# How Context Influences Cross-Device Task Acceptance in Crowd Work

**Danula Hettiachchi,[1] Senuri Wijenayake,[1] Simo Hosio,[2]**
**Vassilis Kostakos,[1] Jorge Goncalves[1]**
[1]The University of Melbourne, [2]University of Oulu
first.last@unimelb.edu.au, simo.hosio@oulu.fi

**Abstract**

Although crowd work is typically completed through desktop or laptop computers by workers at their home, literature has shown that crowdsourcing is feasible through a wide array of computing devices, including smartphones and digital voice assistants. An integrated crowdsourcing platform that operates across multiple devices could provide greater flexibility to workers, but there is little understanding of crowd workers' perceptions on uptaking crowd tasks across multiple contexts through such devices. Using a crowdsourcing survey task, we investigate workers' willingness to accept different types of crowd tasks presented on three device types in different scenarios of varying location, time and social context. Through analysis of over 25,000 responses received from 329 crowd workers on Amazon Mechanical Turk, we show that when tasks are presented on different devices, the task acceptance rate is 80.5% on personal computers, 77.3% on smartphones and 70.7% on digital voice assistants. Our results also show how different contextual factors such as location, social context and time influence workers decision to accept a task on a given device. Our findings provide important insights towards the development of effective task assignment mechanisms for cross-device crowd platforms.

## Introduction

Information workers have used stationary desktop or laptop computers as their primary work tool for decades. A similar trend can be seen in crowd work, with workers typically completing tasks from home and mainly using a desktop workstation (Williams et al. 2019). However, with the advancements in wireless internet technologies and widespread availability of more sophisticated and powerful mobile computing devices (e.g., digital voice assistants, smartphones), digital workers now have more flexibility than ever before to work in different contexts.

Research has shown that crowdsourcing is increasingly conducted via non-traditional devices, such as voice-interaction through smartphones (Vashistha, Sethi, and Anderson 2017), digital voice assistants including smart speakers (Hettiachchi et al. 2020a), situated touch-screen displays (Hosio et al. 2014; Goncalves et al. 2013), as well as low-cost phones (Vashistha, Garg, and Anderson 2019). Given the wide range of crowdsourcing interfaces, workers have the flexibility to complete crowd tasks in a variety of different contexts (Hettiachchi et al. 2020a). These platforms engender additional benefits, such as improved accessibility of crowdsourcing marketplaces for workers with visual impairments (Vashistha, Sethi, and Anderson 2018) (e.g., via voice interaction) and low-income (Vashistha, Garg, and Anderson 2019) (e.g., via the use of low-cost phones or situated touch-screens) to engage in crowd work.

Although crowd work is feasible through many devices, current commercial platforms are primarily built for desktop/laptop access. An integrated crowdsourcing platform that is accessible via different devices, like smartphones and digital voice assistants, has potential for offering various benefits to workers. However, it remains unclear whether – given a choice – crowd workers would be willing to complete different types of tasks on devices other than desktop or laptop computers, particularly when considering different contexts.

Thus, in this study we aim to better understand how workers decide which type of device to use, and particularly how context affects this decision. Through a Human Intelligent Task (HIT) deployed in Amazon Mechanical Turk (MTurk)[1], we collected 25,920 responses from 329 unique crowd workers. Our results indicate that task parameters (*e.g.,* HIT time estimation, available HIT count) and contextual factors (*e.g.,* approximate location, social context) play an important role on workers' decisions to accept or reject tasks. Our findings inform the creation of integrated crowdsourcing platforms and effective cross-device task assignment mechanisms that can increase overall crowd data quality and worker satisfaction.

## Related Work

### Impact of Worker Context

Data quality in crowdsourcing is an important research avenue that has been critical to the widespread adoption of crowdsourcing in both academic and commercial applications. While there are many different data quality enhance-



[1]https://www.mturk.com

ment techniques, the majority of them are centred around matching tasks with workers, improving task design and workflow, or aggregating answers from the crowd (Daniel et al. 2018; Kittur et al. 2013).

Previous work has proposed many different task matching or assignment strategies that capitalise on different factors, such as worker characteristics (e.g., personality (Kazai, Kamps, and Milic-Frayling 2012), skills (Mavridis, Gross-Amblard, and Miklós 2016), cognitive ability (Hettiachchi et al. 2019; Goncalves et al. 2017; Hettiachchi et al. 2020b)) and behavioural traces (Gadiraju et al. 2019; Goyal et al. 2018). However, there is far less research investigating the impact of contextual factors related to the crowd worker's environment. Such contextual factors are of particular importance when the goal is to achieve task assignment or recommendation in a crowdsourcing platform that can be accessed through different types of devices.

For example, Ikeda and Hoashi (2017) show that worker busyness and presence of a companion can impact task acceptance in mobile crowdsourcing. On a related note, as tasks in spatial crowdsourcing are directly related to a specific location, they are typically accessed through smartphones and contextual information plays an important role in task assignment (Gummidi, Xie, and Pedersen 2019). Similarly, contextual factors such as worker location, device sensing capabilities and battery level are critical in crowd sensing applications (Hassani, Haghighi, and Jayaraman 2015).

## Devices for Crowd Work

Several recent studies have explored the characteristics of worker devices and their impact on task performance. Gadiraju et al. (2017) investigated the effect of the work environment on micro-task performance in CrowdFlower. The study which involves workers from the US and India shows that factors like screen resolution and device speed can have an impact on the task completion time. In a study investigating the work-life of crowd workers of MTurk, Williams et al. (2019) report that the number of monitors of the primary work computer is the most productivity defining attribute related to the workspace.

Although micro-task crowdsourcing has been traditionally limited to web interfaces accessed via desktop/laptop computers, crowd workers increasingly use smartphones to complete tasks (Chi, Batra, and Hsu 2018; Chatzimilioudis et al. 2012). Also, recent work has shown the possibility of using a wide variety of devices for crowdsourcing. Crowd work is possible through digital voice assistants through smart speakers (Hettiachchi et al. 2020a), basic phones (Vashistha, Garg, and Anderson 2019), situated touch-screen displays (Hosio et al. 2014; Goncalves et al. 2013) as well as wearable devices like smartwatches (Acer et al. 2019). Hettiachchi et al. (2020a) present a voice-based crowdsourcing platform that works through a digital voice assistant. Results of their lab study show that task accuracy for native English speakers in voice-interaction is similar to the screen-interfaces across five different common crowdsourcing tasks. Vashistha, Garg, and Anderson (2019) use interactive-voice-response (IVR) in basic phones to crowdsource speech transcription tasks. Their application is targeted at economically disadvantaged crowd workers and provides means to engage in crowd work with minimum resources.

While connected crowd platforms that can operate through many devices can be beneficial to crowd workers, there is no work that sheds light on worker perceptions of when tasks are presented and possible to complete on multiple devices.

## Task Search and Acceptance

Crowdsourcing marketplaces typically expose a list of crowd tasks to workers from which they have to choose and accept to work on. While the aim is to provide greater autonomy and agency to workers, searching for a suitable task has become increasingly difficult and time consuming for workers (Chilton et al. 2010). Also, searching for optimal tasks is perceived as unpaid work for crowd workers (Hara et al. 2018).

There are many tools that can help workers find suitable tasks (Kaplan et al. 2018; Williams et al. 2019). For example, Turkopticon is one of the most widely adopted browser extensions that is used to evaluate and review requesters and HITs (Irani and Silberman 2013). Similar tools have been proposed to estimate the time that is needed to complete a task (Saito et al. 2019). However, such tools are limited to web-interfaces and are not always available in other devices, such as smartphones or smart speakers.

On the other hand, task search times can be much longer when interacting with devices such as smartphones and smart speakers when compared to desktop or laptop computers (Hettiachchi et al. 2020a). In smartphones, the amount of information that a worker can obtain at any given time is limited in smartphones due to the screen size. Similarly, voice-interaction limits the amount of information presented on smart speakers (Hettiachchi et al. 2020a).

Therefore, appropriate task assignment and recommendation is quite important for a cross-device crowd platform, especially when workers request tasks through smartphones or smart speakers. In this study, we take the initial steps to understand cross-device task acceptance, which is essential to create an effective task assignment model that can increase the overall data quality and worker satisfaction.

# Study

Our study consists of two main components deployed on MTurk. First, we deployed the main task, where workers marked their stated preference in accepting tasks. Second, we invited workers to complete two different surveys, depending on the number of completed HITs.

## Main Task

To understand workers' preferences in accepting tasks on different devices in various scenarios, we constructed a simple task. As shown in Figure 1, in each HIT, we presented workers with a list of parameters related to a hypothetical task (HIT). These parameters include task characteristics, such as Task Name, number of HITs available as well as contextual parameters, such as workers' approximate location, device and time of the day. Workers were asked to carefully examine the parameters and decide if they would accept and start working on the task. We clarified that they would not be asked to actually complete the given hypothetical task. In addition to the binary response of either accepting or rejecting the given task, workers were asked to indicate through a series of range sliders the extent to which certain factors influenced their decision. Following the design guidelines proposed in the literature, the range sliders had no tick marks in the axis and dynamically displayed the value to users as they move the marker (Matejka et al. 2016; Hosio et al. 2018). As shown in Figure 1, we listed five factors: Location, Device, Time, Social Context and Task details.

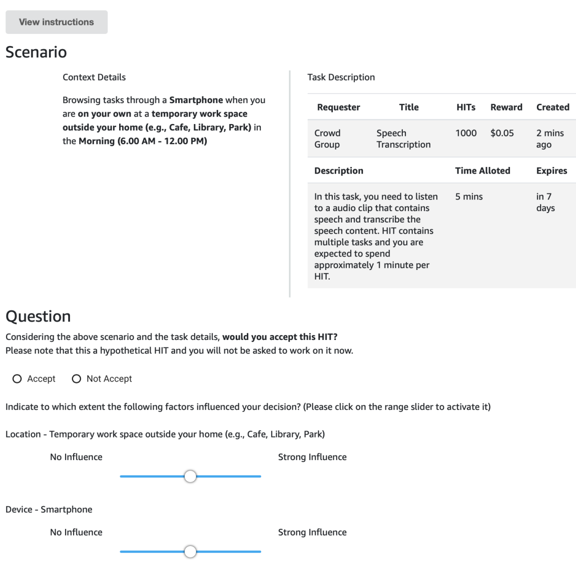


Figure 1: A portion of the HIT which shows the presentation of task parameters and the questions.

Tasks were selected based on typical tasks that are available on crowdsourcing platforms (Difallah et al. 2015) and task taxonomies purposed in the literature (Gadiraju, Kawase, and Dietze 2014). We also ensured that there is an equal number of text-based, audio-based and image-based tasks. As workers consider reward a key factor for accepting tasks (Hara et al. 2018), we kept the reward proportional to the expected time to complete the task. We provided workers with the maximum task time through the 'Time Allotted' parameter. A more realistic estimate of the actual time to complete the HIT was given in the task description. HIT count was set at 1, 10 and 1000 based on common values prevalent in typical marketplaces (Difallah et al. 2015). Requester name, HIT created time, and HIT expiration time were consistent across all HITs.

We created 5,184 HITs by using all possible combinations of the parameters listed in Table 1 and collected five responses per HIT. We set out three pre-qualification requirements for workers. Based on the qualifications, all our workers were from the US and have completed more than 1000 tasks with an approval rate of 95% or higher.

Table 1: Task and Context parameters

| **Parameter** | **Values** |
|---|---|
| Task Type | Sentiment Analysis<br>Information Finding<br>Audio Tagging<br>Speech Transcription<br>Image Classification<br>Bounding Box |
| HITs | 1, 10 or 1000 |
| Reward | $ 0.01,$ 0.05, or $ 0.50 |
| Created | 2 mins ago |
| Time Allotted | 1 min, 5 mins, or 10 mins |
| Expires | in 7 days |
| Time of the day | Morning (6.00 AM - 12.00 PM)<br>Afternoon (12.00 PM - 6.00 PM)<br>Evening (6.00 PM - 12.00 AM)<br>Night (12.00 AM - 6.00 AM) |
| Approximate Location | at home (at your primary workstation)<br>at home (other space different to your primary workstation)<br>at a temporary work space (e.g., Cafe, Library, Park)<br>commuting (e.g, in a Car or Train) |
| Social Context | on your own<br>with family/friends |
| Device | Desktop/Laptop<br>Smartphone<br>Smart speaker |

## Surveys

All workers who completed at least one HIT in our main task were asked to complete a demographic survey. The survey included questions about workers' preferred time to conduct crowd work, the average time they spend on crowd work, crowd work income (as a percentage of total income) and whether they have used voice assistants in general. We also captured basic demographic information such as age, gender, and primary internet device.

Furthermore, we invited workers who completed more than 20 HITs in our main task to complete an additional follow-up survey. We asked workers to provide detailed answers with examples of how different task characteristics and contextual factors impact their task acceptance based on previous crowd work experience. We also queried which factors they would consider if crowdsourcing platforms are available through multiple devices. We further inquired on their preference for task assignment and task recommendation on standard crowd market places as well as on different devices. Workers received USD $1.00 for the completion of this survey.

| Parameter | Estimate | Std. Error | Z value | |
|---|---|---|---|---|
| Intercept | 2.00 | 0.20 | 9.92 | *** |
| Task - Information Finding | −0.33 | 0.07 | −4.75 | *** |
| Task - Audio Tagging | −0.54 | 0.07 | −7.86 | *** |
| Task - Speech Transcription | −0.73 | 0.07 | −10.58 | *** |
| Task - Image Classification | 0.06 | 0.07 | 0.91 | |
| Task - Image Bounding box | −0.58 | 0.07 | −8.45 | *** |
| Time Allotted | 0.04 | 0.01 | 8.11 | *** |
| HIT Count | 0.04 | 0.02 | 2.63 | ** |
| Device - Smartphone | −0.86 | 0.14 | −6.04 | *** |
| Device - Smart speaker | −1.46 | 0.14 | −10.32 | *** |
| Time - Afternoon (12.00 PM - 6.00 PM) | −0.09 | 0.10 | −0.88 | |
| Time - Evening (6.00 PM - 12.00AM) | −0.08 | 0.10 | −0.73 | |
| Time - Night (12.00 AM - 6.00 AM) | −0.48 | 0.10 | −4.81 | *** |
| Social Context - with your family/friends | −0.73 | 0.07 | −10.41 | *** |
| Location - at home (other space different to your primary workstation) | −0.17 | 0.10 | −1.67 | |
| Location - at a temporary work space (e.g., Cafe, Library, Park) | −0.64 | 0.10 | −6.25 | *** |
| Location - commuting (e.g, in a Car or Train) | −0.99 | 0.10 | −9.92 | *** |

Table 2: Fixed effects of Generalised Linear Mixed Model. Significance '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$

| Parameter | Estimate | Std. Error | Z value | |
|---|---|---|---|---|
| Device - Smartphone: Time - Afternoon (12.00 PM - 6.00 PM) | 0.13 | 0.14 | 0.94 | |
| Device - Smart speaker: Time - Afternoon (12.00 PM - 6.00 PM) | 0.07 | 0.14 | 0.55 | |
| Device - Smartphone: Time - Evening (6.00 PM - 12.00AM) | 0.17 | 0.14 | 1.21 | |
| Device - Smart speaker: Time - Evening (6.00 PM - 12.00AM) | −0.08 | 0.14 | −0.55 | |
| Device - Smartphone: Time - Night (12.00 AM - 6.00 AM) | 0.06 | 0.14 | 0.46 | |
| Device - Smart speaker: Time - Night (12.00 AM - 6.00 AM) | 0.13 | 0.14 | 0.96 | |
| Device - Smartphone: Social Context - with your family/friends | 0.41 | 0.10 | 4.27 | *** |
| Device - Smart speaker: Social Context - with your family/friends | 0.46 | 0.10 | 4.81 | *** |
| Device - Smartphone: Location - at home (other space) | 0.18 | 0.14 | 1.28 | |
| Device - Smart speaker: Location - at home (other space) | 0.28 | 0.14 | 2.02 | * |
| Device - Smartphone: Location - at a temporary work space | 0.44 | 0.14 | 3.13 | ** |
| Device - Smart speaker: Location - at a temporary work space | 0.16 | 0.14 | 1.14 | |
| Device - Smartphone: Location - commuting (e.g, in a Car or Train) | 0.40 | 0.14 | 2.91 | ** |
| Device - Smart speaker: Location - commuting (e.g, in a Car or Train) | 0.44 | 0.14 | 3.23 | ** |

Table 3: Interactions of Generalised Linear Mixed Model

## Results

We collected 25,920 responses for our main task with a total of 329 workers contributing to the task. Each worker completed 78.8 tasks on average and spent 51.3 seconds on each single response on average.

### Worker Demographics

60 (18.2%) out of 329 workers completed the demographics survey. The number of answers provided by this subset of workers accounts for 22.0% of the total responses in the main task.

We present an estimation of the worker demographics based on the collected survey responses. Based on self-reported gender, 33 women and 27 men answered the survey with an average age of 38.6 (SD = 10.9) years. Workers reported spending an average of 22.3 hours per week on the Mechanical Turk platform with a majority (86.7%) of them working on Mechanical Turk during both weekdays and weekends. Workers stated earning on average 41.3% of their monthly income from crowd work. Furthermore, 15 workers stated that 90% or more of their monthly income comes from crowd work.

The majority of workers (98.3%) reported that they use a desktop computer or a laptop computer as their primary internet device to complete crowd tasks. Only one worker stated that they use an iPad as their primary device for crowd work. 58.8% of the workers reported to have previously used the mobile version of Mechanical Turk, whereas 62.7% of the workers have used a digital voice assistant. When inquired about the locations from where they complete crowd work, workers mainly mentioned workstation at home, bedroom at home, and living room at home as their primary work location.

### Task Acceptance

In response to our primary question given in the main task, workers decided to accept the given HIT in 19,759 (76.2%) of the cases. To investigate the impact of task and contextual parameters on task acceptance, we fitted a binomial generalised linear mixed model with maximum likelihood (Laplace Approximation) using the R-package lme4 (Bates et al. 2015). We included all the parameters listed in Table 1 and interactions between the device and contextual factors (Time, Location and Social Context). Worker ID, which is unique for each worker, was included as a random effect. Our results indicate significant fixed effects both in terms of the task parameters and contextual factors, and are detailed in the Table 2.

## Impact of Task Parameters

The results indicate strong fixed effects in terms of the task type, time estimation and number of HITs available. In Figure 2, we observe that workers prefer tasks that have an estimated completion time of 1 minute as opposed to very short (10 second) or long (10 minute) HITs. This preference is evident across all devices. However, we note that workers are more reluctant to accept long (10 minute) HITs in smart speakers when compared to other devices.

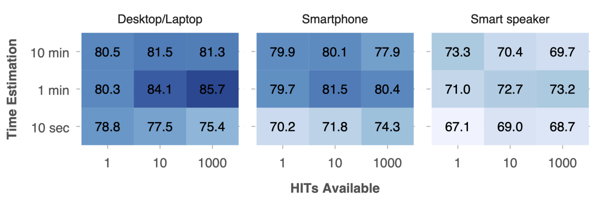


Figure 2: Task Acceptance rate on different devices across varying time estimations and number of HITs available.

As shown in Figure 3, the task acceptance rate also varied by task type, but did not exhibit major variations across devices.

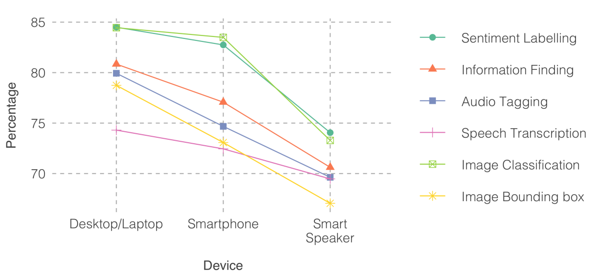


Figure 3: Task Acceptance rate on different devices across task types.

## Impact of Contextual Factors

Our results suggest that approximate location, social context, and device, influenced workers' decision to either accept or reject a given task.

**Approximate Location** When considering tasks presented on Desktop or Laptop computers, results indicate the highest acceptance rate at their primary workstation at home. However, when we examine task acceptance rates on smartphones and smart speakers, the acceptance rate is higher when the workers are at a space within their home different to their primary workstation as compared to the rest of the locations.

**Social Context** With regard to social context, workers are more likely to accept a task when they are on their own (78.3%) as compared to a situation where they are accompanied by family or friends (74.1%). As seen in Figure 4 (middle), this effect is consistent across devices.

**Time** Time of the day did not have a significant impact on workers choice except that workers preferred Morning (78.0%), Afternoon (76.5%) and Evening (77.4%) when compared to Night (73.0%). Similarly, as shown in Figure 4 (bottom), we did not find any meaningful interaction effect between Time and the Type of Device.

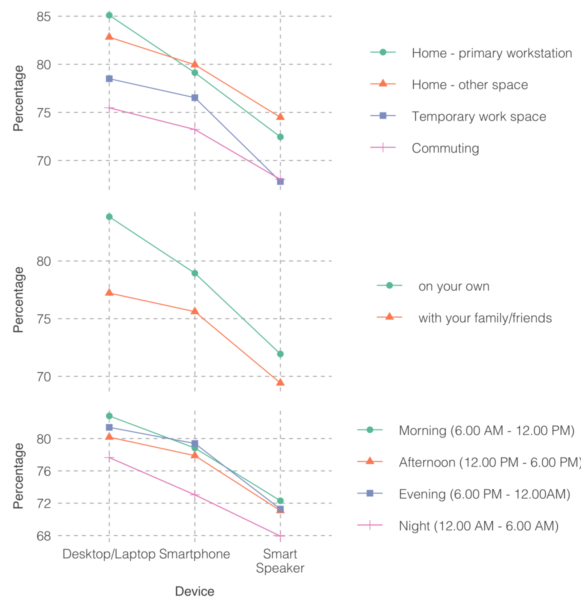


Figure 4: Task acceptance rate across approximate locations (top), social contexts (middle), and time intervals (bottom) on different devices.

**Worker Responses on Contextual Factors** In addition to the binary decision to accept the task, we asked workers to indicate which factors influenced their decision. Figure 5 summarises worker responses. From the response mean values and distributions, we observe that all task and contextual factors influenced the decision when accepting or rejecting given tasks. Task parameters were identified as being slightly more important than contextual factors when rejecting tasks. Also, response distributions (bi-modal distributions in Not Accept and normal distributions in Accept) indicate that workers were more decisive on factors when they did not accept tasks when compared to the cases where they accepted tasks.

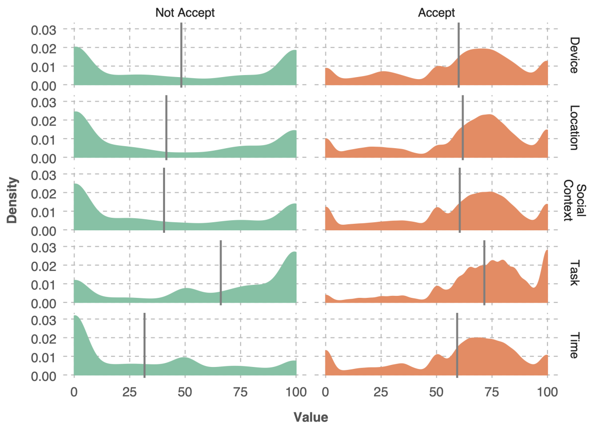


Figure 5: Reported importance of different parameters when deciding whether to accept the given task. The vertical line indicates the mean in each group.

## Follow-up Survey

From 94 invited workers who completed more than 20 HITs in the main task, we received a total of 30 responses to the follow-up survey. Two of the paper's authors individually applied a deductive thematic analysis (Braun and Clarke 2006) to the eligible responses based on the paper's research objectives. Following this, the authors met to discuss their outcomes. Next, we present the main findings of this analysis.

**Impact of Contextual Factors on Task Acceptance** We set out to investigate how contextual factors such as the social context, approximate location, device type, and time of the day impact task acceptance among workers.

First, the social context of workers emerged as a crucial determinant of task acceptance. Participants highlighted how they prefer to work on tasks when they are alone and can adequately concentrate on the task, and would not accept tasks that are audio-based or require higher concentration, when with family or friends;

> "*If I'm alone I will attempt just about any task. However if there are people around, I typically tend to stick to less involved tasks that don't require much concentration, especially those tasks that require listening to audio because it often becomes hard to hear the audio.*" (P8)

Moreover, participants were most likely to work on HITs from their primary workstation at home due to its comfortable and stable setup. However, some participants also note how their location may influence their device preference and the type of tasks they would attempt;

> "*When I have long waits e.g., at doctor's office I will do quick surveys or batch HITS on my phone. So I guess where I am determines the device.*" (P2)

The majority of our participants stressed their preference to complete tasks using a desktop or a laptop computer, over other devices (*e.g., smartphones*), as they offer larger displays and other controls (*e.g., keyboard and mouse*) requiring lesser effort to complete tasks;

> "*I would not do a task if it was not offered on a laptop. The laptop is the best device to use because of the decent size screen and the use of the keyboard and mouse. With a laptop, I have easy access with the click of a mouse and I can use my keyboard to complete some tasks. All the other devices would tire me out faster.*" (P10)

Moreover, if the task was available across multiple devices, participants would consider the compatibility of the device that they are using at the time and the task at hand, when deciding whether to accept a task or not;

> "*[Task acceptance] would depend on what device I'm using currently and how easy or difficult it would be complete that job on that device or if it would be better to switch. Some jobs require a larger screen so if I'm using my phone (rarely) I would want to do that job on a tablet or my Chromebook for example.*" (P7)

**Impact of Task Characteristics on Task Acceptance** We also investigated the impact of task characteristics such as HIT count, reward and requester profile on how workers determine whether or not to accept a task.

We note a significant preference among workers for tasks with a substantial HIT count (in thousands), each HIT requiring a small time period to complete;

> "*I love to complete hits with large hit counts that are fast and easy to complete, allowing me to sit and focus on them for extended periods of time.*" (P1)

Participants explained that completing simple and repetitive tasks allow them to stay focused for a long period of time at once, thereby maximising their earnings per hour;

> "*I love to complete hits with large hit counts that are fast and easy to complete, allowing me to sit and focus on them for extended periods of time.*" (P1)

Furthermore, participants also emphasised how the reward allocated for a task could impact task acceptance. In general, participants were keen on maintaining an acceptable hourly earning and therefore would calculate reward/estimated task time prior to accepting a task.

Moreover, despite general reluctance from participants to accept tasks requiring a device other than a desktop or a laptop, we note that a substantial reward could encourage them to do so;

> "*I love to complete hits with large hit counts that are fast and easy to complete, allowing me to sit and focus on them for extended periods of time.*" (P1)

Additionally, participants were seen to consider the requester profile - especially their approval rate and average pay time - when considering tasks for acceptance;

> "*I steer clear of requesters whose approval percentage is below 90 or whose average pay time is more than 3-4 days. These are signals that my HITs probably won't be approved or paid out, which negatively impacts my worker profile.*" (P8)

Participants also explained how they tend to "only do a few HITs from a requester that has an approval rate under 95% to see if they are approved first" (P12) as a precaution in such cases.

**Preference for Task Recommendation and Assignment** We note both supportive and critical opinions from participants regarding task recommendation and assignment in standard crowd market places. Participants supportive of this notion explained how task recommendations could connect each worker with tasks compatible with their personal skill set, reducing the amount of time they otherwise spend on searching compatible tasks;

> "*It would be nice just to have that option. Viewing tasks that are compatible with me would streamline the amount of tasks I complete because I spend a lot of time searching for HITs to complete.*" (P10)

They also suggested that task recommendations could be based on workers' ratings on completed tasks in addition to tasks they have completed successfully so that recommendations would include a mix of tasks they enjoy as well as tasks they are competent at;

> "*I do think it's nice to be able to sent certain tasks if there was a way to be sure that the recommendation system worked off of something like ratings from the users as opposed to just the history of tasks worked. If it were truly able to work that way then it would make it much easier to be able to jump on a task I enjoy and not have it taken by someone else who may or may not enjoy doing it so that would be a bonus for both the requester and the worker.*" (P14)

Moreover, participants highlighted how task recommendations could be especially important when tasks are offered across multiple devices. For instance, they note how recommending tasks that are compatible with the device currently in use could be valuable;

> "*I would prefer task assignment. I like doing tasks that are compatible to the device at use. There's no point in trying to complete a task that isn't presented on your device in a way that makes it easy for you to complete it.*" (P10)

P3 also commented on how device-based task recommendations could "ensure the task is administered in the most efficient manner, using the most compatible device", resulting in higher quality responses. Participants also emphasised that the opportunity to specify devices that they prefer not to use could "help the right workers get the right HITs and stop so many of the HITs from being picked up and returned" (P14), which would also prevent workers getting frustrated due to device – task incompatibilities.

Alternatively, some participants were critical of task assignment and recommendation as they were sceptic of the platform's ability to account for their personal preferences and other contextual factors;

> "*I definitely would not think [task recommendation] is helpful and in fact it would annoy me. I like to search and scroll through the tasks so I can evaluate them on my own judgement. A platform is simply an AI and it doesn't know any of my other factors, like how much time I have left in a day to complete HITs, what kind of HITs I want to complete today, and what my financial quota is for those HITs.*" (P5)

Some others also emphasised that while task recommendation could assist them find work faster, task assignment would be detrimental to their sense of agency;

> "*I would not like to be assigned work because the whole point of doing MTurk, for me, is functionally being my own boss.*" (P7)

Participants were also concerned that task recommendation may limit them to only certain types of HITs (based on their working history), restricting their opportunities to attempt new and interesting tasks in future.

## Discussion

### Crowd Work Devices

Our qualitative results indicate that crowd workers prefer workstations with desktop or laptop computers, mainly due to usability factors, such as large screen area and familiarity with keyboard-mouse setup. This preference is also evident in Figure 4 (top) through the high rate (85.1%) of task acceptance in desktop/laptop devices when workers are at their primary workstation. This is also in line with the findings of literature that investigate crowd worker preferences (Williams et al. 2019).

We also note that workers were willing to accept 77.3% of the given tasks on smartphones and 70.7% of the tasks on smart speakers as compared to 80.5% on desktop/laptop computers. This receptiveness towards alternative devices in the proposed scenarios shows promising signs with regard to the feasibility of cross-device crowd platforms that involve voice-interaction (Vashistha, Sethi, and Anderson 2017; Hettiachchi et al. 2020a).

Absence of different work tools (*e.g.,* browser extensions that filter tasks (Kaplan et al. 2018)) can make other devices less desirable for crowd workers. Similarly, as observed in our results, complex image related tasks such as image bounding box are less desired on smartphones and smart speakers, due to limited screen-size and restricted interaction options. This is also evident in our results, where workers stated that they are concerned about how easy it would be to complete the task on a device of interest. Therefore, when making tasks available through different devices, it is important to validate if the interaction style (*i.e.,* touch interaction in smartphones and voice interaction in smart speakers) and device capabilities are compatible with the task.

### Worker Context and Task Acceptance

In this work, we explore how contextual factors impact task acceptance in a cross-device scenario. Approximate location and social context appear to be particularly important for workers. When closely examining the task acceptance rates (Figure 4 (top)), for both smartphones and smart speakers, the acceptance rate is higher when at other spaces at home than when at the primary workstation. While extracting the specific worker location is not recommended as it leads to privacy concerns (To, Ghinita, and Shahabi 2014), we show that approximate location is a reasonable alternative that influences task selection on a cross-device platform. We also observe that time of the day is generally not a primary concern for workers except that, unsurprisingly, the task acceptance rate is much lower during the night (12.00AM - 6.00AM).

While mobile crowd work is common (Chi, Batra, and Hsu 2018), an estimated over 40% of our workers have never used the mobile version of MTurk platform. On the other hand, voice-based crowd platforms are not yet commercially available (Hettiachchi et al. 2020a). This limited understanding and exposure to voice-based and other alternative crowdsourcing platforms can be a reason behind less pronounced interaction effects concerning contextual factors and devices.

Our results also indicate that parameters specific to the HIT such as task type, the number of HITs available, and task time estimation, still play a vital role in task selection. Our findings are in line with the crowdsourcing literature (Daniel et al. 2018; Martin et al. 2014) and further confirms that such relationships extend into various crowd work scenarios.

### Integrated Cross-Device Crowdsourcing Platforms

Audio related tasks, like audio annotation and speech transcription, are common in current crowdsourcing marketplaces (Difallah et al. 2015) which has led to an increased exploration of crowdsourcing via voice-interaction (Vashistha, Sethi, and Anderson 2017; Hettiachchi et al. 2020a). On the other hand, smartphones and other mobile computing devices are capable of handling performance intensive tasks and are suitable for sustained work (Chi, Batra, and Hsu 2018).

Crowdsourcing platforms have seen an increase in the number of tasks related to mobile apps. Research has also shown that there is potential to use crowdsourcing for tasks that extend beyond screen-based devices, such as virtual reality experiments or application testing (Ma et al. 2018). Some platforms, such as *Prolific*, even allow mobile app installs as part of the assigned tasks. However, our qualitative results highlight that crowd workers find it inconvenient to switch between devices to complete a task. By allowing workers to browse tasks, accept and work on different devices, a cross-platform crowd marketplace can mitigate the required effort to switch between devices and create a positive crowd work experience for workers.

Literature also reports that workers exhibit multi-tasking behaviour and engage in other tasks like watching TV and chatting online while completing crowd work (Chandler, Mueller, and Paolacci 2014). In fact, some workers prefer to multi-task even though it is not always desired by task requesters (Lascau et al. 2019). Working on devices like smartphones and smart speakers could easily allow workers to facilitate this multi-tasking work style as compared to a workstation. In addition, an always-on device like the smart speaker or a ubiquitous device like the smartphone is helpful for workers in terms of handling interruptions and working in short sessions.

Given the steady growth in crowd work population (Difallah, Filatova, and Ipeirotis 2018) and the availability of a wide array of tasks, we anticipate that crowdsourcing platforms will gradually shift towards natively supporting different types of devices. For example, the popular crowdsourcing platform Amazon MTurk is aiming to increase task compatibility on mobile devices [2] and is well-positioned to extend their platform to smart speakers in the future through the increasingly ubiquitous Amazon Alexa.

### Limitations

We acknowledge several limitations of our study. First, workers who participated in our study have not experienced a fully functional voice-based crowd platform. It is possible that this lack of exposure impacted their decision to either accept or reject tasks on smart speakers. Second, our qualitative data originates from a subset of workers who took part in the main task. We invited 94 workers for the post-survey through a custom qualification in MTurk from which 30 workers completed the task. Third, we do not investigate all possible contextual factors and focus primarily on ones that have been shown to impact crowd work. Nevertheless, we tested over 5,000 unique HITs in our study, which provides a wide array of potential crowd work scenarios. Additional factors would vastly increase this number and lead to an overly complex study design.

## Conclusion

In this paper, we present a study on MTurk aimed at better understanding crowd workers' preferences regarding accepting or rejecting tasks under varying contexts. Our results indicate that task acceptance rate is 80.5% on personal computers, 77.3% on smartphones and 70.7% on digital voice assistants. We also show that contextual factors such as workers approximate location and social context influence their willingness to accept tasks presented on different devices. Further, we discuss how an integrated crowdsourcing platform that operates across different types of devices can bring benefits to crowd workers by allowing for flexibility in terms of work location, convenient task initiation. Further, we argue that the findings of our work can contribute towards creating effective task assignment strategies for future cross-device crowdsourcing platforms. However, further work is needed to examine how task performance varies across devices as well as developing appropriate cross-platform task matching mechanisms.

## References

Acer, U. G.; Broeck, M. v. d.; Forlivesi, C.; Heller, F.; and Kawsar, F. 2019. Scaling Crowdsourcing with Mobile Workforce: A Case Study with Belgian Postal Service. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3(2):35:1–35:32.

[2]https://blog.mturk.com/now-you-can-complete-hits-on-the-new-worker-website-6fab0da9ca80

Bates, D.; Mächler, M.; Bolker, B.; and Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1).

Braun, V., and Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3(2):77–101.

Chandler, J.; Mueller, P.; and Paolacci, G. 2014. Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods* 46(1):112–130.

Chatzimilioudis, G.; Konstantinidis, A.; Laoudias, C.; and Zeinalipour-Yazti, D. 2012. Crowdsourcing with Smartphones. *IEEE Internet Computing* 16(5):36–44.

Chi, P.-Y. P.; Batra, A.; and Hsu, M. 2018. Mobile crowdsourcing in the wild. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '18*, 410–415. New York, New York, USA: ACM.

Chilton, L. B.; Horton, J. J.; Miller, R. C.; and Azenkot, S. 2010. Task search in a human computation market. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, 1–9. ACM.

Daniel, F.; Kucherbaev, P.; Cappiello, C.; Benatallah, B.; and Allahbakhsh, M. 2018. Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions. *ACM Computing Surveys* 51(1):1–40.

Difallah, D. E.; Catasta, M.; Demartini, G.; Ipeirotis, P. G.; and Cudré-Mauroux, P. 2015. The Dynamics of Micro-Task Crowdsourcing: The Case of Amazon MTurk. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, 238–247. Geneva, Switzerland: IW3C2.

Difallah, D.; Filatova, E.; and Ipeirotis, P. 2018. Demographics and dynamics of mechanical turk workers. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, WSDM '18, 135–143. ACM.

Gadiraju, U.; Checco, A.; Gupta, N.; and Demartini, G. 2017. Modus Operandi of Crowd Workers: The Invisible Role of Microtask Work Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1(3):1–29.

Gadiraju, U.; Demartini, G.; Kawase, R.; and Dietze, S. 2019. Crowd Anatomy Beyond the Good and Bad: Behavioral Traces for Crowd Worker Modeling and Pre-selection. *Computer Supported Cooperative Work (CSCW)* 28(5):815–841.

Gadiraju, U.; Kawase, R.; and Dietze, S. 2014. A Taxonomy of Microtasks on the Web. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, 218–223. ACM.

Goncalves, J.; Ferreira, D.; Hosio, S.; Liu, Y.; Rogstadius, J.; Kukka, H.; and Kostakos, V. 2013. Crowdsourcing on the Spot: Altruistic Use of Public Displays, Feasibility, Performance, and Behaviours. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, 753–762. ACM.

Goncalves, J.; Feldman, M.; Hu, S.; Kostakos, V.; and Bernstein, A. 2017. Task Routing and Assignment in Crowdsourcing Based on Cognitive Abilities. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, 1023–1031. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee.

Goyal, T.; Mcdonnell, T.; Kutlu, M.; Elsayed, T.; and Lease, M. 2018. Your Behavior Signals Your Reliability: Modeling Crowd Behavioral Traces to Ensure Quality Relevance Annotations. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*, HCOMP, 41–49.

Gummidi, S. R. B.; Xie, X.; and Pedersen, T. B. 2019. A survey of spatial crowdsourcing. *ACM Transactions on Database Systems* 44(2).

Hara, K.; Adams, A.; Milland, K.; Savage, S.; Callison-Burch, C.; and Bigham, J. P. 2018. A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 449:1–449:14. ACM.

Hassani, A.; Haghighi, P. D.; and Jayaraman, P. P. 2015. Context-Aware Recruitment Scheme for Opportunistic Mobile Crowdsensing. In *2015 IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS)*, 266–273. IEEE.

Hettiachchi, D.; van Berkel, N.; Hosio, S.; Kostakos, V.; and Goncalves, J. 2019. Effect of Cognitive Abilities on Crowdsourcing Task Performance. In *Human-Computer Interaction – INTERACT 2019*, 442–464. Springer.

Hettiachchi, D.; Sarsenbayeva, Z.; Allison, F.; van Berkel, N.; Dingler, T.; Marini, G.; Kostakos, V.; and Goncalves, J. 2020a. "Hi! I am the Crowd Tasker" Crowdsourcing through Digital Voice Assistants. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20. New York, NY, USA: ACM.

Hettiachchi, D.; van Berkel, N.; Kostakos, V.; and Goncalves, J. 2020b. CrowdCog: A Cognitive Skill based System for Heterogeneous Task Assignment and Recommendation in Crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction* 4(CSCW2).

Hosio, S.; Goncalves, J.; Lehdonvirta, V.; Ferreira, D.; and Kostakos, V. 2014. Situated crowdsourcing using a market model. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST 2014, 55–64.

Hosio, S. J.; Karppinen, J.; Takala, E.-P.; Takatalo, J.; Goncalves, J.; van Berkel, N.; Konomi, S.; and Kostakos, V. 2018. Crowdsourcing Treatments for Low Back Pain. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, volume 2018-April, 1–12. New York, New York, USA: ACM Press.

Ikeda, K., and Hoashi, K. 2017. Crowdsourcing GO: Effect of worker situation on mobile crowdsourcing performance. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1142–1153. ACM.

Irani, L. C., and Silberman, M. S. 2013. Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 611. ACM Press.

Kaplan, T.; Saito, S.; Hara, K.; and Bigham, J. P. 2018. Striving to Earn More: A Survey of Work Strategies and Tool Use Among Crowd Workers. In *Proceedings of the Sixth AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*. AAAI.

Kazai, G.; Kamps, J.; and Milic-Frayling, N. 2012. The face of quality in crowdsourcing relevance labels. In *Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12*, CIKM '12, 2583. ACM.

Kittur, A.; Nickerson, J. V.; Bernstein, M.; Gerber, E.; Shaw, A.; Zimmerman, J.; Lease, M.; and Horton, J. 2013. The Future of Crowd Work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, 1301–1318. ACM.

Lascau, L.; Gould, S. J. J.; Cox, A. L.; Karmannaya, E.; and Brumby, D. P. 2019. Monotasking or multitasking: Designing for crowdworkers' preferences. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–14. New York, New York, USA: ACM.

Ma, X.; Cackett, M.; Park, L.; Chien, E.; and Naaman, M. 2018. Web-Based VR Experiments Powered by the Crowd. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, 33–43. ACM Press.

Martin, D.; Hanrahan, B. V.; O'Neill, J.; and Gupta, N. 2014. Being a turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*, CSCW '14, 224–235. ACM Press.

Matejka, J.; Glueck, M.; Grossman, T.; and Fitzmaurice, G. 2016. The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5421–5432. New York, NY, USA: ACM.

Mavridis, P.; Gross-Amblard, D.; and Miklós, Z. 2016. Using Hierarchical Skills for Optimized Task Assignment in Knowledge-Intensive Crowdsourcing. In *Proceedings of the 25th International Conference on World Wide Web - WWW '16*, 843–853. ACM Press.

Saito, S.; Nakano, T.; Chiang, C. W.; Kobayashi, T.; Savage, S.; and Bigham, J. P. 2019. TurkScanner: Predicting the hourly wage of microtasks. In *The Web Conference 2019 - Proceedings of the World Wide Web Conference*, WWW 2019, 3187–3193.

To, H.; Ghinita, G.; and Shahabi, C. 2014. Framework for protecting worker location privacy in spatial crowdsourcing. *Proceedings of the VLDB Endowment*.

Vashistha, A.; Garg, A.; and Anderson, R. 2019. ReCall: Crowdsourcing on Basic Phones to Financially Sustain Voice Forums. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 169:1–169:13. ACM.

Vashistha, A.; Sethi, P.; and Anderson, R. 2017. Respeak: A Voice-based, Crowd-powered Speech Transcription System. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, 1855–1866. ACM.

Vashistha, A.; Sethi, P.; and Anderson, R. 2018. BSpeak: An Accessible Voice-based Crowdsourcing Marketplace for Low-Income Blind People. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, CHI '18, 1–13. ACM.

Williams, A. C.; Mark, G.; Milland, K.; Lank, E.; and Law, E. 2019. The perpetual work life of crowdworkers: How tooling practices increase fragmentation in crowdwork. *Proceedings of the ACM on Human-Computer Interaction* 3(CSCW).