Running Out of Time: The Impact and Value of Flexibility in On-Demand Crowdwork

Ming Yin, Siddharth Suri, Mary L. Gray
Microsoft Research
Ming.Yin, suri, mlg@microsoft.com

ABSTRACT
With a seemingly endless stream of tasks, on-demand labor markets appear to offer workers flexibility in when and how much they work. This research argues that platforms afford workers far less flexibility than widely believed. A large part of the "inflexibility" comes from tight deadlines imposed on tasks, leaving workers little control over their work schedules. We experimentally examined the impact of offering workers control of their time in on-demand crowdwork. We found that granting higher "in-task flexibility" dramatically affected the temporal dynamics of worker behavior and produced a larger amount of work with similar quality. In a second experiment, we measured the compensating differential and found that workers would give up significant compensation to control their time, indicating workers attach substantial value to in-task flexibility. Our results suggest that designing tasks which give workers direct control of their time within tasks benefits both buyers and sellers of on-demand crowdwork.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Flexibility; Crowdsourcing; On-Demand Crowdwork.

INTRODUCTION
The past decade produced a sharp spike in the growth of contingent employment, particularly a rise in "on-demand work"—jobs sourced, scheduled, managed, shipped and billed through APIs provided by online platforms [13, 22, 7]. Pundits—champions and critics alike—claim that on-demand jobs offer *flexibility* to work as much as and whenever a worker wants and, in the process, jettison the set hours and commitments that come with traditional workplaces [20, 23].

At first glance, the supposed flexibility of on-demand labor markets might seem appealing—work opportunities streaming in night and day across a platform where workers can choose whatever time, location, and manner to work that is most convenient for them. Recent studies, however, have begun to ask: Are online labor markets as flexible as they are assumed to be? The very nature of "on-demand" work dictates that the volume and kinds of work opportunities available through on-demand platforms constantly fluctuates [4]. A significant portion of workers treat on-demand work as a major source of income, often to mitigate the impact of changes in their other income sources. This pushes workers to work whenever there is demand from requesters rather than enjoying the flexibility of determining their own working schedule.

Even if a worker can fully decide when she would like to participate in on-demand work, she may still face additional constraints *within* each individual task that she works on. For example, consider *crowdwork* that is completed by on-demand labor through crowdsourcing platforms like Amazon Mechanical Turk (MTurk). Requesters must select the maximum amount of time assigned to the task, a key parameter called the "time allotted," before they can post their tasks. A worker needs to complete the task within this time limit in order to get paid; otherwise, the task will expire and she may not be able to accept the task again.

Although it appears obvious that the time allotted parameter would substantially affect worker's work experience in a task, it is not clear to requesters how they should set this parameter properly. Anecdotally, we have heard that requesters often set the time allotted parameter by doubling or tripling their estimated time to complete the task without thinking about whether they need a task returned that quickly, or leave it as MTurk's default value of 1 hour regardless of how long the task actually takes. These somewhat arbitrarily set time limits on individual tasks inevitably create a degree of inflexibility for workers. Requesters may let tight deadlines for tasks stand even when they have not fully considered or know how to precisely gauge the "time allotted" for tasks. Also, hard limits on "time allotted" parameters leave workers ill-equipped to deal with occasional interruptions in the tasks, like the need for restroom breaks, checking on a child asleep in an adjacent room, or picking up a phone call, without having the tasks expire. Workers may also find it challenging to schedule multiple tasks and must pass up some tasks that they find interesting or well-prepared to do due to time conflicts [3]. In other words, time limits on on-demand tasks leave workers constantly facing the risk of running out of time in tasks, and

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ISBN 978-1-4503-5620-6/18/04…$15.00

DOI: https://doi.org/10.1145/3173574.3174004

workers also have to take such risk into consideration when deciding which tasks to take in the first place.

Previous studies of traditional workplaces suggest that increasing a worker's temporal flexibility, that is, her ability to control the scheduling of her work, can significantly improve both the work output and the worker experience of a job [1, 12, 18]. It is therefore natural to ask *if*, like traditional work, on-demand work may benefit from worker-focused approaches to "flexibility" that offer workers explicit control of their time to complete on-demand tasks when these approaches are feasible. On the one hand, the benefits associated with greater worker control of their time may be generalizable across different work environments. On the other hand, it is unclear how providing more temporal flexibility affects workers when tasks, like the ones in on-demand work, take a small amount of time, especially given that workers in on-demand work are believed to already enjoy a lot of freedom in scheduling their work.

As a first step towards answering this question, in this paper, we focus on *on-demand crowdwork* and address two main questions related to better understanding flexibility in on-demand crowdwork:

1. What is the impact of granting more flexibility in on-demand crowdwork on worker behavior?
2. What is the value that workers attach to flexibility in on-demand crowdwork?

We present two randomized behavioral experiments to answer the two questions above. While it would be difficult to experimentally manipulate and control when all on-demand tasks post to the market to measure scheduling flexibility, it is relatively easy to manage the flexibility *within* a task by allotting a different amount of time to the task. We refer to this as allowing for *in-task flexibility*. We conducted both experiments using *batches* of sentiment analysis HITs, that is, we posted multiple sentiment analysis HITs for workers to complete in the same HIT group. We made this choice as batches of sentiment analysis tasks are popular on crowdsourcing platforms [11]. In additon, batch tasks leave more room for requesters to adjust within task flexibility levels compared to other types of tasks like surveys where requesters may set tight time limits to ensure undivided attention from workers.

In the first study, we show that workers will indeed exercise control over their work time when more in-task flexibility is granted in crowdwork, which is reflected by the dramatic changes in the temporal dynamics of worker behavior, such as when they start to work on a task or how long they spend on a task. In addition, higher levels of in-task flexibility also lead to a larger quantity of work without sacrificing work quality.

In the second study, we measure how much workers value controlling their time in crowdwork using two different methods. First, we infer the value through workers' reported preferences between pairs of tasks with different prices and/or amounts of time allotted. Second, we estimate the *compensating differential*, that is, the extra amount of money needed to pay workers to complete the same amount of work when given less in-task flexibility as they would complete when given more in-task flexibility. Both methods suggest that workers attach substantial value to control of one's time through temporal flexibility within on-demand tasks. For example, we estimate the compensating differential to be at least \$0.86/hour, which means that on average, workers equate the ability to control scheduling their work with a financial compensation of at least \$0.86/hour. Interestingly, we also find that workers who report spending more time earning money from on-demand platforms than their peers place a higher value on in-task flexibility.

In sum, our results suggest that on-demand workers may experience constraints on flexibility as limiting, and would value greater control of their time in tasks where feasible.

RELATED WORK

The impact of *job flexibility* on workers has been extensively studied within traditional companies and organizations in the organizational behavior and psychology literature. While the broad term of "job flexibility" includes various dimensions, like work schedule and work location, the concept of "*temporal flexibility*" or "*work time control*" specifically refers to flexibility applied to time working on specific tasks. Temporal flexibility can be further divided into sub-dimensions, such as control over when to start and end the workday (i.e., "flextime"), when to take breaks, and when to take days off or work overtime [1, 12, 17].

A significant body of research demonstrates the relationship between temporal flexibility in traditional workplaces and job-related outcomes that impact various aspects of workers' lives. For example, an increase in flextime has been associated with positive effects on worker productivity and job satisfaction [1]. Flexible working arrangements, such as self-scheduled shifts, arguably, improve workers' health and wellbeing [12]. Research has also shown that organizational interventions designed to promote greater employee control over work time not only reduced employee's perceived stress [16], but also lowered their turnover intentions [17]. In addition, there is further evidence that giving workers greater temporal flexibility improves workers' work-life balance [9, 18].

Two main mechanisms appear to explain why temporal flexibility significantly influences job-related outcomes and a worker's quality of life. First, the *time-regulation mechanism* suggests that work time control allows workers to better regulate their time demands, such as reduce or avoid work-family conflict [5, 21]. Second, the *recovery-regulation mechanism* indicates that temporal flexibility may give workers the opportunities to lessen their fatigue from work by taking breaks as needed or prevent overload in the first place [2, 18].

Our work builds on but is distinct from the studies above in four key aspects. First, we focus on studying flexibility in a relatively novel work environment: *on-demand labor markets*. These platform-based jobs are often composed of discrete projects, typically small-sized tasks, done without the support mechanisms or constraints of a physical worksite but believed to be more flexible than traditional employment. It is thus important to see whether online labor markets are as flexible as their proponents say they are. Second, we examine *if* applying worker-focused flexibility to on-demand crowdwork can produce similar, positive outcomes for work output and worker behavior as it does in traditional work. To the best of our knowledge, our work is the first study to answer this question. Third, we focus on examining the impact and value of *in-task flexibility*, which is reflected by the amount of time allotted to each task and describes workers' freedom in controlling their work time *within* individual tasks. In-task flexibility in on-demand crowdwork is analogous to the flexibility a worker might have to complete each discrete project at a traditional job (e.g., whether a tight deadline is imposed on a project or not) rather than the flextime, offering specific guidance to those designing and gauging the time allotted for tasks. Our study, thus, specifically explores how in-task flexibility, a parameter easily controlled by the task requester, impacts on-demand workers. Finally, in addition to examining the effects of granting more flexibility in on-demand crowdwork, we also adopt an innovative approach to *quantitatively* measure the economic value that workers attach to flexibility in the work, following the methodology in [24, 6].

STUDY 1: EXAMINING THE IMPACT OF FLEXIBILITY

In our first study, we designed and conducted an online experiment on Amazon Mechanical Turk to understand whether and how granting workers with more in-task flexibility can influence the temporal dynamics of worker behavior as well as the work quantity and quality.

Tasks and Treatments

We used a sentiment analysis task for this experiment because it is a very common type of task on MTurk [11]. Specifically, each task contained a 150–200 word Amazon customer review for an automobile related product. Workers were asked to classify whether the review was positive or negative. The set of customer reviews in the tasks was taken from [15] providing us with the ground truth for all reviews used in the experiment. Through a pilot study we found that it took workers less than 30 seconds, on average, to read one review and determine the sentiment in it.

Our experiment consisted of six treatments arranged in a 3×2 design as along these two factors:

- *time allotted*: the amount of time allotted in a task, with three possible levels—1 minute, 1 hour, and 1 day;
- *provision of time estimate*: whether or not to provide an estimate of the task completion time in a task—if yes, we told workers that it takes roughly 30 seconds on average to complete one sentiment analysis task; otherwise, we did not provide this information.

Time allotted is used as a proxy for the level of flexibility workers get in the task. Intuitively, the more time allotted in a task, the more control workers would have over their work time in the task. Importantly, because our pilot study suggested completing one sentiment analysis task typically took less than 30 seconds, we granted workers enough time to complete the task for all three levels of time allotted.

One may wonder how workers would interpret the amount of time allotted in a task. For example, it is possible that workers may view tasks with a larger amount of time allotted as more time-consuming rather than more temporally flexible. Therefore, by varying whether a task completion time estimate was provided, we made the in-task flexibility more salient for workers in some treatments. Such two-factor factorial design (3×2) thus enables us to obtain an in-depth understanding on how workers perceive and are affected by the amount of flexibility granted within a task.

Experimental Procedure

We conducted our experiment in two phases similar to Section 5 of [10]. The first phase was the recruiting phase, in which we posted a 20-cent participant recruiting HIT for future sentiment analysis tasks on MTurk. Workers who were interested in completing some future sentiment analysis tasks (i.e., phase 2) could sign up by answering three survey questions about their usage of MTurk. We asked each participant for the number of years they have been using MTurk, the number of hours spent working on MTurk in the last week, and the number of income sources the worker has outside of MTurk. We then assigned each worker who completed the recruiting HIT to one of the 6 treatments uniformly at random. Next, in the second phase of the experiment, each worker who signed up through the recruiting HIT was provided with a batch of 100 sentiment analysis HITs, where each HIT contained one sentiment analysis task. Depending on the treatment that the worker was assigned, the amount of time allotted to each task was either 1 minute, 1 hour or 1 day, and a completion time estimate may or may not have been included in the instructions of the task. Importantly, using MTurk qualifications, we ensured that workers were only able to see and work on HITs in the treatment that they were assigned. Designing our HIT in 2 phases allowed us to randomly assign workers into treatments and made it so that workers in one treatment could not see another treatment. We posted the sentiment analysis HITs for all treatments at the same time and they were available on MTurk for 24 hours. Workers were asked to complete as many of these HITs as they want, and they were told upfront that the payment for each HIT was 5 cents.

Experimental Data

While workers were working on the sentiment analysis tasks, we recorded data on the temporal dynamics of their working behavior by keeping a detailed log of how each worker interacted with each task. In particular, when worker i accepted a task t, we recorded a timestamp a_i^t as the time for *task acceptance*. Once a worker accepted a task, the task would be automatically added into her *task queue* on MTurk. A worker can enter and leave any task in her queue as many times as she wants, as long as the time allotted for the task has not been reached. Depending on how worker i interacted with task t after she accepted it, we recorded some additional timestamps. One possible scenario is that worker i had task t open in her browser ever since she accepted it, in which case the only other timestamp (if any) we collected for the worker on this task is s_i^t, the time for *task submission*[1]. Another possibility is that the worker left task t (i.e., closed the browser tab which contained task t) after accepting it and later came back to work on it

[1]Note that having a task open in the browser does not imply that the worker is working on the task. For example, the worker can work on one task while keeping other tasks open in her browser, or the worker can take a break within a task. It is difficult to precisely determine if and when the worker actually works on a task.

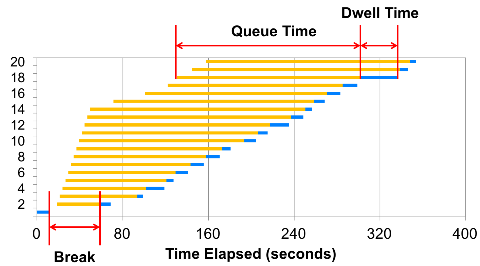


Figure 1: An example showing the metrics of working behavior. In this example, the worker completed 20 tasks in total, and each task is represented as a horizontal bar. The leftmost end of each bar represents the task acceptance time. The rightmost end of each bar represents the task submission time. Each bar may be further divided into an orange part and a blue part. The transition point from orange to blue in each bar (if any) represents the time when the worker re-enters the task for the last time.

from her task queue. In this case, besides the task submission timestamp s_i^t, we also kept another sequence of timestamps $r_i^t(j)$, $1 \leq j \leq n_i^t$, with $r_i^t(j)$ representing the time when worker i *re-entered* task t from her task queue for the j-th time, and n_i^t was the total number of times that worker i re-entered task t. Naturally, we have $a_i^t < r_i^t(1) < \cdots < r_i^t(n_i^t) < s_i^t$.

We next sorted all the tasks that worker i completed according to the increasing order of the task acceptance time, and defined a few metrics for measuring the temporal dynamics of worker behavior in the tasks:

- *queue time (q_i^t)*: the amount of time elapsed between worker i accepting task t and entering the task for the last time. When $n_i^t = 0$, $q_i^t = 0$; otherwise, $q_i^t = r_i^t(n_i^t) - a_i^t$.
- *dwell time (d_i^t)*: the amount of time elapsed between worker i entering task t for the last time and submitting the task. When $n_i^t = 0$, $d_i^t = s_i^t - a_i^t$; otherwise, $d_i^t = s_i^t - r_i^t(n_i^t)$.
- *between-task break*: the amount of time between worker i submitting one task (e.g., task t) and entering the next task (e.g., task $t+1$) for the last time.

Figure 1 gives a visual example on how the above working behavior metrics are defined.

Furthermore, to examine the impact of granting more in-task flexibility on work quantity, we used two metrics, the number of tasks that a worker *accepted* and *completed*. Given that we only presented 100 sentiment analysis HITs to each worker, both metrics had an upper bound of 100. With respect to work quality in the tasks, since we had ground truth data on the sentiment of each review, we calculated each worker's *accuracy* (averaged over all tasks the worker completed) and used this as the metric for work quality.

Research Questions and Hypotheses

We aim to answer three research questions (Q1–Q3) through Study 1. After each question we state our hypotheses.

Q1: How does granting workers more in-task flexibility affect worker behavior on tasks?

Granting workers higher levels of in-task flexibility empowers them to choose when to start, work on, and finish the tasks; hence, if workers indeed take advantage of the flexibility that they get within each task, we are likely to observe changes in the temporal dynamics of their working behavior. For example, a worker may decide to start working on a task later so that she can first complete work, either from within or outside of the on-demand platform, that is more urgent or time-sensitive. The degree to which workers utilize the control they are granted through in-task flexibility over when to start working on a task can be measured using the amount of time workers put their tasks in queues. Higher levels of in-task flexibility may also give workers the possibility of controlling the tempo of their work pace, including intentionally working slower or taking breaks within a task to deal with interruptions or recover from fatigue. As a result, with more in-task flexibility, workers may dwell on a task for a longer period of time while workers' needs for taking breaks between subsequent tasks may decrease. We conjecture that:

[H1] Workers will increase the queue time for a task when the amount of time allotted in the task is larger.

[H2] Workers will dwell on a task for longer time when the amount of time allotted in a task is larger.

[H3] Workers will take fewer between-task breaks when the amount of time allotted in each task is larger.

[H4] The first time that a worker needs to take a between-task break is later when the amount of time allotted in each task is larger.

Possible changes in the temporal dynamics of worker behavior may, in fact, signal that workers experience constraints on in-task flexibility as limiting and imply a fundamentally different way of working that workers prefer when higher levels of in-task flexibility are provided. That is, instead of racing the clock that is set by the requesters, workers may be able to work at their own pace. As a result, more in-task flexibility may allow workers to choose a time to work that is optimal for *them*. We thus hypothesize that in doing so, workers can get more done with at least as good accuracy.

Q2: How does granting workers more in-task flexibility affect work quantity?

[H5] Workers will accept and complete more tasks when the amount of time allotted in a task is larger.

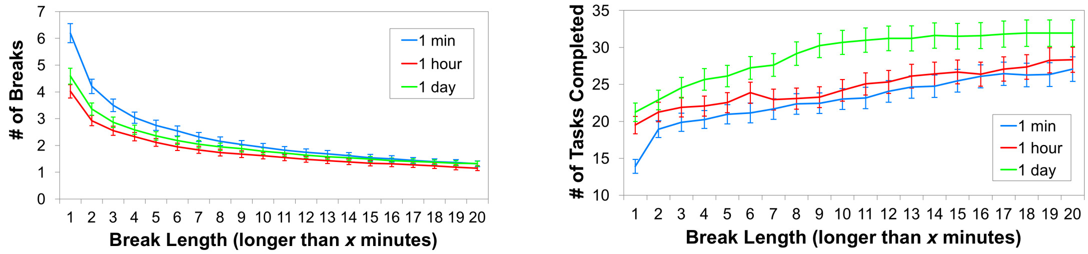
[H6] The effect of time allotted on work quantity is larger when an estimate on the task completion time is provided in the task.

Q3: How does granting workers more in-task flexibility affect the quality of work produced?

[H7] Workers will maintain or increase their accuracy when the amount of time allotted in a task is larger.

Experimental Results

In total 1,999 workers signed up through the recruiting HIT in phase 1 of our experiment, of these, 1,379 workers accepted at least one sentiment analysis task in phase 2. We ran a



(a) Number of between-task breaks

(b) Timing of the first between-task break

Figure 2: 2a: The average number of between-task breaks a worker takes that are longer than *x* minutes across treatments with different time allotted. 2b: The average number of tasks a worker has completed before she takes her first between-task break that is longer than *x* minutes (averaged over workers who take at least one such break). Error bars show standard errors of the mean.

	Avg. queue time (SD)	**Avg. dwell time (SD)**
1 min	0.03 (0.01)	22.75 (0.35)
1 hour	17.28 (4.35)	47.12 (2.27)
1 day	254.38 (110.74)	66.18 (5.23)

Table 1: Mean values and standard deviations of worker queue time and dwell time (in seconds) across different treatments.

chi-square test to check for significant differences in the percentage of workers choosing to participate in phase 2 across the six treatments and found none ($p = 0.71$). Similarly, for workers who participated in phase 2 of our experiment, we observed no significant difference in their responses to survey questions in the recruiting HIT across the six treatments.

Impact on Working Behavior

We start by examining the impact of in-task flexibility on worker behavior (Q1) to see whether and how workers leverage the flexibility granted in tasks. As we do not observe significant differences in working behavior between treatments with or without the provision of a task time estimate, we conduct our analyses on data after combining treatments with the same time allotted level together.

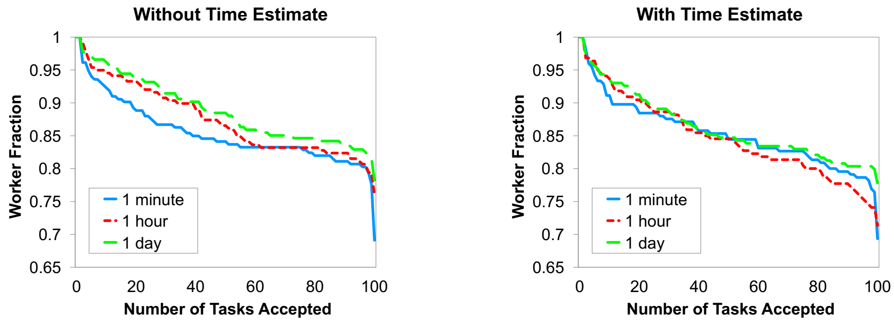
We first attempt to understand whether granting more in-task flexibility has any impact on worker's behavior *within* individual tasks. For each worker, we calculate her *average queue time* and *average dwell time* over all tasks that she completed. Table 1 compares these two metrics across treatments with different amounts of time allotted. The table shows that granting more flexibility within a task leads to *substantial* increases in both worker's queue time and dwell time, and such an increase is confirmed to be statistically significant through one-way Kruskal Wallis ANOVA tests (for queue time, $\eta^2 = 0.088$, $p < 10^{-26}$; for dwell time, $\eta^2 = 0.209$, $p < 10^{-61}$). These results are in line with our hypotheses H1 and H2—with more in-task flexibility, workers have greater control on when they *start* working on a task (hence longer queue time), and workers may also intentionally work slowly or take breaks *in* a task if needed (hence longer dwell time).

Next, we move on to examine the impact of in-task flexibility on how workers behave *between* tasks. Figure 2a compares the average number of breaks of different lengths that a worker takes between subsequent tasks when the time allotted in a task differs. In general, we find that compared to workers who are assigned to the 1-minute tasks, workers in the 1-hour or 1-day treatments seem to take far fewer breaks between tasks, especially when the length of the break is relatively short. For example, when workers are allotted 1 minute for the task, they need to take significantly more short-length, 1-to-5-minute breaks ($\eta^2 = 0.047$, $p = 5.32 \times 10^{-15}$) and medium-length, 5-to-10-minute breaks ($\eta^2 = 0.006$, $p = 0.005$), between subsequent tasks. Figure 2b further shows that workers in treatments with longer time allotted also take their first between-task break after completing more tasks, and this difference is also statistically significant (e.g., $p < 0.05$ for the differences across treatments in the number of tasks completed before the first break that is longer than 1, 5, 10 or 30 minutes). Put another way, with more in-task flexibility, workers not only take *fewer* breaks between tasks, but also take breaks *later*. This is consistent with our hypothesis H3 and H4—when more time is allotted in a task, workers can take breaks within the task to deal with interruptions or recover from fatigue, hence taking breaks between subsequent tasks become less necessary.

Together, our analyses reveal some significant changes in the temporal dynamics of worker behavior when sufficient flexibility is granted in the tasks. This implies that workers actually exercise control over their work time and thus leverage in-task flexibility when provided. It's worthwhile to note that such dramatic changes are mostly caused by a small fraction of workers on a small number of tasks. For example, in 1-day treatments, about 10% of the workers have an average queue time that is longer than 2 minutes, and the percentage of tasks that workers put in their queues for more than 2 minutes is about 5%. Correlating workers' responses to survey questions in the recruiting HIT with their working behavior, we find that workers who spend more hours on MTurk seem to leverage the in-task flexibility to a larger degree by dwelling on tasks or putting tasks in queues for a significantly longer period of time. Such difference is not observed when we divide workers by their experience levels or number of income sources outside of MTurk. These observations suggest that it is predominantly those workers who spend lots of time working on MTurk that take advantage of the extra in-task flexibility. We will examine how the needs for flexibility differ within the on-demand worker population more carefully in Study 2.

Impact on Work Quantity

Our second goal is to understand the impact of in-task flexibility on work quantity to answer Q2. We first measure how granting more in-task flexibility affects the number of tasks



(a) Without provision of the completion time estimate (b) With provision of the completion time estimate

Figure 3: Retention curves showing the fraction of workers who accepted at least x tasks.

workers *accept*. Figures 3a and 3b show the curves of the fraction of workers who accepted at least x ($0 \leq x \leq 100$) tasks for treatments without or with a completion time estimate, respectively. Visually, we find that given a fixed x, the fraction of workers who accepted at least x tasks tend to be higher in treatments where time allotted for each task is longer. The biggest difference occurs between workers in the 1-minute treatment and workers in the 1-day treatment when task time estimate is *not* provided.

The data on the number of tasks a worker accepted is highly skewed. Figure 3 shows that 70%–80% of the workers accepted all 100 tasks that we offered to them. To statistically analyze these over-dispersed count data, we used negative binomial regressions (as done in [6]) and the results are reported in Table 2. We first consider only the *main effect* of in-task flexibility on the number of tasks workers accept, using the 1-minute treatment without the estimate on task completion time as the reference (Model 1). According to the regression results, providing extra time within a task increases the number of tasks a worker accepted, and such increase is statistically significant when a long period of time (i.e., 1 day) is allotted to a 30-second task.

Model 1 suggests the provision of the task time estimate significantly *decreases* the number of tasks workers accept. Furthermore, as Model 2 in Table 2 shows, the *interaction* between time allotted and the provision of time estimate is observed to be significantly *negative*. This suggests that making the difference between time allotted and the actual time cost in a task more salient does *not* help workers perceive the flexibility in a task more accurately and work quantity is not further improved. Overall, our examination of the impact of in-task flexibility on task acceptance supports H5 and rejects H6.

We conducted similar analyses to understand the impact of in-task flexibility on the number of tasks a worker *completed* (i.e., Models 3 and 4 in Table 2). Again, we find that granting extra time in a task leads to a marginally significant increase in the number of tasks a worker submits, while the provision of task time estimate does not affect task submission much. Thus, we can conclude from our analyses on both task acceptance and submission that more in-task flexibility leads to a larger quantity of work (i.e., H5 is supported), yet the magnitude of such positive impact does not increase with the provision of time cost estimation of tasks (i.e., H6 is rejected).

	# of accepted (Model 1)	**# of accepted (Model 2)**	**# of submitted (Model 3)**	**# of submitted (Model 4)**
Intercept	4.456^{***} (0.006)	4.446^{***} (0.007)	4.388^{***} (0.006)	4.394^{***} (0.007)
w/ estimate	-0.021^{***} (0.006)	-0.002 (0.010)	0.005 (0.006)	-0.006 (0.011)
1 hour	0.009 (0.007)	0.024^{*} (0.010)	$0.013^{\dagger}$ (0.007)	0.008 (0.010)
1 day	0.026^{***} (0.007)	0.039^{***} (0.010)	$0.014^{\dagger}$ (0.007)	0.003 (0.010)
w/ estimate × 1 hour		-0.030^{*} (0.014)		0.010 (0.015)
w/ estimate × 1 day		-0.028^{*} (0.028)		0.023 (0.015)

Table 2: Regression results for work quantity. Coefficients and standard errors are reported. Significance levels: † ($p < 0.1$), * ($p < 0.05$), *** ($p < 0.001$).

We provide three possible explanations for why H6 is rejected. First, our original hypothesis, that the provision of a task time estimate would increase the positive effect of in-task flexibility on work quantity, was based on the assumption that workers may use time allotted in a task as a proxy for the actual time required to do the task. Had this assumption held then telling workers an estimated time cost of a task could have helped workers realize that they are able to control their time in the task rather than the task is time-consuming. The conclusions from our data analyses imply that for simple tasks like sentiment analysis, it is likely that workers can quickly estimate the time cost themselves through working on a few of these tasks. So, there is no need for them to use the time allotted to infer a task's time cost in lieu of their own evaluation of the task's demands.

Second, while the provision of a task completion time estimate makes the difference between the actual time cost of a task and the time allotted to it salient to the workers, workers do not know *why* they get extra amount of time in these short tasks. As such, workers may worry about the potential mismatch between the requester's expectation and their own understandings of the tasks, and thus may hesitate to accept more tasks in order to minimize their risk of getting their work rejected.

Finally, it is also possible that some workers find the task time estimate we provide is not consistent with their own experiences and thus decide to stop accepting tasks due to psychological factors like mistrust to the requester or low levels of self-efficacy. The latter can be especially true if a worker finds that for her, completing a sentiment analysis task takes significantly longer than 30 seconds.

Impact on Work Quality

Lastly, we explore the relationship between in-task flexibility and the quality of work produced (Q3). Figure 4 displays the average accuracy for workers in each of the six treatments in our experiment, which shows an upward trend as more time is allotted in a task. One-way Kruksal Wallis ANOVA tests on the data further indicate that the differences in worker accuracy across treatments with different time allotted is marginally significant when task time estimate is not provided ($p = 0.096$) and insignificant when such estimate is provided ($p = 0.184$). Interestingly, we also find that workers are significantly more accurate when task time estimate is provided ($p = 5.080 \times 10^{-5}$). We conjecture that this is because workers interpret the provision of time estimate as a signal of the requester's familiarity with his tasks and thus workers choose to consciously keep producing high-quality work to satisfy the requester. Overall these results are consistent with H7, suggesting that granting more in-task flexibility does not harm the work quality.

STUDY 2: MEASURING THE VALUE OF FLEXIBILITY

In Study 1, we observe that workers actively respond to more flexibility in tasks with significant changes in working behavior, increased work quantity and similar work quality. These results appear to imply that on-demand workers *value* the flexibility within the tasks, at least to some degree. Naturally, one may ask *to what degrees* do workers value in-task flexibility. The purpose of Study 2 is to answer this question by quantitatively measuring the economic value that workers attach to in-task flexibility.

Survey-Based Value Estimation: A Pilot Study

To get an initial rough estimate of how much workers value in-task flexibility in on-demand crowdwork, we first conducted a pilot study in which we surveyed 399 MTurk workers. Workers in the survey were asked to indicate their preference between pairs of proposed task designs. For any two tasks presented to workers in a pair, the task content was the same (i.e., a 30-second sentiment analysis task) but the pay rate and/or time allotted varied. More specifically, about half of the workers were asked if they would prefer completing the task in 1 minute to earn 5 cents or completing the task in 1 day to earn $1 \leq x \leq 5$ cents. The other half of the workers were asked if they would prefer completing the task in 1 day for 5 cents or completing it in 1 minute for $5 \leq y \leq 9$ cents. Workers' reported preferences allow us to get a rough estimate of how much workers value in-task flexibility. For the purposes of this survey we can describe a task as a tuple consisting of the time allotted and the pay rate. So, for example, if a worker prefers (1 minute, 5 cent) to (1 day, x cents) for $x \in \{1,2\}$ but prefers (1 day, x' cents) to (1 minute, 5 cents) for $x' \in \{3,4,5\}$, then this worker's value for in-task flexibility is somewhere between 2 and 3 cents.

Taking the lower bound of each worker's value, we find that on average, workers who choose between (1 minute, 5 cents) and (1 day, x cents) are willing to forego at least 0.69 cents to get more control of their time for this 30-second task, which is

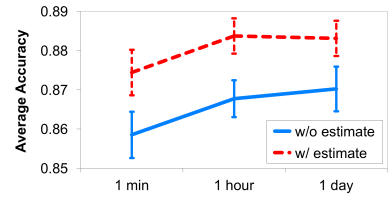


Figure 4: Workers' average accuracy in different treatments. Error bars represent standard errors of the mean.

equivalent to \$0.83/hour (i.e., $0.0069 \times 3600/30$). Conversely, workers ask for at least an additional \$1.21/hour if they have to give up the flexibility that they have already been provided (i.e., switch from (1 day, 5 cents) to (1 minute, y cents)).

Measuring Value as Compensating Differential: Experimental Design

Our pilot survey suggests that workers attach substantial economic value to in-task flexibility. However, one limitation of estimating the value of flexibility through a survey is that workers indicated their preferences by imagining their choices between pairs of tasks. Preferences elicited this way are called "*stated preferences*," which may be different from "*revealed preferences*" which are reflected by worker's actual decisions. To compute worker's revealed preferences we designed and conducted an experiment to compute the *compensating differential* of in-task flexibility, that is, the extra amount of money requesters need to pay workers for them to complete the same amount of work in the absence of in-task flexibility as that in the presence of in-task flexibility. We adopt a similar approach that was previously used to quantify the value of a clean and clear user interface [24] and estimate the economic costs of annoying display ads [6].

The task we used in this experiment was again a sentiment analysis task, but with a different set of automobile product reviews than that used in Study 1. We created a set of six treatments, in a 2×3 design, defined by two dimensions: *time allotted*, with 1 minute and 1 day as the two possible levels; and *task price*, with the three levels of 3 cents, 4 cents, and 5 cents. This range of task prices was chosen as it roughly equals an effective hourly wage of \$5.4–\$9/hour, which is typical for MTurk. Since Study 1 informed us that the provision of task completion time estimate does not help workers to perceive in-task flexibility more accurately, we did not provide such information in any of the six treatments in Study 2.

Like Study 1, we conducted this experiment in two phases. In the first phase, we posted a 20-cent HIT on MTurk to recruit workers who are interested in completing some sentiment analysis tasks in the future (i.e., phase 2). Each worker who took this recruiting HIT answered two questions:

- *Activity level*: In the past week, how many hours have you spent both searching for and doing tasks on MTurk?
- *Income goal*: Do you set a daily target for the total amount of income you want to earn from completing tasks on MTurk?

Our analyses on the temporal dynamics of worker behavior in Study 1 suggests that workers who spend more hours on

Price / Time Allotted	1 minute	1 day
3 cents	59.43 (39.08)	65.89 (39.08)
4 cents	64.54 (36.28)	75.02 (33.54)
5 cents	72.52 (34.64)	76.64 (33.48)

Table 3: Mean values and standard deviations (in parentheses) of the number of submitted tasks across the six treatments in Study 2.

the platform tend to leverage in-task flexibility more. Thus, in Study 2, we attempt to understand the relationship between a worker's engagement with on-demand platforms and her value of in-task flexibility more directly. We use a worker's responses to the above two questions as a proxy of worker's engagement level with MTurk. After the first phase of the experiment, we randomly assigned each worker who submitted the recruiting HIT into one of the six treatments. Then, in the second phase of our experiment workers got the opportunity to work on a batch of 100 sentiment analysis tasks. The price and time allotted for each task a worker could work on was decided by the treatment the worker was assigned, and workers were not able to see tasks from other treatments on MTurk. Here again, we used this two-phase design to randomly assign workers to treatments and ensure that workers could only see their own treatment. Workers were instructed to complete as many of these sentiment analysis tasks as they want. As we will compute the compensating differential based on the impact of in-task flexibility on task submission, the major dependent variable we used in this experiment is the number of tasks a worker completed, although we also recorded all the relevant data as used in Study 1.

Experimental Results

Overall 1,800 workers signed up in the first phase of our experiment, and 1,079 workers submitted at least one task in phase 2. Across the six treatments, no significant differences were found in the portion of workers who participated in our second phase experiment or the responses to survey questions.

As a robustness check, we first repeated all analyses that we conducted in Study 1 on the data that we collected through Study 2. Our analyses support all results that we get from Study 1: With higher levels of in-task flexibility, workers tend to put tasks in queues and dwell on tasks for longer periods of time[2], take significantly fewer number of between-task breaks and take these breaks later. Work quantity increases when more time is allotted as reflected by the significant increases in task acceptance and submission[3]. Granting extra amount of time in a task also increases the average worker accuracy from 0.889 in 1-minute treatments to 0.894 in 1-day treatments, although it is not statistically significant ($p = 0.712$). In other words, our findings in Study 1 are robust.

[2]For queue time, the effect sizes as measured by Cohen's d are 0.257, 0.166 and 0.210, respectively, for the 3 comparisons for 3 task prices (i.e., 3, 4 or 5 cents), and the largest p-value is 2.30×10^{-8}; similarly, for dwell time, the three Cohen's d values are 0.144, 0.611, and 0.320, respectively, and the largest p-value is 8.94×10^{-7}.

[3]Across the 3 comparisons for 3 task prices, for task acceptance (or submission), the smallest Cohen's d is 0.066 (or 0.121) and the largest p-value is 0.049 (or 0.030).

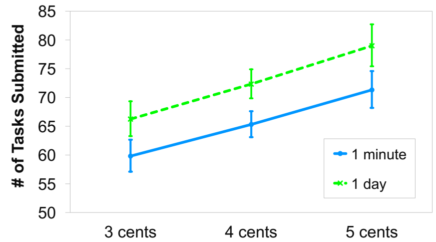


Figure 5: The number of submitted tasks in different treatments predicted by the negative binomial model in Table 4 (left section). Error bars extend one standard error above and below the predicted values.

Workers Attach Substantial Value to Flexibility

Table 3 summarizes the mean values and standard deviations for the number of tasks workers submitted across the six treatments in Study 2. In general, we find higher task price and more in-task flexibility leads to increased number of submitted tasks. The large standard deviations indicate the over-dispersion of the data. Thus, following a similar approach as in [6], we fit a negative binomial regression model to this over-dispersed data and compute the compensating differential based on the model prediction. Table 4 (left section) shows the fitted model, and Figure 5 plots the model prediction in the original scale. The regression results confirm that increasing the task price significantly increases the number of tasks workers submit, and so does the provision of extra amount of time in the tasks.

As an example, we compute the compensating differential of in-task flexibility for workers who were assigned to the (1 minute, 4 cents) treatment. Based on the model predictions, we estimate the effect of a task price increase on the number of tasks submitted by averaging the increase in the number of submitted tasks when raising the task price from 3 to 4 cents, and from 4 to 5 cents, while fixing the amount of time allotted in a task at 1 minute. Doing so suggests that a 1 cent pay raise leads to an average of 5.76 additional task submissions. Moreover, when the task price is fixed at 4 cents, moving from a 1-minute task to a 1-day task implies an average increase of 7.02 task submissions. Next we calculate how much more we would need to pay someone in the (1 minute, 4 cents) treatment to do the same amount of work as someone in the (1 day, 4 cents) treatment. We find that for workers who work on 4-cent tasks with a 1 minute time limit, the pay raise required to match the effect of allotting extra time (i.e., 1 day) in each task is 1.22 cents (i.e., $1 \times 7.02/5.76$) per task. Our data suggests that on average, workers who were assigned to the (1 minute, 4 cents) treatment dwell on a task for 18.9 seconds, which means that workers equate the flexibility within the tasks with a financial compensation of \$2.32/hour (i.e., $0.0122 \times 3600/18.9$).

One may question whether this is an overestimate of worker's value of in-task flexibility, because workers in the (1 minute, 4 cents) treatment have little freedom in controlling their work time and have to complete the tasks in a rather fast pace. Thus, we instead use 51.2 seconds—the average amount of time that workers in the (1 day, 4 cents) treatment dwell on a task—as the reference for the time cost of a task. Even with this reference, we can still estimate that workers ask for an addi-

	All Data	Activity: [0, 10]	Activity: [11, 30]	Activity: 30+	Goal: no	Goal: yes
Intercept	3.827^{***} (0.126)	3.871^{***} (0.225)	3.854^{***} (0.030)	3.863^{***} (0.039)	3.903^{***} (0.307)	3.824^{***} (0.137)
1 day	0.102^{*} (0.049)	-0.009 (0.090)	0.108^{***} (0.011)	0.218^{***} (0.014)	0.029 (0.118)	0.117^{*} (0.054)
price	0.088^{**} (0.030)	0.066 (0.055)	0.096^{***} (0.007)	0.069^{***} (0.009)	0.061 (0.073)	0.091^{**} (0.033)
Number of observations	1079	377	425	277	221	858

Table 4: Negative binomial regression results for the number of tasks workers submitted in Study 2. Left section: regression model fitted using all data in Study 2. Middle section: regression models fitted using data from workers with different activity levels. Right section: regression models fitted using data from workers who have/do not have income goals. Coefficients and standard errors are reported. Significance levels: * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$).

tional \$0.86/hour (i.e., $0.0122 \times 3600/51.2$) for completing the same number of tasks when tasks have little flexibility as they would complete when tasks have sufficient flexibility. This estimate coincides with the \$0.83/hour we calculated from the stated preferences gathered from our survey. Therefore, our experimental results confirm that workers attach substantial economic value to flexibility in on-demand crowdwork.

Engagement Levels with Platforms and the Value of Flexibility

We next move on to examine whether different workers value flexibility to different degrees. In particular, we are interested in understanding how worker's value of flexibility changes with the level of engagement with on-demand platforms. Our conjecture is that the more engaged a worker is with an on-demand platform, the more she values the flexibility of on-demand work. We used a worker's responses to the two survey questions in our first-phase recruiting HIT to approximate the degree to which this worker engages with the on-demand platform MTurk. Intuitively, the more hours a worker spends on MTurk, the more engaged she is with MTurk. Similarly, workers who set a daily income goal for completing tasks on MTurk are likely to engage with MTurk to a larger degree.

We first test whether workers who spend different amounts of time on MTurk value flexibility differently. All workers who participated in our second phase experiment are divided into 3, roughly equal sized, groups based on the amount of time they spent on MTurk in the previous week: less than 10 hours, 11 to 30 hours, or more than 30 hours. To determine how much each group values in-task flexibility we then fit negative binomial regression models for each group of workers separately, and the results are reported in Table 4 (middle section). According to the regression results, in-task flexibility did *not* have any significant impact on the number of tasks a worker submitted, for workers who spend less than 10 hours a week on MTurk. On the other hand, in-task flexibility significantly increases the number of task submissions for workers who spend more than 10 hours a week on MTurk, and the magnitude of such increase is the largest for workers who spend more than 30 hours a week on MTurk. Through a similar calculation of the compensating differential, we estimate that workers who work on MTurk for less than 10 hours a week do not have a positive economic value to in-task flexibility (at least, not one that we could detect), while workers who spend 11 to 30 hours (or, 30 or more hours) a week on MTurk equate the flexibility in tasks with a financial compensation of \$0.98/hour (or, \$2.37/hour)[4]. These results are consistent with our conjecture—the more time a worker spends on an on-demand platform, the more she depends on managing her time on it, and hence the more she values flexibility in on-demand tasks.

Similar conclusions can also be made when we separately examine workers who have or do not have daily income targets for working on MTurk. The negative binomial regression results are presented in Table 4 (right section). We find that in-task flexibility does not significantly influence the number of tasks submitted for workers who do not have daily income targets. On the other hand, workers who set daily income targets submitted significantly more tasks when high levels of flexibility are granted in the task, and they value in-task flexibility as much as \$0.92/hour.

DISCUSSION

In this paper, we experimentally examine the impact and value of flexibility in on-demand crowdwork. Our experimental results suggest that granting more in-task flexibility in on-demand crowdwork leads to dramatic changes in the temporal dynamics of worker behavior, significant improvement of work quantity and similar work quality. Furthermore, we have also quantitatively measured the economic value that workers attach to in-task flexibility and showed that workers equate the ability to control their own time within on-demand tasks with substantial financial compensation.

Does More Flexibility Mean Lower Speed?

A natural question one may ask is whether the provision of flexibility in on-demand crowdwork implies a decrease in task completion speed. In our study, while we do find that it takes individual workers a longer period of time to complete the same amount of work when more time is allotted in the task, on the aggregate level, we do not observe such a decrease as individuals also complete more tasks with more in-task flexibility. For example, in Study 1, the amount of time it takes workers to complete the first $N = 1,000$ (or 10,000) tasks is 28.2, 31.5 and 22.2 minutes (or 3.1, 2.8 or 2.9 hours) for the 1 minute, 1 hour, and 1 day treatments, respectively. These comparisons are robust over a wide range of N. It is plausible that

[4]Within each worker group, the average amount of time workers in (1 day, 4 cents) treatment dwell on a task is used as the reference for the time cost of a task.

this observation may result from the current on-demand crowdwork environment. Specifically, as a lot of tasks in on-demand platforms today provide limited in-task flexibility, workers are accustomed to work at a pace that is as fast as possible. Therefore, providing extra flexibility in the work only leads a fraction of workers to exercise control over their work time on a small number of tasks and thus does not immediately result in a decrease in overall task completion speed. Understanding the long-term effects of providing higher levels of flexibility in on-demand crowdwork, as well as thoroughly examining the possible trade-off between task flexibility and completion speed, is an important research direction for the future.

Limitations

In this research, we examined the impact of flexibility on worker behavior, work quantity and quality. Further research is needed to understand how in-task flexibility affects various aspects of worker's life (e.g., worker's perceived stress). In addition, our estimation of the value of flexibility depends on how sensitive workers are to changes in task price in the price range we studied. Arguably, humans respond to the change in financial incentives differently when the relative change in the magnitude of incentives differs. Examining the value of flexibility for other price ranges is another important, future direction.

Our examination of the impact and value of flexibility in on-demand crowdwork is conducted based on a particular type of task (i.e., batches of sentiment analysis tasks). While it is a popular type of task on on-demand platforms and is representative of many simple tasks that require human intuition and judgment (e.g., search query relevance, image annotation, etc.), it is unclear whether and how our results can be generalized to other types of tasks (e.g., survey tasks) that are structurally different and may be less compatible with granting workers more in-task flexibility.

We have observed that workers who spend more time working on on-demand platforms exercise more control of their work time when more in-task flexibility is granted, and they also value in-task flexibility more. It is known that some on-demand crowdworkers, especially those expert ones (who are likely to spend more time on on-demand platforms), leverage various scripts and tools to manage and schedule their work [14, 8]. The impact of in-task flexibility on on-demand crowdworkers may be partly mediated by worker's usage of scheduling tools. Further research is needed to examine this possible mediation.

Finally, this study focuses on examining in-task flexibility in on-demand crowdwork, and we leave the question of whether in-task flexibility and across-task flexibility are interchangeable, and their respective roles in impacting on-demand workers for future study.

Practical Implications for Requesters and Workers

Our study suggests that the promise of flexibility in on-demand crowdwork falls short in giving workers sufficient control over the timing of their work. Workers' willingness to put a premium on controlling the scheduling of their work is a clear signal of this. These findings have important implications for both requesters of labor and on-demand platform workers.

For requesters, as higher levels of in-task flexibility are associated with the same or better work outcomes and it is valued by workers, they should consider providing more such flexibility in their work whenever possible. Importantly, the lesson here is not, "we could give workers extra time and get them to work for less." Trying to lower market pricing by offering workers extra time could backfire. The most active workers, reliant on platform earnings, would need to work even harder to make up the pay, incenting them to boycott requesters who make them choose between taking bathroom breaks and higher wages. We urge requesters to introduce more flexibility in their work in two ways. First, carefully experiment with and gauge the "time allotted" parameters for tasks rather than simply leave them as the default. For a large number of tasks on MTurk, the requesters choose to set the time allotted to be the default value of 1 hour. Our experimental results indicate that putting a 1 minute time limit on a 30-second task is not optimal for encouraging higher levels of work quantity and quality. Similarly, blindly setting time allotted to 1 hour, even for tasks that cost significant amounts of time (e.g., 30 minutes or more) can be suboptimal as well. Second, delineate urgent tasks from tasks that could include workers control over their pace and task deadlines. Doing so would not only give workers more flexibility that they value but also produce better work outcomes in on-demand crowdwork.

From the workers' point of view, as we have mentioned earlier, not all workers exercised their freedom to control their work schedule when we granted flexibility in tasks. As we encourage requesters to consider increasing the temporal flexibility of on-demand crowdwork, it is also beneficial for workers to learn how they can best utilize such flexibility to both increase their efficiency and improve their working experience. In addition, workers can also consider to take actions collectively [19] to request for more flexibility in the work, or even a redesign of the interaction mechanism of on-demand platforms to give workers a role in negotiating time limits of the work.

CONCLUSION

Flexibility has long been assumed to be a key feature and benefit of on-demand crowdwork. This paper questions this common perception by examining whether there is sufficient flexibility in on-demand crowdwork and what it might mean to design in-task flexibility into on-demand work. Our study suggests that higher levels of flexibility can be afforded in on-demand crowdwork by providing workers more control of their work time within individual tasks. Through two randomized behavioral experiments, we find that granting more flexibility in tasks significantly influences the ways workers work and leads to higher work quantity and similar work quality, and workers also attach substantial value to the flexibility provided to them. Together, these results highlight the importance and benefits of allowing workers to control their own time in individual tasks in on-demand crowdwork.

ACKNOWLEDGEMENT

We thank all the workers who participated in our studies and the CHI 2018 anonymous reviewers for many helpful comments.

REFERENCES

1. Boris B. Baltes, Thomas E. Briggs, Joseph W. Huff, Julie A. Wright, and George A. Neuman. 1999. Flexible and Compressed Workweek Schedules: A Meta-Analysis of Their Effects on Work-Related Criteria. *Journal of Applied Psychology* 84, 4 (1999), 496–513.
2. Giovanni Costa. 2003. Shift work and occupational medicine: An overview. *Occupational Medicine* 53, 2 (2003), 83–88.
3. Xuefei Deng, KD Joshi, and Robert D. Galliers. 2016. The duality of empowerment and marginalization in microtask crowdsourcing: Giving voice to the less powerful through value sensitive design. *MIS Quarterly* 40, 2 (2016), 279–302.
4. Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, Panagiotis G. Ipeirotis, and Philippe Cudré-Mauroux. 2015. The dynamics of micro-task crowdsourcing: The case of Amazon MTurk. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 238–247.
5. Sabine AE Geurts and Evangelia Demerouti. 2003. Work/non-work interface: A review of theories and findings. *The Handbook of Work and Health Psychology* 2 (2003), 279–312.
6. Daniel G. Goldstein, Siddharth Suri, R Preston McAfee, Matthew Ekstrand-Abueg, and Fernando Diaz. 2014. The economic and cognitive costs of annoying display advertisements. *Journal of Marketing Research* 51, 6 (2014), 742–752.
7. Mary L. Gray and Siddharth Suri. 2018. *On-Demand: Crowds, Platform Economies, and the Future of Work in Precarious Times*. Forthcoming.
8. Mary L Gray, Siddharth Suri, Syed Shoaib Ali, and Deepti Kulkarni. 2016. The crowd is a collaborative network. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 134–147.
9. E. Jeffrey Hill, Alan J. Hawkins, Maria Ferris, and Michelle Weitzman. 2001. Finding an extra day a week: The positive influence of perceived job flexibility on work and family life balance. *Family Relations* 50, 1 (2001), 49–58.
10. Chien-Ju Ho, Aleksandrs Slivkins, Siddharth Suri, and Jennifer Wortman Vaughan. 2015. Incentivizing high quality crowdwork. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 419–429.
11. Ayush Jain, Akash Das Sarma, Aditya Parameswaran, and Jennifer Widom. 2017. Understanding workers, developing effective tasks, and enhancing marketplace dynamics: a study of a large crowdsourcing marketplace. *Proceedings of the VLDB Endowment* 10, 7 (2017), 829–840.
12. Kerry Joyce, Roman Pabayo, Julia A. Critchley, and Clare Bambra. 2010. Flexible working conditions and their effects on employee health and wellbeing. *The Cochrane Library* (2010).
13. Lawrence F Katz and Alan B Krueger. 2016. *The rise and nature of alternative work arrangements in the United States, 1995-2015*. Technical Report. National Bureau of Economic Research.
14. David Martin, Benjamin V Hanrahan, Jacki O'Neill, and Neha Gupta. 2014. Being a turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 224–235.
15. Julian McAuley and Jure Leskovec. 2013. Hidden factors and hidden topics: Understanding rating dimensions with review text. In *Proceedings of the 7th ACM Conference on Recommender Systems*. ACM, 165–172.
16. Phyllis Moen, Erin L. Kelly, Wen Fan, Shi-Rong Lee, David Almeida, Ellen Ernst Kossek, and Orfeu M. Buxton. 2016a. Does a flexibility/support organizational initiative improve high-tech employees' well-being? Evidence from the work, family, and health network. *American Sociological Review* 81, 1 (2016), 134–164.
17. Phyllis Moen, Erin L. Kelly, Shi-Rong Lee, J Michael Oakes, Wen Fan, Jeremy Bray, David Almeida, Leslie Hammer, David Hurtado, and Orfeu Buxton. 2016b. Can a Flexibility/Support Initiative Reduce Turnover Intentions and Exits? Results from the Work, Family, and Health Network. *Social Problems* (2016), spw033.
18. Hylco H. Nijp, Debby GJ Beckers, Sabine AE Geurts, Philip Tucker, and Michiel AJ Kompier. 2012. Systematic review on the association between employee worktime control and work-non-work balance, health and well-being, and job-related outcomes. *Scandinavian Journal of Work, Environment & Health* (2012), 299–313.
19. Niloufar Salehi, Lilly C Irani, Michael S Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, and others. 2015. We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1621–1630.
20. Trebor Scholz. 2016. *Uberworked and Underpaid: How Workers are Disrupting the Digital Economy*. John Wiley & Sons.
21. Kristen M. Shockley and Tammy D. Allen. 2007. When flexibility helps: Another look at the availability of flexible work arrangements and work–family conflict. *Journal of Vocational Behavior* 71, 3 (2007), 479–493.
22. Aaron Smith. 2016. *Gig Work, Online Selling and Home Sharing*. Technical Report. Pew Research Center.
23. Arun Sundararajan. 2016. *The Sharing Economy: The End of Employment and the Rise of Crowd-based Capitalism*. MIT Press.
24. Michael Toomim, Travis Kriplean, Claus Pörtner, and James Landay. 2011. Utility of human-computer interactions: Toward a science of preference measurement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2275–2284.