7-2018

Striving to earn more: A survey of work strategies and tool use among crowd workers

Toni KAPLAN
Carnegie Mellon University

Susumu SAITO
Waseda University

Kotaro HARA
Singapore Management University, kotarohara@smu.edu.sg

Jeffrey P. BIGHAM
Carnegie Mellon University

Citation

KAPLAN, Toni; SAITO, Susumu; HARA, Kotaro; and BIGHAM, Jeffrey P.. Striving to earn more: A survey of work strategies and tool use among crowd workers. (2018). *Proceedings of the 6th AAAI Conference on Human Computation and Crowdsourcing HCOMP 2018: Zürich, Switzerland, July 5-8*. 70-78.
Available at: https://ink.library.smu.edu.sg/sis_research/4333

Striving to Earn More: A Survey of Work Strategies and Tool Use Among Crowd Workers

Toni Kaplan,[1*] Susumu Saito,[2*] Kotaro Hara,[3] Jeffrey P. Bigham[1]
[1]Carnegie Mellon University, Pittsburgh, PA
[2]Waseda University, Tokyo, Japan
[3]Singapore Management University, Singapore
tkaplan@andrew.cmu.edu, susumu@pcl.cs.waseda.ac.jp, kotarohara@smu.edu.sg, jbigham@cs.cmu.edu

Abstract

Earning money is a primary motivation for workers on Amazon Mechanical Turk, but earning a good wage is difficult because work that pays well is not easily identified and can be time-consuming to find. We explored the strategies that both low- and high-earning workers use to find and complete tasks via a survey of 360 workers. Nearly all workers surveyed had earning money as their primary goal, and workers used many of the same tools (browser extensions and scripts) and strategies in an attempt to earn more money, regardless of earning level. However, high-earning workers used more tools, were more involved in worker communities, and more heavily used batch completion strategies. A natural next step is to use automated systems to assist workers with finding and completing tasks. Workers found this idea interesting, but expressed concerns about impact on the quality of their work and whether using automated tools to support them would violate platform rules. We conclude with ideas for future work in supporting workers to earn more and design considerations for such tools.

Introduction

Crowd work is an increasingly important component of the digital economy. It provides an opportunity for people to earn income by completing online tasks issued by task *requesters* via crowd work marketplaces. Types of tasks vary widely; common tasks include video and audio transcription, translation, image tagging, data retrieval, and usability testing of websites (Ipeirotis 2010; Difallah et al. 2015). We focus here on the Amazon Mechanical Turk (AMT) crowd work marketplace due to both its scale and its ubiquity in research and machine learning applications.

Much of the prior research examining crowdsourcing marketplaces from the workers' perspective emphasize low wages and an uneven distribution of power between workers and requesters. Crowd workers on AMT are not provided a fixed hourly wage. Instead, earnings are allotted based solely on *human intelligence tasks* (HITs) completed and approved by requesters in a piece rate. Low per-task rewards and unpaid task search time contribute to more than half of the AMT workers currently earning less than $5 an hour (Ipeirotis 2010; Irani and Silberman 2013; 2016; Hitlin 2016; Horton and Chilton 2010; Martin et al. 2014; Hara et al. 2018).

Prior research suggests that workers use online communities and external tools to aid their work (Mason and Suri 2012; Schmidt 2015; Huang and Bigham 2017). For example, to avoid unfair requesters, workers use tools like Turkopticon (Irani and Silberman 2013), Crowd Workers (Callison-Burch 2014), and online forums such as TurkerNation[1]. To reduce the unpaid work due to task search time, people employ strategies like "Preview and Accept" (known by workers as PandA), to accept a manually specified batch of similar HITs in parallel, assuring they have a constant stream of HITs to progress through during their work session. Researchers have also created prototype tools that automatically queues and visualizes available work (Hanrahan et al. 2015). Outside of the research community, crowd workers themselves have produced a plethora of tools to help augment the process of filtering HITs and automate the queuing process. However, there is limited research on what techniques and tools the workers are currently using, how it affects their income, and how strategies may differ between novice and experienced workers.

In this paper, we seek to better understand the challenges crowd workers face in wage-efficient task selection, and what strategies, tools and information high-earning workers are using to overcome these obstacles. We conducted a survey on AMT to explore how low- and high-earning workers are leveraging information about HITs to select tasks to complete, and to make inferences about where further research could be best focused to improve crowd workers' earnings. We examined the task-selection habits and types of external tools utilized by high-earning workers in comparison to their low-earning peers. By investigating these factors, we aim to provide informed design considerations for future tools and task-recommendation systems for improving crowd workers' earnings.

In summary, our contributions are:

- An overview of worker strategies and tool use, which may inform future research in teaching workers to be more effective and tools to help workers earn higher wages;

[*]These authors contributed equally to this work

[1]http://turkernation.com/

- An analysis of how high- and low-earning crowd workers different in working strategies, engagement with social communities, and tool usage; and,
- Design considerations for researchers and others developing tools to support crowd workers, especially tools that would bring to bear automated technology for recommending tasks for workers to do and help workers complete tasks more efficiently.

Related Work

The AMT marketplace demonstrates a severely uneven distribution of power and information between workers and requesters (Salehi et al. 2015; Irani and Silberman 2013). Requesters have the ability to reject submitted tasks. Workers are not compensated at all if work is rejected. Generally the mechanism is used by requesters to avoid compensating workers for poor quality or incomplete work. This is a point of contention, as requesters are able to keep data from uncompensated work. Previous explorations of ethics in crowd work have noted that workers feel this is an unfair practice, as requesters can independently and subjectively curate results (Martin et al. 2014). Researchers have made efforts to promote a more fair crowd work marketplace by addressing such unethical treatment of workers. Approaches ranged from working collaboratively with crowd workers to amass a collection of letters to Jeff Bezos (Salehi et al. 2015), the founder of AMT, to the creation of a separate crowdsourcing marketplace based on open-governance in which the workers needs and rights are prioritized (Gaikwad et al. 2015). While these efforts push for fair treatment of crowd workers, reasonable wages, and open communication, the collective letter has not had significant effects on the environment and the open platform is in a process of taking off.

The information imbalance between requesters and workers on AMT limits workers ability to effectively filter and search for HITs that will be completable and wage-efficient (Chilton et al. 2010). While requesters are able to judge workers by a number of metrics within AMT (*e.g.,* by qualifications, location, number of tasks completed), workers do not have access to similar information regarding requesters. The current AMT search interface allows workers to sort tasks by criteria like creation date and reward amount, but does not provide more advanced features like sorting and filtering available HITs by wage efficiency, level of difficulty, interests or other preference metrics (Silberman, Irani, and Ross 2010). The lack of such advanced search features limit workers' ability to judge the quality of requesters and the wage and feasibility their tasks. Tasks may be impossible to complete (*e.g.,* due to unclear instructions, interface glitches or insufficient time to complete a task). This forces a worker to abandon it or return it to the pool of available tasks, resulting in wasted and unpaid work time. This makes it difficult for workers to optimize wages, forcing them to balance per-task reward with completion time, while also attempting to minimize unpaid time spent searching for tasks.

Prior work has explored ways to overcome this information imbalance. Arguably, the most widely adopted tool is Turkopticon (Irani and Silberman 2013), a browser extension that enables crowd workers to collaboratively rate and review requesters. The ratings and reviews are publicly visible and can be used to avoid a specific requester if s/he has a poor reputation. The information from Turkopticon's API is integrated into tools designed by workers themselves. For example, a browser extension such as HitScraper allows its users to filter and prioritize search results based on Turkopticon ratings. Although we know these tools are widely used among crowd workers, to our knowledge, there has not been formal research investigating the types and prevalence of worker tools or their impact on workers' income. This leaves us with little information about how workers are currently addressing the challenges they face earning a viable wage in the workplace, providing little foundation from which to develop new tools. In this paper, we investigate the current state-of-the-art in worker tools and strategies on AMT to providing necessary grounding for future tool development.

In the research survey presented here, we examine the role of HIT content, search features, and tools in wage-efficient task selection. HIT-content based task selection on AMT is sparsely studied and rarely implemented. Previous task-recommendation systems have leveraged information readily available in the AMT search, such as task keywords, reward, qualifications, etc., in combination with Turkopticon ratings to queue wage-efficient tasks (Hanrahan et al. 2015; Alsayasneh et al. 2018). We hypothesize that efficient HIT selection hinges on additional content dependent factors that affect work duration, such as the types of media included in the HIT and the type and number of inputs required.

Method

We created and deployed a survey to gather information about AMT worker earnings and demographics, HIT selection criteria, work strategies and tools. The survey was created and hosted using Qualtrics[2], and 400 HITs including the survey were posted to Amazon Mechanical Turk for United States-based workers to complete. The survey contained 67 required questions and took between 10 and 30 minutes to complete. Participants were compensated \$3.50 upon completion to provide a mean hourly wage of \$10.

We staggered the release of HITs in order to sample workers with varying levels of crowd work experience, as follows: The first batch of 100 HITs was made available to workers with over 10,000 HITs completed. The following three batches of 100 HITs were made available to workers with more than 5000, 1000, and then 100 HITs completed. The survey was limited to workers in the United States, and was posted from January 23, 2018 to January 31, 2018.

The survey began with general demographic questions, including gender, age, employment status, education level and income. The following survey sections included questions on AMT related demographic information, such as time spent working and estimated earnings. Workers were asked if they had the Masters Qualification on AMT. A "Masters Qualification" is a qualification that is automatically granted to a selection of workers by AMT based on statistical models used to identify workers who "consistently

[2]https://www.qualtrics.com/

Table 1: Description of Mechanical Turk related browser extension tools (as of February 2018)

Extension name	Description
Turkopticon	A web platform (with API) for reviewing and evaluating requesters and HITs. Also refers to a browser extension that displays pop-ups of the evaluation status on AMT search pages.
Panda Crazy	A userscript that provides an interface for managing and PandA-ing batches of HITs.
MTurk Suite	An extension enhancing AMT pages with features from various scripts and extensions. Includes of Turkopticon, Turkerview, and minor work history and earnings tracking features.
HIT Scraper	A userscript that provides a an augmented search interface for HITs. Hit Scraper includes additonal search filters and can automatication search for new HITs at set intervals.
MTurk Engine	An extension combining HIT Scraper and Panda Crazy features, with an automatic HIT watcher and improved dashboard for managing earnings.
Turkmaster	A userscript that adds a side bar in Mechanical Turk dashboard page. Automatically runs a watcher for new HITs based saved requesters and search keywords. Also supports PandAing HITs.
Greasemonkey/ Tampermonkey	Extensions that enable userscripts. (Required for some userscripts, such as HIT Scraper, HITForker, Overwatch, Panda Crazy and Turkmaster)

demonstrate a high degree of success in performing a wide range of HITs across a large number of Requesters" [3]. We also asked if workers felt day of the week was a factor in earnings on AMT, and, if so, which days were the best and worst for earnings.

Workers were then asked 5-point Likert scale questions about what factors they consider when selecting HITs. These included rating the importance of HIT reward, HIT media type, predicted HIT completion time, and recommendations from other workers when selecting a HIT. We asked similar questions about reasons why a worker may choose to avoid or return a HIT, and why they may choose to end a work session. Workers were then asked about preferred task types, their usage of AMT related tools (see Table 1), and website forums (see Table 2). Questions regarding AMT tools and websites also included and "Other" option with a text field in which participants could provide additional details. We also asked workers to indicate how time-consuming and also how frustrating they found task search, spending time on returned HITs, and spending time on rejected HITs. Four additional questions were asked to gauge worker sentiment about the possible future of automation in crowd work. Each set of Likert scale responses were followed by optional open-ended fields in which workers could provide additional comments.

[3]https://www.mturk.com/worker/help

The survey closed with more specific questions regarding workers' experience and income on Mechanical Turk. Workers were asked to access their AMT dashboard and report the number of HITs approved/rejected/pending, their HIT approval rate, earnings from 2017 and total AMT earnings. These values are available in the AMT dashboard interface, and thus should be more reliable than self-reported estimated wages.

All 400 survey HITs were completed. Of these, 360 were kept for analysis. Forty responses were omitted due to violations of our spam filtering and validation criteria, which checked worker responses for non-zero total AMT earnings and internal consistency (*e.g.,* workers' reported approval rate should be consistent with the their reported approved HITs divided by reported total HITs submitted). Researchers then manually evaluated the optional open-ended responses to identify obvious spammers (e.g. random strings, repeated questions, consistently unrelated responses). All but two remaining persons completed at least one open-ended meaningful response ("no, none, and nope" were not considered meaningful responses). No additional spammers were identified. All 360 remaining responses reported a HIT acceptance rate within 1% of what would be expected based on their reported HIT submission history, and thus were deemed valid responses.

Results

In this section, we provide and discuss the results of the survey. We first describe high-level results such as the survey respondents' demographics, their income levels, and the tools they use. We then perform a more detailed analysis to uncover how and why workers selected particular tasks, the challenges they face, and tools they use. To investigate the effects of external tools and work strategies on workers' earnings, we split the workers into 2 groups based on their total reported 2017 earnings on AMT and compare between groups when relevant. We use total income as opposed to hourly wage as it is available in the AMT dashboard and therefore not prone to estimation errors among reliable respondents. We compute the median 2017 earnings ($948.18) among the workers who responded our survey, and assign them to the high-earning group if they earn more than the median, and low-earning group otherwise. This results in 180 respondents in each group.

We then define the top 10% of earners in our survey as high-earning extremes and further examine how their habits and strategies differ in comparison to the top 50% of workers. Via these additional comparisons, we aim to further elucidate successful work strategies.

Demographics. The composition of our survey respondents is similar to the worker demographics reflected in prior research (Ross et al. 2010). Women represented 47.8% of respondents, and the most common age group was 25-34, comprising 39.7% of respondents. More than half the respondents (61.7%) reported that they are employed full-time, and 50.2% reported having completed a four year degree or higher. Reported approximate household income (from all sources, including AMT) ranged from "Less than $10,000" to "Over $150,000." The median income bracket

Table 2: Description of Mechanical Turk related website forums (as of February 2018)

Website name	Description
MTurk Crowd (https://www.mturkcrowd.com/)	A community with forum topics such as sharing HIT links, requesters' reputation, scripts/extensions, and AMT news. There are "mentors" for novice workers. 1,130,000+ messages have been posted and 5,200+ members have joined.
Mturk Forum (http://www.mturkforum.com/)	A community with forum topics such as sharing HIT links, requesters' reputation, worker know-hows and habits. The largest platform among our choices; 1,650,000+ messages have been posted and 64,000+ members have joined.
Mturkgrind (http://www.mturkgrind.com)	A community with multiple forum topics such as sharing HIT links and other general discussions. Posts have slowed significantly in the past year. 1,100,000+ messages have been posted and 14,000+ members have joined.
[Reddit] Hits Worth Turking For (https://www.reddit.com/r/HITsWorthTurkingFor/)	A community with a single forum, for sharing good HIT links between workers. 42,000+ members have joined.
[Reddit] Hits NOT Worth Turking For (https://www.reddit.com/r/hNOTwtf/)	A community with a single forum, for warning other workers about bad HITs. 500+ members have joined.
[Reddit] Amazon Mechanical Turk (https://www.reddit.com/r/mturk/)	A community with a single forum, for general conversations/discussions (e.g., various comments on HITs, tips for better tasking, warnings for bad requesters, etc.) 26,000+ members have joined.
Turker Hub (https://turkerhub.com/)	A community with forum topics such as sharing HIT links, scripts/extensions, and wiki information. The newest among our choices; established in Nov. 2016. 559,000+ messages have been posted and 2,200+ members have joined.
Turker Nation (http://turkernation.com/)	A community with multiple forum topics such as sharing HIT links (by workers/requesters) and other general discussions. This forum has 640,000+ posts and 20,000+ members.
HIT Notifier (http://hitnotifier.com/)	Aggregates good HIT links posted on Turker Hub, MTurk Crowd, MTurk Forum, and HITs Worth Turking For and provides an audio notification when new recommended HITs appear.

was $40-49,000, and 3% of total respondents reported less than $10,000.

Income Tracking. Of 360 workers, 258 (71.7% of workers) reported that they think about their earnings per day. This was the most common measurement interval, followed by wages per week (35% of workers), and earnings per hour (17.2% of workers).

Reported Earnings. Self-reported hourly workers' earnings averaged $5.12 per hour ($SD = 3.23$) and ranged between $0.01 and $25 per hour. Seventeen percent of respondents (62 workers) reported earnings above the current United States federal minimum wage ($7.25 per hour). Note that given the above details on tracked earnings, hourly reported earning alone may not be an effective means of describing workers' earnings. Another measure of hourly earnings can be computed per respondent by dividing daily earnings by average hours worked per day, resulting in a calculated hourly wage. The average calculated hourly wage was $4.73 ($SD = 3.27$) and ranged between $0.01 and $26.67. Given average calculated hourly wage, 16.39% of workers reported earnings above the federal minimum hourly wage.

Self-reported daily workers' earnings averaged $17.3 ($SD = 16.84$) and ranged between $0.03 and $100 per day. The low daily earnings may be due to the low hours worked per day. Reported hours worked per day ranged between .5 to 15 hours ($SD = 2.41$), and averaged 3.8 hours per day.

These figures are slightly higher than those reported in previous research (Ross et al. 2010; Hara et al. 2018). We believe this is due to the staggered distribution of the survey based on the number of HITs a worker has had approved, which resulted in an increase of experienced worker respondents. In fact, individual Spearman non-parametric correlations indicate a positive correlation between experience ($r(360) = .39,\ p < .001$) and hourly earnings, as well as between experience and daily earnings ($r(360) = .58,\ p < .001$), suggesting that these figures are slightly inflated due to the sampling method that we employed.

Impact of Day of the Week. Eighty-nine percent (321/360) of respondents agreed that day of the week "Probably" or "Definitely" had an effect on their earnings and opportunities on AMT. Workers reported the most profitable day was Monday (31%), which was closely followed by Tuesday (29%). The least profitable days were Sunday (59%) and Saturday (34%).

While more low-earners found Sunday to be the least profitable day (60.57%), followed by Saturday (32%), equal amounts of high-earners found Saturday (45.71%) and Sunday (45.71%) unprofitable. Our survey data does not allow us to investigate *why* workers think they earn more early in the week. We suspect this is because requesters who are not active during weekends become more active early in the week, so there is a greater number and variety of HITs available to workers.

PandA Strategy. "Preview and Accept" (PandA) is a strategy to reduce unpaid work and task search time, in which workers automatically accept a worker-specified batch of similar HITs in parallel, assuring they have a constant stream of HITs to work through. PandA is a work strategy facilitated and augmented by a wide array of extensions and scripts. In total, 156 workers (43.3%) reported using the PandA strategy in their work.

A Chi-square test of independence comparing the frequency of PandA strategy use between the high and low-earning groups showed PandA was more prevalent among the high-earning group ($\chi^2(1) = 23.927,\ p < .0001$). 101 of 180 workers in the high-earning group reported using PandA, in comparison to only 55 workers in the low-earning group. Given the prevalence of PandA usage overall and among high-earners, we believe that this strategy is one of the most important factors in efficient work on AMT, and that support for this strategy should be a design consideration for future crowd worker tools.

Extension Usage. See Figure 1. 213 (59.2%) of respondents reported using extensions to aid their work on AMT. The number of extensions used ranged from 0 to 8 and averaged 2.2 ($SD = 2.24$). Among workers using at least one extension, the average number of extensions used was 3.75. The most commonly used extensions were Tampermonkey, Turkopticon and MTurk Suite. "Other" extensions included HITForker (12) , Turkerview (4), Overwatch (4), HIT Database (4) and Task Archive (4). Note that HITForker, HIT Database and Overwatch are Greasemonkey scripts. Four high-earning workers also reported using their own custom scripts.

High-earning workers were more likely to use scripts such as MTurk Engine and Tampermonkey. A Wilcoxon Rank-Sum Test indicated that high-earners used significantly more extensions, $Mdn = 3$, than low-earners, $Mdn = 0$ ($Z = 4.49,\ p < .0001$).

Social Platform Usage. See Figure 1. More than 60% of workers (222 respondents) reported at least occasionally posting or browsing in AMT related online social spaces. The most popular social platform among workers was the MTurk subreddit where 99 of the surveyed workers used the platform, followed by the HITsWorthTurkingFor subreddit with 80 users, MTurk Crowd with 77 users, and Turker Hub with 49 users. "Other" websites included Facebook groups (7) and the Turkopticon website (5).

MTurk Crowd was significantly more popular among high-earners ($Z = 2.44, p < .05$). Twenty-seven percent (48) of high-earners used MTurk Crowd, in comparison to 16.11% (29) of low-earners. Similarly, Turker Hub was more popular among high-earners, with 20% (36) high-earners using the site, while only 7.2% (13) of low-earners ($Z = 3.53, p < .001$) used Turker Hub.

Task Search: Time and Frustration. 30% of respondents indicated via 5-point Likert scale that finding HITs to complete was "4 - Very" or "5 - Extremely" time consuming. Results did not differ significantly between high- and low-earners ($Z = .30,\ p = .766$). Regarding frustration, 22% of participants (81) reported that task search was "4 - Very" or "5 - Extremely" frustrating.

Notably, the most important reason for both high and low-earning workers ending a work session was that workers "Can't find more HITs worth doing." Nearly half of participants (48%) indicated that this was a "5 - Extremely Important" motivation in ending a work session. In combination, these findings suggest that the search for HITs on AMT poses challenges for workers of all levels, and improvement to the task search and selection process could potentially improve earnings for all workers.

Rejected / Returned Tasks: Time and Frustration. 44% (161) of participants indicated via 5-point Likert scale that having to return a HIT was "4 - Very Time Consuming" or "5 - Extremely Time Consuming." Similarly, 58% (205) of participants indicated that having to return a HIT was "4 - Very Frustrating" or "5 - Extremely Frustrating."

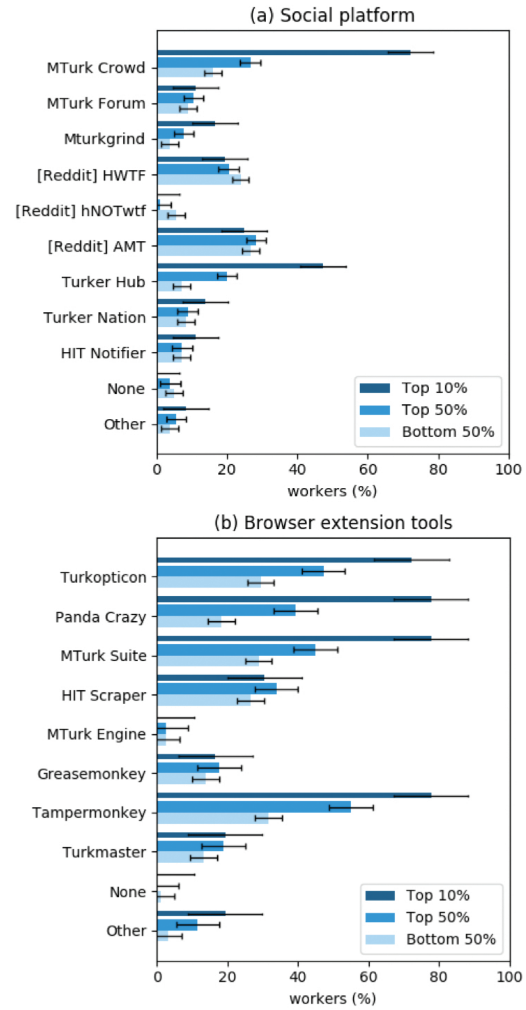

Figure 1: Workers used a number of browser extensions and social websites related to their work. High-earning workers were more likely to use extensions and used more extensions overall. High-earning workers also made heavier use of social web sites related to their work. Error bars represent standard error.

62% of workers found rejected HITs "Extremely Time Consuming" and 80% of workers indicated they rejected HITs are "Extremely Frustrating." This means that workers found that Rejected HITs were the most time consuming as well as the most frustrating.

There were no reliable differences between the high- and low-earning groups in level of frustration ($Z = 1.81,\ p = .0707$) or reported time consumption ($Z = .43,\ p = .6670$) for rejected tasks, nor were there any differences in frustration ($Z = -1.04,\ p = .2975$) or reported time consumption ($Z = -1.48,\ p = .1380$) for returned tasks.

Masters Qualification. 37 (10.28%) workers reported they had the Masters qualification. The majority, 28 (75.68%) of them were in the high-earning group. A chi-square test of independence was performed to examine the

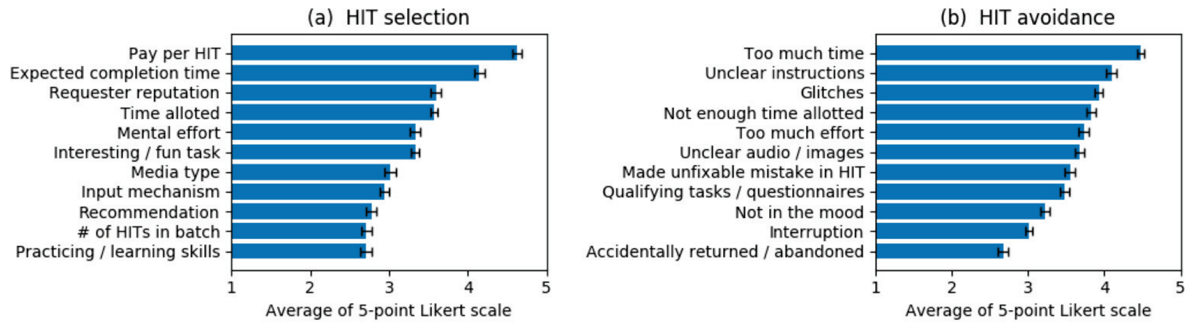

Figure 2: (a) HIT selection / (b) HIT avoidance criteria of all workers. While some of the features used to select or avoid HITs are readily available on the platform (*e.g.*, pay per HIT, Time allotted), others are only available with the use of extensions (*e.g.*, Requester reputation), and yet others require workers to guess (*e.g.*, expected completion time, unclear instructions). Error bars represent standard error.

relation between earnings group (high- *vs.* low-earning) and Masters Qualification status (with *vs.* without Masters Qualification). This was significant, ($\chi^2(1) = 10.87,\ p < .01$). High-earning workers were more likely to have the Masters Qualification than low-earners. This may be due to increased access to wage efficient tasks among those with Masters Qualification. Workers with Masters qualifications reported working an average of just under 2.5 years on AMT before achieving the qualification. This time period ranged between 1 and 5 years of work on AMT.

HIT Type Preference. The most popular HIT type was surveys and extended reading tasks, while the least popular was image transcriptions. High-earners had less extreme preferences overall across all HIT types, $M = 2.25$ on a 5-point Likert scale from 1-Not at All Preferred to 5-Extremely Preferred in comparison to the low-earning group, $M = 2.41$. A Wilcoxon Rank-Sum Test indicated low-earners were significantly more likely to prefer surveys ($Z = -4.08,\ p < .0001$) and image transcriptions ($Z = -2.93,\ p < .01$) in comparison to high-earners.

HIT Selection Criteria. Survey respondents indicated the importance of HIT selection criteria on 5-point Likert scales, ranging from 1 - Not at all Important to 5 - Extremely Important. See Figure 2(a). Results indicated that the most important HIT selection criteria was "Pay per HIT", followed by "Expected Task Completion Time" and then "Requester Reputation." Given importance of HIT pay and time per HIT in task selection, we can infer that workers are concerned with wage in addition to earnings.

The least important were "Opportunities to Learn New Skills" and the "Number of HITs Available in a Batch". The low importance reported for the number of HITs in a batch is surprising, given the prevalence of the PandA technique for quickly working through HITs in a batch. In addition, 54 unique respondents (35 in high-earning group and 19 in low-earning group) mentioned working on batches of HITs as part of their work strategy in the open-response questions. Given this, we believe that workers are working through batches of HITs, but generally batches are fairly abundant, and batch size is not something that workers must deliberately consider. Instead, in the open-ended responses, workers seemed more concerned about their personal opportunity to seize HITs in a good quality batch. One respondent clarifies, "I prefer to have something I can work on consistently for a long period of time more than anything, which I'm not sure is answered by any of the above options. It kind of matches "Number of HITs available in batch" but 10000 HITs can be taken in 10 minutes, whereas a batch of 200 might last an hour." Seven workers expressed sentiments about how task quality and requester reputation can take precedence over batch size, with users noting that "when trying a batch with a new requester, I will usually only do 5-10 hits at the most until they approve." Others mentioned previewing multiple HITs in the batch before accepting, only accepting batches from a requester they have worked with in the past, or accepting batches only from requesters with high T.O. ratings.

High-earners were significantly less likely to consider the type of media ($Z = -3.02,\ p < .01$), input mechanism ($Z = -2.62,\ p < .001$), opportunities to learn new skills ($Z = -2.57,\ p < .05$), or their interest in the task ($Z = -2.17,\ p < .05$) when selecting a HIT. This suggests that workers who are less selective about types of HITs to work on tend to earn more, though we cannot argue causal relationship between the two. That is, it is not clear if workers being less selective is enabling them to earn more, or if there is a hidden factor affecting worker selectivity and/or earnings.

HIT Avoidance / Return / Abandonment Criteria. Survey respondents indicated the importance of various factors in their decision to avoid, return or abandon (ARA) a HIT. Responses were via 5-point Likert scales, ranging from 1 - Not at all Important to 5 - Extremely Important (Figure 2(b)).

Results indicated that the most important ARA factor was that a task "Requires too much Time for the Pay", followed by "Unclear Instructions" and then "Glitches". The least important were "Accidents Resulting In Returns / Rejections" and "Interrupted Work".

Unclear instructions ($Z = -3.51,\ p < .001$), Unclear Audio / Images: ($Z = -3.21,\ p < .05$), Glitches ($Z =$

$-2.61,\ p < .05$) and Not Being in the Mood For this Type of Task ($Z = -2.21,\ p < .05$) were significantly more important ARA decision factors for low-earners than high-earners.

Automation. Workers were asked via Likert scale the extent to which they agreed or disagreed with statements about automation. Sentiment toward the use of automation differed between the high and low-earning groups ($Z = -2.09,\ p < .05$), with low-earning workers being more inclined to use a tool that automates some of the work in a HIT ($M = 4.17,\ SD = 1.18$), than high-earners ($M = 3.85,\ SD = 1.40$). In open-ended responses, 78 workers expressed various concerns regarding automation. The most common concern, mentioned by 18 respondents, was the role of the human in Human Intelligence Tasks (HITs). One worker concisely summarized that, "the whole purpose of a HIT is to complete a Human Intelligence Task, which by definition is a task that cannot or should not be automated." Other workers noted that they are being paid for their "opinions and thoughts" and, if they were using AI, they would "feel that it wasn't really [their] work." Seventeen workers expressed a lack of trust in the quality of AI output, worrying that they "wouldn't trust it to work correctly," noting that if they "don't trust it, it would add more time to go back and check to see if it was right." Twelve participants also mentioned that use of work automation tools would violate the AMT terms of service, and nine participants reported it would be a violation of their personal ethics. Twelve workers discussed how automation would be unfair to requesters. Workers specified that requesters post work on AMT expecting and valuing a human response, and using automation "doesn't feel right towards the requesters" because they "aren't trying to hire robots."

Workers were generally somewhat concerned ($M = 3.43$) about how automation could effect the availability of tasks on AMT, and this did not differ reliably between high- and low-earning workers ($Z = -1.62,\ p = .1037$). When asked about the possibility of automated systems completing the types of work currently on AMT, only 34% of respondents agreed that this would eventually be plausible in the future. In the open-ended questions workers emphasized that some tasks would always require human input, such as academic or opinion surveys and tasks involving evaluating art or music.

To gauge workers' awareness of their role in AI and machine learning, participants were asked if they felt that their work is being used to improve automated systems. The majority, 52% of respondents indicated that they did not think or did not know if that their work was being used to improve automated systems.

High-Earning Extremes. We define the top 10% of earners in our survey as high-earning extremes and further examine what habits and strategies these workers are using ("90-100%" of Figure 3.) The top 10% of workers was comprised of 36 people whose earnings ranged from $8,500 to $26,593 ($M = 13,030.29,\ SD = 4,818.12$). Their estimated hourly wage averaged $46.81 and varied between $20 and $100 ($SD = 23.27$).

The PandA work strategy was very common among the high-earning extremes, with 33 of the 36 (91.7%) high-earning extreme workers reporting using PandA. A Chi-square test of independence comparing the frequency of PandA between the high-earning and high-earning extreme group showed PandA was more prevalent among the high-earning extremes group ($\chi^2(1) = 22.63,\ p < .0001$).

Figure 3: Distribution of workers' total earnings in 2017 (split into 10 groups based on earnings.) We define the top 10%, indicated as "90-100%", as high-earning extremes.

The high-earning extremes were also more likely than high-earners to using browser scripts or extensions when working on AMT ($\chi^2(1) = 11.47,\ p < .001$), with 91.7% of high-earning extremes using scripts of extensions to augment their work experience. High-earning extremes also reported using a greater number of extensions ($M = 4.11,\ SD = 1.96$) than high-earning workers ($M = 2.42,\ SD = 2.42$) ($Z = 3.48,\ p < .001$).

The most popular extensions used among the high-earning extremes were Tampermonkey (77.78%), MTurk Suite (77.78%), Panda Crazy (77.78%) and Turkopticon (72.22%). The usage of Panda Crazy is significantly higher among high-earning extreme workers than high-earners ($Z = 5.21,\ p < .001$).

The importance of pay rate was evident in task selection based on the open ended questions. One respondent noted that they, "don't care what the task is, as long as it pays at least $12 an hour." Twenty of the 36 high-earning extremes included similar sentiments in their open-ended responses.

MTurk Crowd was significantly more popular among high-earning extremes. Over 70.22% (26) over the high-earning extreme workers used MTurk Crowd. This was significantly more than the 20% (36) of high-earners who used the site ($Z = 6.86,\ p < .0001$). Turker Hub was also more popular among high-earning extremes ($Z = 4.52,\ p < .0001$). Forty-seven percent (17) of high-earning extremes used Turker Hub, in comparison to only 16.11% (29) of high-earners who used the website.

Three workers mentioned how private qualifications affected their earnings. Qualifications on AMT can be assigned to an AMT worker based on demographics, number of HITs completed, qualification tasks (e.g. demonstration of language proficiency) or assigned to workers as needed by requesters (private qualifications). For a private qualification a requester might assign a custom qualification to a set of workers who completed part 1 of a study, or who had done quality work in the past. Then they may post new HITs restricted to only workers with that qualification. Given that requesters desire a specific subset of workers, these tasks generally pay higher than those without prior qualifications.

Open-ended responses among high-earning extreme workers also included multiple references to workers tracking their HITs and earnings history and their previous work per requester. In addition, four high-earning extreme workers mentioned a Greasemonkey / Tampermonkey script called MTurk HIT Database which provides this functionality. They were the only workers surveyed who mentioned this script. These responses may indicate the high-earner extreme workers are leveraging information about their previous work to inform current work selection patterns.

Discussion

Extensions and Tools. Our survey results indicate that extension and tool usage is prevalent among AMT workers. High-earning workers are generally using more extensions in their work, and they are using tools that facilitate the PandA strategy for queuing batch work. High-earners are also using these tools to monitor their previous work. Future tools may benefit from supporting similar batch work strategies and work tracking practices among crowd workers, especially if they can extend these successful strategies to low-earning workers who do not yet use these tools.

Since workers using extensions are, on average, using more than one extension to facilitate their work on AMT, extensions and tools should be designed to run in parallel with other scripts. If an extension conflicts with another popular tool, or provides redundant information that clutters the AMT interface, it will likely not be favored among workers.

Task Selection. Workers are frustrated with unpaid time spent working on HITs that are eventually returned or rejected by the requester. To explore why workers are experiencing HIT returns and rejections, we examined HIT avoidance, rejection and abandonment factors. The most important reasons for ARA can be broken down into two categories: poor compensation (for time and/or effort) and impossible HITs. HITs may be impossible to complete due to unclear instructions or media, glitches, or qualification tasks embedded within the HIT. Embedded qualification tasks may involve a pre-survey, allowing only those who meet certain demographic criteria to proceed with and be paid for the HIT. Those who do not meet the criteria are forced to return the HIT without receiving compensation for time spent on the embedded qualifying task. Presently, it is difficult for workers to determine whether a task is completable and will provide worthwhile compensation via the existing AMT interface without wasting time attempting HITs. Workers are currently using extensions to address this concern, using information from other workers to identify recommended tasks and make inferences about task completion time (and thus pay rate). Still, the results of this survey indicate the task selection strategies are not adequately reducing unpaid work time and frustration, and there is room for improvement of worker tools. Future systems should include a means for predicting a HITs completability and wage.

Sentiment Towards Automation. In our survey, we asked workers about the role they thought automation could play in their work. Workers expressed concerns about the ethics of using automation for partial task completion in a marketplace focused on "human" output. Workers noted that they would feel they were "cheating the requester" and that they may spend too much time checking over automated output to assure quality. From this feedback, we take away design concerns in creating automated systems for crowd workers. These systems should not make the worker feel as if they are being replaced or dishonest. We propose that providing auto-fill options for workers as they progress through a task, instead of providing automatically generated output upon page load, may help workers complete tasks quicker without compromising their output quality or minimizing their personal contribution. Automation should likely be oriented not toward the human intelligence part of the task, but rather to the mechanics of completing it.

Future Work

Microtask Recommendation. Our analysis shows that workers' earnings would likely benefit from access to a constant stream of tasks, as seen in the PandA technique, in which workers must manually identify batches to queue. Future work might therefore look to develop systems that automatically identify and recommend wage-efficient and completeable tasks to queue, reducing task searching time. We believe there is an opportunity for Machine Learning (ML) to reduce unpaid work time. For instance, an ML model might be trained to predict feasibility and completion time of HITs based on HIT content (HTML) and metadata. While previous approaches to AMT task recommendation (Hanrahan et al. 2015) exist, there are opportunities to utilize HIT content in addition to HIT metadata to make better predictions (and, thus, pay rate). Task feasibility may be able to be determined via web automation, enabling our system to identify and recommend worthwhile tasks. A recommendation system for HITs would reduce search time, unpaid work time, and frustration due to returned or rejected tasks. Later iterations of this automated task recommendation system may capture and leverage information about workers' personal work history to recommend similar and preferred tasks.

Masters Qualification. Another beneficial focus would be to help workers achieve the Masters Qualification, which significantly impacted the earning potential for workers. Unfortunately, Mechanical Turk does not clearly state the requirements to get Masters Qualification, although the AMT documentation notes that their statistical models consider the "variety of tasks" preformed. Perhaps, workers could be empowered by capturing and aggregating worker performance and task selection behavior, and then analyzing it to understand what leads to Masters Qualification attainment. Work selection trends identified here could then be incorporated into task recommendation.

Conclusion

In this paper, we explored the strategies that low- and high-earning workers use to find and complete tasks. Workers identified pay per HIT as their primary task selection factor, and used a variety of worker tools in an attempt to earn higher wages, regardless of their earning level. High-earning workers used more tools and were more involved in worker communities. High earners were also more likely to use batch completion strategies. Through our survey, rejected and returned HITs appeared as key factors in unnecessary unpaid work time and worker frustration.

These findings suggest several avenues of future research in optimizing task selection for improved wages and qualification achievement. Notably, automated task recommendation systems may benefit from collecting HIT content information that allows for automatic feasibility evaluation and work time predictions. Such measures would reduce unpaid work time and improve user access to wage-efficient HITs. Although workers were wary of using automation in their work in general, they seemed more open to using automation to improve efficiency in finding work and completing other tasks unrelated to the perceived core human intelligence task. We believe these augmentations are likely to improve the overall crowdwork experience, and lead to more workers achieving the higher wages that they seek.

References

Alsayasneh, M.; Amer-Yahia, S.; Gaussier, E.; Leroy, V.; Pilourdault, J.; Borromeo, R. M.; Toyama, M.; and Renders, J.-M. 2018. Personalized and diverse task composition in crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering* 30(1):128–141.

Callison-Burch, C. 2014. Crowd-workers: Aggregating information across turkers to help them find higher paying work. In *Second AAAI Conference on Human Computation and Crowdsourcing*.

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.

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*, 238–247. International World Wide Web Conferences Steering Committee.

Gaikwad, S. N.; Morina, D.; Nistala, R.; Agarwal, M.; Cossette, A.; Bhanu, R.; Savage, S.; Narwal, V.; Rajpal, K.; Regino, J.; et al. 2015. Daemo: A self-governed crowdsourcing marketplace. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, 101–102. ACM.

Hanrahan, B. V.; Willamowski, J. K.; Swaminathan, S.; and Martin, D. B. 2015. Turkbench: Rendering the market for turkers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1613–1616. ACM.

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 SIGCHI Conference on Human Factors in Computing Systems*, CHI '18. New York, NY, USA: ACM.

Hitlin, P. 2016. Research in the crowdsourcing age, a case study. *Pew Research Center. http://www. pewinternet. org/2016/07/11/research-in-the-crowdsourcing-age-a-casestudy*.

Horton, J. J., and Chilton, L. B. 2010. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on Electronic commerce*, 209–218. ACM.

Huang, T.-H. K., and Bigham, J. P. 2017. A 10-month-long deployment study of on-demand recruiting for low-latency crowdsourcing. In *In Proceedings of The fifth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2017). AAAI, AAAI*.

Ipeirotis, P. G. 2010. Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students* 17(2):16–21.

Irani, L. C., and Silberman, M. 2013. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 611–620. ACM.

Irani, L. C., and Silberman, M. 2016. Stories we tell about labor: Turkopticon and the trouble with design. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 4573–4586. ACM.

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*, 224–235. ACM.

Mason, W., and Suri, S. 2012. Conducting behavioral research on amazons mechanical turk. *Behavior research methods* 44(1):1–23.

Ross, J.; Irani, L.; Silberman, M.; Zaldivar, A.; and Tomlinson, B. 2010. Who are the crowdworkers?: shifting demographics in mechanical turk. In *CHI'10 extended abstracts on Human factors in computing systems*, 2863–2872. ACM.

Salehi, N.; Irani, L. C.; Bernstein, M. S.; Alkhatib, A.; Ogbe, E.; Milland, K.; et al. 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*, 1621–1630. ACM.

Schmidt, G. B. 2015. Fifty days an mturk worker: The social and motivational context for amazon mechanical turk workers. *Industrial and Organizational Psychology* 8(2):165–171.

Silberman, M.; Irani, L.; and Ross, J. 2010. Ethics and tactics of professional crowdwork. *XRDS: Crossroads, The ACM Magazine for Students* 17(2):39–43.