their current and succeeding activities, we found that their task choices were influenced more by their current and preceding ones. There were only a few effects of succeeding-activity attributes on task attributes, whereas many attributes of current and preceding activities were influential on task attributes. In particular, as shown in Fig. 6, most attributes of current and preceding activities were positively correlated to task attributes; the exception being complexity. That is, the more complex a person's current and preceding activity were, the less likely s/he was to perform a mobile crowdsourcing task that required high attention and high mental effort.

Finally, and also interestingly, each task attribute was positively correlated with a corresponding attribute of the participant's current activity, again, the exception being complexity. This implies that people may tend to choose tasks that are in some way similar to what they are already doing, e.g., select crowdsourcing tasks that require high attention when their attention levels have already been elevated by non-crowdsourcing tasks. All the correlations relevant to these three trends were very strong (all p-values < .001).

5.5 Relations of Stress Level and Energy Level to Task Choice

We found differential effects of stress level and energy level on task choice. Overall, the influence of stress mainly emerged in WITHIN, BETWEEN or PRECEDED. During such moments, the more stress the participants felt, the more likely they were to choose tasks that were relatively complex and required relatively high mental effort. These findings seem unintuitive at first glance. However, it should be borne in mind that task choice only arose when a task session had been initiated, either self-initiated or prompted. In other words, they were less likely to initiate tasks when they felt stressed, but once they did choose to start a task, it appeared to release some of the stress that had built up during their previous or ongoing activity. In their interviews, many participants mentioned treating tasks as a way to refresh themselves (as will be discussed further below). On the other hand, the influence of individuals' energy levels mainly emerged when they were engaged in an ongoing activity or when they perceived that another was upcoming. Specifically, when an activity was upcoming, participants who felt more energetic were more likely to choose tasks that were more complex and required more attention.

6 QUALITATIVE FINDINGS

6.1 Preferences Regarding "Free Time" for Task Execution

Our interview data regarding how participants selected when to perform mobile crowdsourcing tasks were quite coherent with the quantitative results presented above. All the interviewees mentioned that they performed tasks when they had free time, and frequently mentioned their preference for leveraging transition times between activities to task execution. Many said they initiated tasks without prompting during such periods. The main reason

Task	Complexity	Attention	Physical Effort	Mental Effort	Stress	Energy	Interact or Not
Condition 1: Within an Activity							
Complexity	-.01	-.01	.02	-.01	.10	.03	-.02
Attention	-.10	.07	-.02	.08	.00	.08	-.08
Physical Effort	-.03	-.00	.07	.00	.08	.02	.05
Mental Effort	-.10	-.06	.03	.13	.10	-.01	-.14
Condition 2: Preceded by an Activity							
Complexity	-.03	-.04	.07	.06	.07	.04	-.08
Attention	-.09	.02	.04	.06	.07	.03	-.03
Physical Effort	-.00	-.03	.04	.04	.04	-.00	-.01
Mental Effort	-.01	-.03	.09	.09	.11	-.01	-.05
Condition 3: Succeeded by an Activity							
Complexity	.05	-.03	.03	-.03	.01	.05	-.06
Attention	.01	.00	.06	-.04	.03	.06	-.03
Physical Effort	.00	-.02	-.01	.01	.05	.03	.02
Mental Effort	.02	-.06	-.02	.06	.04	.02	-.02

Figure 6: Regression results of all models of the four task attributes across three conditions: 1) within an activity (WITHIN), 2) preceded by an activity (PRECEDED + BETWEEN), and 3) succeeded by an activity (SUCCEEDED + BETWEEN). Note that the direction of effects and their significance levels are color coded: with green and purple representing positive and negative effects, respectively, and darker shades of those colors representing stronger effects, with the darkest shade of each color indicating p < .001.

they gave for considering these periods opportune was that, during them, they could not engage in other, longer unscheduled activities they were interested in, because their next scheduled activities were starting too soon. In other words, performing mobile crowdsourcing tasks was intended to kill time until the next activity began, or while physically moving towards it. *"Between when I had my lunch and when I got back to work, I leaned toward doing the tasks before starting my work"* (P16). Likewise, P06 said: *"When the bus would not be arriving for three to five minutes, I would do the tasks too. It's when you really just wanted to start 'phubbing', just do something to spend that time."*

Some interviewees confirmed that, as shown in the quantitative data, they performed crowdsourcing tasks in the midst of some other activity, and gave particular reasons for doing so. A common reason was feeling bored by the other activity, as P12 explained: *"I was in class, but the content was not appealing. So I did those missions."* Some participants reported that they felt "stuck" in certain situations, i.e., wanting to leave but feeling forced to stay until they ended. P28 likened these periods to *"being in a moving car [. . .] forced to sit there and not go anywhere."* Despite participants not actually performing any actions themselves at such times, they reported themselves as being in the prevailing activity in an ESM.

Apart from feelings of boredom, several participants reported that they had a tendency to polychronicity, i.e., preferring to multitask. *"I think work time is a good time for these [crowdsourcing tasks]. I'm supposed to do my work. But if you can do something extra at the same time, you feel that you've earned some time"* (P26). Interestingly, a few mentioned that they paused their primary tasks and started doing simple crowdsourcing tasks to *"shift their gears"*, i.e., refresh themselves when feeling tired by the current activity. As P02 put it, *"Like I was attending to the talk, but I was really sleepy. Then I wanted to do it to see if that helped me recover."*

The interviewees also offered some explanations of why they preferred not to perform crowdsourcing tasks while they were on long breaks. The main reason was that they wanted to rest, or engage in pursuits more clearly designated as leisure. As P17 stated, *"When I was not doing anything, I wanted to get some sleep or just veg out. When I was not going to do anything, there was a high chance that I'd just ignore it [the prompt]."*

6.2 Inopportune Moments for Crowdsourcing Tasks

The interviewees mentioned a variety of moments they found inopportune for performing crowdsourcing tasks and/or for receiving reminder prompts. Two general types of uninterruptible moments were commonly mentioned. The first was not being physically available, i.e., involved in an activity that could not be paused because it would need to be restarted from the beginning if a notification were dealt with. The second type was when they were concentrating on something too hard to be able to switch to crowdsourcing tasks. Nevertheless, some participants said it was fine to receive notifications at such times, because they *"could judge for themselves whether the current activity could be interrupted"* (P10). In addition, some noted that their reluctance to receive task prompts was because they were pondering something important (e.g., an exam or meeting with a professor) or were anxious about it. Receiving prompts in such circumstances made them feel annoyed. This resonated with our quantitative results reported earlier: that higher stress levels were correlated with a lower likelihood of task execution.

6.3 Reasons behind Task Choices

The interviewees mentioned a variety of factors that underlay their task choices, including how much time they had available, how tired they were, whether a given task required thinking, the perceived convenience of performing it, how many times they had done it before, whether they had one hand or both hands available, and whether they perceived the task was worth doing.

Specifically, they tended to perform tasks that required less mental effort and fewer actions when they were in a hurry. As P02 commented: *"If I were about to go to a place after a meal, which is a little bit rushed, I preferred tasks that were easier and that took only a few clicks. But if I just arrived at some place and I had a bunch of time, I would choose tasks that I rarely do."* When participants felt tired, they also tended to choose easier tasks, such as after workout/exercise or studying/working, as P15 noted: *"[I]t was time for relaxing. I don't want something complicated, but an easier one, like Image Labeling."*

However, the interviewees said they were more willing to choose a more difficult crowdsourcing task that took more thinking when they felt more energetic – *"I chose a different one when I wasn't tired, like the ones that are brain-draining"* (P10) – or when they perceived themselves to have more free time: *"I chose Sentiment Evaluation only when I had plenty of time. You have to read its text and comprehend it. Some of them are ambiguous, and it really takes you some time to think about it, not like the image label, which is more intuitive"* (P28). A few participants also mentioned their tendency to choose tasks that were similar to their current activity, out of a desire to maintain continuity. As P23 put it, *"If the thing I just did took mental effort, then I'd feel that I could also do some [crowdsourcing] tasks that take some thinking."* Several participants also performed tasks for specific purposes, such as leaving comments only when they wanted to praise or complain about a place: *"I left the comments if I was really angry about the shop"* (P24). A few tended to perform certain tasks that helped them improve their skills, such as P6, who attempted to improve his English through performing translation-related tasks. Finally, several participants mentioned their tendency to switch between tasks, mainly to keep up a sense of novelty: *"If I did a lot of image-label verification yesterday, I might get bored with it and change to another one today. I feel like I'd want to balance it somehow"* (P06).

6.4 Perceptions and Experiences of Task Reminders

Participants were generally pleased to see the task reminders issued by the research app, because these allowed them to enter the crowdsourcing apps directly. Several participants mentioned that our notifications appealed to their attention to click on them. For instance, *"I really rely on notifications, because without them, I need to find the app and perform the tasks. [...] Whenever I see the reminder on the notification drawer, I have an impulse to click it and do the tasks"* (P26). Nevertheless, reminder prompts that appeared when the users were actively using their phones for achieving specific, unrelated aims were generally unwelcome and considered annoying. As P17 explained, *"[t]exting is a continuous back-and-forth activity. Moments like these are not convenient for the tasks, because I'd feel disrupted."* Similarly, P15 stated, *"[w]hen I'm using my apps for doing something, like chatting with my friends or looking for restaurants, my goal was not doing the tasks, and I wouldn't want to do them."* And a few participants mentioned that they sometimes missed the prompts because they tended to dismiss all notifications at once.

7 DISCUSSION

7.1 "Free Time" for Mobile Crowdsourcing Tasks

Our participants commonly mentioned "free time" when reflecting on when they would choose to perform crowdsourcing tasks. What is particularly interesting to us is that, in saying this, they were mostly referring to transitions between activities rather than long spans of free time that were not perceived as directly preceded or followed by specific activities. These self-reported patterns were well supported by the quantitative results. Participants in transitional, as distinct from truly idle or unoccupied moments were more likely to proactively initiate the tasks; and such self-initiation was, in turn, correlated with longer task times and greater effort devoted. This was unexpected; we had assumed that when there was no activity at hand, individuals would be more willing to perform mobile crowdsourcing tasks. However, the participants consistently said that they regarded lengthy periods of free time as being for rest, and not "free time" of the sort they liked to use for tasks.

Their main stated reasons for considering transitions to be "free time" varied. Some cited personal preferences for making efficient use of their time. Others said that, since they could not have fit any lengthier activity into the sorts of brief time periods, they chose to perform mobile crowdsourcing tasks instead.

We were also somewhat surprised to find that being mid-activity was not necessarily considered a bad moment for task performance, even though full execution of tasks at such moments was less likely than it was at others. While some participants had a tendency to multitask and thus to feel they had "earned" time to do tasks, others felt bored by their other activity or felt "stuck" in it, and wanted to make time pass more quickly for those reasons, referred to as microwaiting in prior literature [17]. Interestingly, some participants treated mobile crowdsourcing tasks as refreshment during work. Taken together, these results suggest that "free time" for mobile crowdsourcing tasks, as conceived of by platform users, was not the same as periods of pure inactivity. Rather, it consisted of moments when these users felt they had to kill short amounts of time, either until their next activities commenced, or until their current dull or otherwise unwelcome activity ended. Prior research on opportune moments [42], breakpoints [1, 33, 34], and microwaiting moments [17] has deemed transitions to be good moments for handling notifications. Our findings not only support those results, but add that it is precisely such transitional periods – and not long, free-form breaks – that are most likely to be considered "free time", at least in the specific case of mobile crowdsourcing tasks.

7.2 The Influence of Activity Attributes and Personal

Conditions on Task Execution, Effort Devoted, and Task Choice Prior research has suggested that when smartphone users' engagement in their existing activity and the complexity of that activity are both high, they have lower receptivity to external interruptions [39]. Our research, despite using a different set of attributes – namely, physical effort, mental effort, attention, and complexity – to describe individuals' activity perceptions, resonates strongly with that prior body of work [19, 39]: one notable example being our finding that the attention required by an activity was negatively correlated with task execution.

On the other hand, while previous research indicates the negative effects of fatigue on the time it took to execute tasks, and task validity, out results were not only consistent with previous ones [14], but suggest different roles played by stress levels and energy levels on task performance. That is, while the former was influential on whether users would initiate a task in the first place, the latter had more bearing on how much time and effort they were willing to devote to a task session once it had already started. Moreover, stress that was caused specifically by an ongoing or preceding activity was found to influence users' task choices, whereas their energy level mattered to their task choice more if an activity was upcoming. In a similar vein, the different roles played by preceding and succeeding activities were also highlighted While preceding activities were more influential on task choices, succeeding ones were more influential on task execution.

Even more interesting were the positive correlations we found between the attributes of people's current activities and the attributes of the tasks they chose. That is, our participants were more likely to choose tasks whose attributes were similar to those of whatever else they happened to be doing. This phenomenon was also supported by our qualitative findings. On the whole, these findings suggest not only that activity context and personal condition do matter to task execution, effort devoted, and task choices, but also that there are specific relationships and correspondences between these factors that future research should pay close attention to. And our qualitative findings further suggest that even more factors were in play, that our ESM questionnaires did not capture. Those factors, too, would be worthwhile for future researchers to take into account.

8 DESIGN IMPLICATION

Our results have various design implications for future platforms or services seeking mobile crowdsourcing contributions via prompting mechanisms on smartphones. Our high-level recommendation is that such future entities, assuming that they send a fixed number of prompts per day to avoid overwhelming users, should schedule such prompts based on the recipients' calendar information, prioritizing short periods when they are more likely to feel they are "killing time", as opposed to lengthy periods without any scheduled activity, which they are likely to deem themselves "at rest". According to our results, the former, opportune moments are more likely to happen during transitions, especially between two activities. Prompts could usefully be sent when such a transition is due to start, or perhaps a little bit earlier. A sense of "killing time" is, however, also likely to arise in mid-activity if users become bored. Crowdsourcing services/platforms should be able to detect when their users are bored during an activity: for example, intensive phone usage might be a signal that a user has become disengaged from the non-crowdsourcing task at hand, and thus may feel s/he has free time for a crowdsourcing one. However, boredom detection is far from easy [24, 42], and sending a prompt whenever people are actively using their phones risks disrupting and annoying them. A sophisticated learning-based model that can distinguish between task-oriented active use and time-killing active use will therefore be needed. Yet, long before such technology is perfected, it will remain worthwhile to deliver crowdsourcing-task prompts. Our observations in this paper imply that the key to avoiding a sense of interruption and disruption on the part of platform users will be to provide them with a point of entry for self-initiation of tasks. We therefore recommend that task prompts appear silently, without generating any alert or pop-up. This is because, when users are motivated to self-initiate tasks – regardless of whether they merely want to kill time, or treat the task itself as a form of refreshment, or have one of the other motivations cited in prior studies [8] – they are also more likely to perform more tasks of a wider variety of types. The role of the reminder prompt should therefore be mainly to provide a "portal" into the crowdsourcing platform/service, rather than to hector them into performing particular tasks at particular times.

Managers of platforms/services that offer crowdsourcing tasks with differing attributes should also bear in mind that the same user is likely to prefer different task types at different times. Based on our results, we recommend that such platforms/services take account of the attributes of each users' existing and proximate activities, as well as their stress and energy levels, when determining which tasks to invite them to perform, according to the attributes of those tasks. Our Figures 4 and 5 provide, respectively, a reference for users' average ratings of four attributes of a variety of tasks, and the important attributes of current and temporally proximate activities that affect users' task choices. Crowdsourcing entities should also seek to understand their users' task choices using machine learning. Finally, we recommend that prompting mechanisms monitor the status of users' phones' notification centers. This is because, when a user has aggregated a large number of notifications, s/he is more likely to dismiss all of them [53]. Based on individual users' notification-checking and dismissal patterns, various prompting mechanisms can be developed that are tailored to individual differences.

9 LIMITATION

The current study is subject to several limitations. First, its design was inherently incapable of capturing the situations the participants were in when they failed or refused to answer ESM questionnaires. On the whole, it seems likely moments at which they would not or could not answer such questionnaires were also moments in which they would not have performed mobile crowdsourcing tasks: a known tradeoff of the ESM approach. For this reason, the task-execution likelihood was probably overestimated here.

Second, in our study, participants could not choose their preferred time range to receive task reminders according to their routine. In addition, as comparing task performance in different contexts made sending multiple reminders inevitable, this was likely to enhance our participants' awareness of mobile crowdsourcing tasks, resulting in that their task engagement might have been higher than when receiving only one task reminder per day.

Third, in the context of mobile crowdsourcing that did not provide any monetary benefit (e.g., Google Crowdsource, at the time of writing), our results might not be generalizable to them. It was likely that the active users of such platforms/services had other, strong motivations to offer crowdsourcing data, and were willing to perform tasks in a wider range of situations, than their paid counterparts. Our participants, despite their diversity in terms of how often they contributed (ranging from more than five times a day to just once a week), might not have been representative of the wider population of mobile crowdsourcing users [8].

Fourth, although there was no restriction on people's travel in Taiwan during the COVID-19 period, the pandemic situation might have lowered participants' mobility to some extent. This might have in turn lowered the frequency of participants receiving task prompts from the Local Guides platform.

Fifth, since we only asked our participants to rate task attributes once per task session, if there was more than one task in a session, their rating had to be taken as a summary of all the tasks within it. This could have caused bias when we analyzed both task attributes and users' task time and effort devoted. Nevertheless, perhaps because many of the crowdsourcing tasks our participants performed could be completed very quickly, and they tended to perform a series of tasks in each session, we observed that most tasks in any given session were identical (only 11% of all sessions contained two or more different types of tasks).

Sixth, in this study, participants received compensation based on the number of tasks completed. Incentives could influence motivation and likely increased the proactive execution rate of mobile crowdsourcing tasks; but from another perspective, the positive impact of providing incentives was in line with the expectations of mobile-crowdsourcing platforms.

Finally, we did not measure contribution quality, due to the voluminous amount of video data that was coded, and because it was not our focus. However, it might be worthwhile for future research to investigate the influence of context on the quality of users' data contributions to commercial crowdsourcing apps.

10 CONCLUSION

In this paper, we have primarily tried to understand how activity context and personal condition influenced smartphone users' task execution, effort devoted, and task choices when contributing to commercial mobile crowdsourcing services on their own smartphones in the wild. Our results indicate that the participants were more likely to proactively perform tasks during transitions between non-crowdsourcing activity than either during an ongoing activity or during long periods of idle time; and that these observed choices/actions closely matched their self-reported preferences. We also have shown that the measured attributes of the tasks they chose to perform were correlated with the attributes of the current and proximate activities, as well as with two personal conditions: stress level and energy level. More importantly, we have highlighted the differential impacts of preceding and succeeding activities, and of stress level and energy level, on task execution, task effort devoted, and task choice. Through interviews, we were also able to obtain insights into why participants chose to perform certain tasks at certain times. Our quantitative and qualitative findings have important design implications for future mobile crowdsourcing task-prompting mechanisms, and are themselves an important contribution to knowledge at the intersection of opportune-moments and mobile-crowdsourcing research.

ACKNOWLEDGMENTS

We thank Dr. Peggy Chi for her help with our study. We also thank the Ministry of Science and Technology, R.O.C. for their support (MOST 108-2218-E-009-050).

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 2004 conference on Human factors in computing systems (CHI '04)*. ACM Press, Vienna, Austria, 271–278. https://doi.org/10.1145/985692.985727

[2] Elena Agapie, Jaime Teevan, and Andrés Monroy-Hernández. 2015. Crowdsourcing in the field: A case study using local crowds for event reporting. In *Third AAAI Conference on Human Computation and Crowdsourcing (HCOMP '15)*.

[3] Brian P. Bailey and Joseph A. Konstan. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior* 22, 4 (July 2006), 685–708. https://doi.org/10.1016/j.chb.2005.12.009

[4] Francesco Calabrese, Massimo Colonna, Piero Lovisolo, Dario Parata, and Carlo Ratti. 2011. Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome. *IEEE Transactions on Intelligent Transportation Systems* 12, 1 (March 2011), 141–151. https://doi.org/10.1109/TITS.2010.2074196

[5] Yung-Ju Chang, Gaurav Paruthi, Hsin Ying Wu, Hsin Yu Lin, and Mark W. Newman. 2017. An investigation of using mobile and situated crowdsourcing to collect annotated travel activity data in real-word settings. *International Journal of Human Computer Studies* 102 (June 2017), 81–102. https://doi.org/10.1016/j.ijhcs.2016.11.001

[6] 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* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 6–15. https://doi.org/10.1145/2785830.2785852

[7] Yung-Ju Chang, Chu-Yuan Yang, Ying-Hsuan Kuo, Wen-Hao Cheng, Chun-Liang Yang, Fang-Yu Lin, I-Hui Yeh, Chih-Kuan Hsieh, Ching-Yu Hsieh, and Yu-Shuen Wang. 2019. Tourgether: Exploring Tourists' Real-Time Sharing of Experiences as a Means of Encouraging Point-of-Interest Exploration. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 128 (Dec. 2019), 25 pages. https://doi.org/10.1145/3369832

[8] Pei-Yu (Peggy) Chi, Anurag Batra, and Maxwell Hsu. 2018. Mobile crowdsourcing in the wild: challenges from a global community. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, 410–415. https://doi.org/10.1145/3236112.3236176

[9] Pei-Yu (Peggy) Chi, Matthew Long, Akshay Gaur, Abhimanyu Deora, Anurag Batra, and Daphne Luong. 2019. Crowdsourcing Images for Global Diversity. 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, 1–10. https://doi.org/10.1145/3338286.3347546

[10] 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

[11] Kapil Garg, Yongsung Kim, Darren Gergle, and Haoqi Zhang. 2019. 4X: A Hybrid Approach for Scaffolding Data Collection and Interest in Low-Effort Participatory Sensing. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 90:1–90:28. https://doi.org/10.1145/3359192

[12] Jorge Goncalves, Simo Hosio, Niels van Berkel, Furqan Ahmed, and Vassilis Kostakos. 2017. CrowdPickUp: Crowdsourcing Task Pickup in the Wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (Sept. 2017), 51:1–51:22. https://doi.org/10.1145/3130916

[13] Ching-Yu Hsieh, Yi-Shyuan Chiang, Hung-Yu Chiu, and Yung-Ju Chang. 2020. Bridging the Virtual and Real Worlds: A Preliminary Study of Messaging Notifications in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376228

[14] Kazushi Ikeda and Keiichiro Hoashi. 2017. Crowdsourcing GO: Effect of Worker Situation on Mobile Crowdsourcing Performance. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1142–1153. https://doi.org/10.1145/3025453.3025917

[15] Shamsi T. Iqbal and Brian P. Bailey. 2005. Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1489–1492. https://doi.org/10.1145/1056808.1056948

[16] Shamsi T. Iqbal and Brian P. Bailey. 2008. Effects of intelligent notification management on users and their tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. Association for Computing Machinery, New York, NY, USA, 93–102. https://doi.org/10.1145/1357054.1357070

[17] Ellen Isaacs, Nicholas Yee, Diane J Schiano, Nathan Good, Nicolas Ducheneaut, and Victoria Bellotti. 2009. Mobile Microwaiting Moments: The Role of Context in Receptivity to Content While on the Go. http://www.izix.com/pubs/Isaacs-InfoOnTheGo-2009.pdf

[18] Thivya Kandappu, Nikita Jaiman, Randy Tandriansyah, Archan Misra, Shih-Fen Cheng, Cen Chen, Hoong Chuin Lau, Deepthi Chander, and Koustuv Dasgupta. 2016. TASKer: behavioral insights via campus-based experimental mobile crowdsourcing. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 392–402. https://doi.org/10.1145/2971648.2971690

[19] Ioanna Katidioti, Jelmer P. Borst, Marieke K. van Vugt, and Niels A. Taatgen. 2016. Interrupt me: External interruptions are less disruptive than self-interruptions. *Computers in Human Behavior* 63 (Oct. 2016), 906–915. https://doi.org/10.1016/j.chb.2016.06.037

[20] Yongsung Kim, Darren Gergle, and Haoqi Zhang. 2018. Hit-or-Wait: Coordinating Opportunistic Low-effort Contributions to Achieve Global Outcomes in On-the-go Crowdsourcing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173670

[21] Stephen F. King and Paul Brown. 2007. Fix my street or else: using the internet to voice local public service concerns. In *Proceedings of the 1st international conference on Theory and practice of electronic governance (ICEGOV '07)*. Association for Computing Machinery, New York, NY, USA, 72–80. https://doi.org/10.1145/1328057.1328076

[22] Alisa Kongthon, Choochart Haruechaiyasak, Jaruwat Pailai, and Sarawoot Kongyoung. 2014. The Role of Social Media During a Natural Disaster: A Case Study of the 2011 Thai Flood. *International Journal of Innovation and Technology Management* 11, 03 (April 2014), 1440012. https://doi.org/10.1142/S0219877014400124

[23] Terry K. Koo and Mae Y. Li. 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* 15, 2 (June 2016), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

[24] Kostadin Kushlev, Bruno Cardoso, and Martin Pielot. 2017. Too tense for candy crush: affect influences user engagement with proactively suggested content. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, Vienna Austria, 1–6. https://doi.org/10.1145/3098279.3098569

[25] Florian Künzler, Varun Mishra, Jan-Niklas Kramer, David Kotz, Elgar Fleisch, and Tobias Kowatsch. 2019. Exploring the State-of-Receptivity for mHealth Interventions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (Dec. 2019), 140:1–140:27. https://doi.org/10.1145/3369805

[26] 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

[27] Nicolas Maisonneuve, Matthias Stevens, and Bartek Ochab. 2010. Participatory noise pollution monitoring using mobile phones. *Information Polity* 15, 1,2 (Jan. 2010), 51–71. https://doi.org/10.3233/IP-2010-0200

[28] Gloria Mark, Shamsi Iqbal, Mary Czerwinski, and Paul Johns. 2015. Focused, Aroused, but so Distractible: Temporal Perspectives on Multitasking and Communications. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM Press, Vancouver, BC, Canada, 903–916. https://doi.org/10.1145/2675133.2675221

[29] Daniel C. McFarlane. 1999. Coordinating the interruption of people in human–computer interaction. In *IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT '99)*. IOS Press, Edinburgh, UK, 295–303. Google-Books-ID: yXehjiOd_kkC.

[30] 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

[31] 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)*. ACM, New York, NY, USA, 1021–1032. https://doi.org/10.1145/2858036.2858566

[32] Tadashi Okoshi, Jin Nakazawa, and Hideyuki Tokuda. 2014. Attelia: Sensing User's Attention Status on Smart Phones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) *(UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 139–142. https://doi.org/10.1145/2638728.2638802

[33] Tadashi Okoshi, Julian Ramos, Hiroki Nozaki, Jin Nakazawa, Anind K. Dey, and Hideyuki Tokuda. 2015. Attelia: Reducing user's cognitive load due to interruptive notifications on smart phones. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom '15)*. IEEE, Los Alamitos, CA, USA, 96–104. https://doi.org/10.1109/PERCOM.2015.7146515 ISSN: null.

[34] 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* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 475–486. https://doi.org/10.1145/2750858.2807517

[35] Tadashi Okoshi, Kota Tsubouchi, Masaya Taji, Takanori Ichikawa, and Hideyuki Tokuda. 2017. Attention and engagement-awareness in the wild: A large-scale study with adaptive notifications. In *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 100–110. https://doi.org/10.1109/PERCOM.2017.7917856 ISSN: 2474-249X.

[36] Tadashi Okoshi, Kota Tsubouchi, and Hideyuki Tokuda. 2019. Real-World Product Deployment of Adaptive Push Notification Scheduling on Smartphones. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. ACM, Anchorage AK USA, 2792–2800. https://doi.org/10.1145/3292500.3330732

[37] 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)*. ACM, Portland Oregon USA, 541–554. https://doi.org/10.1145/2998181.2998189

[38] 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)*. ACM, New York, NY, USA, 897–908. https://doi.org/10.1145/2632048.2632062

[39] 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)*. ACM, New York, NY, USA, 1100–1105. https://doi.org/10.1145/2786567.2794336

[40] Maria Clara Pestana and Vaninha Vieira. 2018. Context-Aware Task Distribution for Mobile Crowdsourcing. In *Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems (IHC 2018)*. ACM, New York, NY, USA, 1–8. https://doi.org/10.1145/3274192.3274206

[41] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond Interruptibility: Predicting Opportune Moments to Engage Mobile Phone Users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (Sept. 2017), 1–25. https://doi.org/10.1145/3130956

[42] 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)*. ACM, Osaka, Japan, 825–836. https://doi.org/10.1145/2750858.2804252

[43] Benjamin Poppinga, Wilko Heuten, and Susanne Boll. 2014. Sensor-Based Identification of Opportune Moments for Triggering Notifications. *IEEE Pervasive Computing* 13, 1 (Jan. 2014), 22–29. https://doi.org/10.1109/MPRV.2014.15

[44] Kayela Robertson, Cody Rosasco, Kyle Feuz, Maureen Schmitter-Edgecombe, and Diane Cook. 2015. Prompting technologies: A comparison of time-based and context-aware transition-based prompting. *Technology and health care : official journal of the European Society for Engineering and Medicine* 23, 6 (2015), 745–756. https://doi.org/10.3233/THC-151033

[45] Manaswi Saha, Michael Saugstad, Hanuma Teja Maddali, Aileen Zeng, Ryan Holland, Steven Bower, Aditya Dash, Sage Chen, Anthony Li, Kotaro Hara, and Jon Froehlich. 2019. Project Sidewalk: A Web-based Crowdsourcing Tool for Collecting Sidewalk Accessibility Data At Scale. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300292

[46] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web (WWW '10)*. Association for Computing Machinery, New York, NY, USA, 851–860. https://doi.org/10.1145/1772690.1772777

[47] Hillol Sarker, Moushumi Sharmin, Amin Ahsan Ali, Md. Mahbubur Rahman, Rummana Bari, Syed Monowar Hossain, and Santosh Kumar. 2014. Assessing the Availability of Users to Engage in Just-in-time Intervention in the Natural Environment. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 909–920. https://doi.org/10.1145/2632048.2636082

[48] SeeClickFix. 2008. *SeeClickFix | 311 Request and Work Management Software*. https://seeclickfix.com

[49] Brian L. Sullivan, Christopher L. Wood, Marshall J. Iliff, Rick E. Bonney, Daniel Fink, and Steve Kelling. 2009. eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation* 142, 10 (Oct. 2009), 2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006

[50] 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

[51] 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

[52] Heli Väätäjä and Paul Egglestone. 2012. Briefing news reporting with mobile assignments: perceptions, needs and challenges. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 485–494. https://doi.org/10.1145/2145204.2145280

[53] 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

[54] Huimin Wen, Zhongwei Hu, Jifu Guo, Liyun Zhu, and Jianping Sun. 2008. Operational Analysis on Beijing Road Network during the Olympic Games. *Journal of Transportation Systems Engineering and Information Technology* 8, 6 (Dec. 2008), 32–37. https://doi.org/10.1016/S1570-6672(09)60003-9

[55] Tilo Westermann, Ina Wechsung, and Sebastian Möller. 2016. Smartphone Notifications in Context: a Case Study on Receptivity by the Example of an Advertising Service. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 2355–2361. https://doi.org/10.1145/2851581.2892383

[56] Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. 2018. Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. Association for Computing Machinery, New York, NY, USA, 585–596. https://doi.org/10.1145/3196709.3196730