Understanding Crowdsourcing Requesters' Wage Setting Behaviors

Kotaro Hara
kotarohara@smu.edu.sg
Singapore Management University
Singapore

Yudai Tanaka
yudaitanaka@uchicago.edu
University of Chicago
Chicago, Illinois, USA

ABSTRACT

Requesters on crowdsourcing platforms like Amazon Mechanical Turk (AMT) compensate workers inadequately. One potential reason for the underpayment is that the AMT's requester interface provides limited information about estimated wages, preventing requesters from knowing if they are offering a fair piece-rate reward. To assess if presenting wage information affects requesters' reward setting behaviors, we conducted a controlled study with 63 participants. We had three levels for a between-subjects factor in a mixed design study, where we provided participants with: no wage information, wage point estimate, and wage distribution. Each participant had three stages of adjusting the reward and controlling the estimated wage. Our analysis with Bayesian growth curve modeling suggests that the estimated wage derived from the participant-set reward increased from \$2.56/h to \$2.69/h and \$2.33/h to \$2.74/h when we provided point estimate and distribution information respectively. The wage decreased from \$2.06/h to \$1.99/h in the control condition.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

ACM Reference Format:
Kotaro Hara and Yudai Tanaka. 2022. Understanding Crowdsourcing Requesters' Wage Setting Behaviors. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29-May 5, 2022, New Orleans, LA, USA*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3491101.3519660

1 INTRODUCTION

Crowd workers on Amazon Mechanical Turk (AMT) are underpaid. Prior work reported that workers' median hourly wage is approximately US\$2 per hour [10, 11]. The report suggested that the majority of requesters pay under US\$5 per hour, and those who post a large number of tasks tend to undercompensate. This is problematic as income generation is the primary motivation of crowd workers [1]. Oft-spoken narrative is that the wage stays low because requesters try to minimize the crowdsourcing cost by reducing piece-rate payment to introduce redundancy for quality control [16]. Although accurate, we believe this narrative is not the complete picture; we believe requesters unknowingly pay the low wage because they have limited information about the hourly wage when they set the piece-rate reward for the work they request. If this is the case, we could visualize the estimated wage of crowdsourcing tasks to better inform the requesters and nudge them to pay a fairer wage to the workers.

To test this hypothesis and explore the design space, we investigated if requesters' wage-setting behaviors change when we present the estimated hourly wage of performing crowdsourcing tasks through a controlled study. We designed three interfaces; one design that did not present wage information (*Control* interface), one design that presented a point estimate of the wage (*Point Estimate* interface), and another design that presented the estimated wage distribution (*Distribution* interface). We recruited 63 participants from a local university and asked them to perform three trials of wage-setting tasks. In the first trial, wage information was hidden in all conditions. In the second and third trials, the wage was presented to the participants in the treatment conditions. We adjusted the compensation for the study participants based on the rewards that the participants set in the trials, giving them incentive to earn more compensation.

In the treatment groups, the estimated wage for crowd workers increased from \$2.56/h to \$2.69/h (*Point Estimate*) and \$2.33/h to \$2.74/h (*Distribution*) indicating that the participants in the treatment groups set increased piece-rate reward after seeing the wage information. In contrast, the participants in the *Control* group slightly decreased the piece-rate reward, dropping the estimated worker hourly wage. Our study suggests that requester interface design could influence users' behaviors and nudge them to pay higher wages.

2 BACKGROUND

Research has shown that workers on micro-task crowdsourcing platforms like AMT are underpaid [1, 17, 20, 27, 28]. Workers typically earn a fraction of the U.S. minimum wage [13, 14, 17, 18, 28]—about \$2 per hour [10, 11]. This is problematic as the primary motivation of the workers is income generation [1, 2, 25, 28]. For paid crowdsourcing to strive as a viable workplace as described by Kittur *et al.* [24]—particularly for those who have been discriminated in the existing work environment [6, 9, 12, 33, 38]—while offering requesters labor that is highly flexible and productive, we need to better understand why requester underpay workers. One explanation of the low wage is requesters' self-interests in minimizing the cost of crowd work. Ipeirotis explained that due to the presence of spammers,

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04…\$15.00
https://doi.org/10.1145/3491101.3519660

requesters utilize an additional layer of quality control (*e.g.*, majority voting [32]) . This redundancy may increase the hourly rate of payment. To negate the effect of the increased cost, requesters decrease piece-rate reward. The low wage is permitted by the lack of platforms' minimum payment guidelines, with exceptions like Prolific. Although external parties recommend and facilitate paying fair wage [35, 37, 41], it is not clear if the practice is well adopted.

While the structure of the platform likely plays a large role in keeping the worker wage low, we also believe that the above narrative depicts requesters too egoistic, heartlessly taking advantage of workers because they can. In his essay, Sen argued that people do not make a judgement based solely on self-interest, but also on commitment and sympathy [36], challenging Edgeworth's conception *"every agent is actuated only by self-interest"* [5]. Thus, our question is *"are crowdsourcing requesters indeed driven to maximize their profit, or do they embrace irrationalities like commitment and sympathy to adjust the reward that they pay?"* People's irrationalities have been studied in prior crowdsourcing research [21, 29]. For example, Kaufmann *et al.* identified AMT workers work on micro tasks not only for extrinsic reward, but also for intrinsic motivation like having fun [21]. However, much work focused on crowd workers' extrinsic and intrinsic motivation. We know little about how requesters set rewards.

One explanation for the low payment is that the requesters do not know whether they compensate workers fairly. On AMT, for instance, if you create external HITs through API calls, you are responsible for calculating the hourly rate based on the piece-rate reward that you set and the work duration. Unless you are committed, you end up not knowing what hourly wage you are setting. In fact, Whiting *et al.* estimates that requesters underpay unknowingly for 68% of the time [41]. We thus speculate, if requesters don't know how much they are paying, would providing wage information affect their wage-setting behaviors? Our work draws design inspirations from the literature on persuasive technology and information visualization to support people's decision making. The existing research has shown that persuasive technologies could nudge people to perform prosocial activities [31]. For instance, Froehlich *et al.* designed eco-feedback visualizations guiding people to act in a more environmentally responsible manner [8]. Kay *et al.* designed a series of visualizations to inform when a bus is arriving, investigating whether different visualizations affect people's behaviors differently [7, 22]. Our goal is to investigate if presenting estimated worker wages via persuasive visualization can nudge requesters to behave in a more prosocial way that is not necessarily advantageous to their personal welfare.

3 HOURLY WAGE MODELING

In the controlled study described below, we want to present estimated wages to our participants who act as crowdsourcing requesters to see if showing them such information affects their wage-setting behaviors. The notion of fair wage is often discussed in terms of hourly wage (*e.g.,* minimum wage), but AMT employs a piece-rate reward mechanism—*i.e.,* a worker earns a preset amount of reward if they successfully complete the task regardless of how long the task took to complete. This required us to convert rewards into hourly wages by estimating task durations. In the controlled study, we could get the reward directly by recording what the participants set. However, we needed to estimate durations of micro-tasks so that we could present the estimated hourly wages to the participants.

Wage prediction is hard due to the variety of digital work and variability in task durations [4, 10, 15]. But prior work reports some success in duration and wage prediction (*e.g.,* [26, 34]). Learning from the prior work, we model task duration and hourly wage of image classification tasks—a simple and common task type on AMT [4, 10, 15]. We designed an interface in which we had crowd workers observe an image and answer yes/no questions—see Appendix.

We posted the image classification tasks on AMT. We recorded image classification time—the duration between the time that image was shown to the worker and the worker clicked "yes" or "no." N=80 workers independently performed N=2037 image classification tasks. We set the reward to US$0.50 per HIT. The number of image classification tasks that were bulked in one batch (*i.e.*, HIT) varied between HITs in the process of adjusting the hourly wage to pay US$7.25 per hour. The workers completed one image classification in 7.19 seconds on average (SD=34.18, median=2.47s). The distribution of task completion time was long-tail, which fit well with a log-normal distribution.

We fit the data to a log-normal distribution to obtain its parameters using SciPy's `stats.lognorm.fit()` method [19]. The location and scale parameters are 0.01 and 2.94, and $\mu = 1.08$ and $\sigma = 0.98$. The theoretical mean and theoretical median are 4.75s and 2.94s respectively[1]. The theoretical mean is underestimated compared to the empirical mean (7.19s) due to a few records of abnormally long durations. But theoretical and empirical medians are close to each other (2.94s and 2.47s), suggesting that the model represents the empirical data reasonably well. We thus use log-normal distribution to model the task completion time in the subsequent sections.

4 APPARATUS

We designed an interactive prototype of the requester interface that allowed its users to explore task information and set reward and tasks per assignment—*i.e.*, a number of image classification tasks bundled in one assignment (on AMT, a worker is compensated by successfully completing an assignment). The interface was developed using JavaScript, CSS, and Scala. At the top of the interface (Figure 1a), high-level description of the tasks was presented (*e.g.*, title, task type, task preview, project budget). At the bottom, the interface presented assignment-level information: total number of tasks to be completed, tasks per assignment, number of assignments to be posted, and reward per assignment (Figure 1b). In the study, we asked each study participant to adjust tasks per assignment and reward per assignment.

The bottom-right part of the interface had three design variations (Figure 1c and Figure 2). One design only presented the total cost and budget (*Control* interface; Figure 2a). The second design presented point estimates of assignment duration and hourly wage to communicate the information that may be important for setting reward (*Point Estimate* interface; Figure 2b). Inspired by prior work [22], the third design presented a distribution of the

[1] mean = $\exp(\mu + \sigma^2)/2$ and median = $\exp(\mu)$ for log-normal data

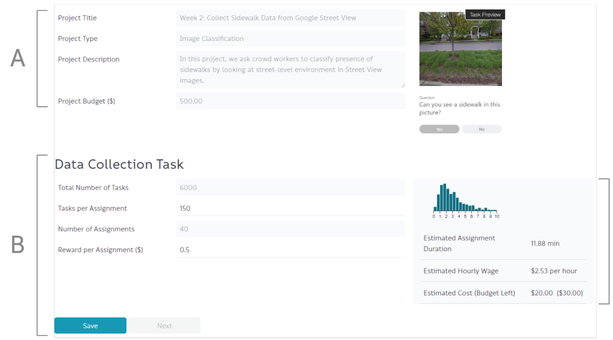

Figure 1: A prototype interface for posting tasks. (a) *Project Description Pane*: Title, type of crowdsourcing tasks, description, and project budget are presented. (b) *Task Information Pane*: The total number of tasks to be completed, a number of image classification tasks per assignment, number of assignments, and reward per assignment are shown here. The participants were asked to adjust two parameters in this pane: tasks per assignment and reward per assignment. (c) *Wage and Cost Pane*: Wage and cost information is presented in this pane. This part of the interface varies between study conditions.

Figure 2: Three different interface conditions: (a) *Control*: Only the total cost and remaining budget were presented. (b) *Point Estimate*: In addition to the budget, the point estimates of assignment duration and hourly wage were presented. (c) *Distribution*: In addition to the information presented in the Point Estimate interface, the distribution of the hourly wage was presented.

estimated hourly wage (*Distribution* interface; Figure 2c). The assignment duration and hourly wage information presented in the *Point Estimate* and *Distribution* interfaces were computed using the user-specified reward and number of tasks per assignment, as well as the log-normal model of task completion time that we described above. Estimating the assignment completion time was done as follows; we calculated the theoretical mean of the task duration (*i.e.*, 4.75s) and multiplied it by the number of tasks per assignment that the user-specified. The point estimate of the hourly wage was calculated by dividing the reward by the estimated assignment completion time. We plotted the hourly wage distribution on the fly by taking the following steps. From the log-normal task duration model, we sampled a number of task duration records specified in the number of tasks per assignment by the user. We summed all the sampled durations to compute the estimated assignment completion time. We divided the user-set per assignment reward by the assignment completion time to compute a single record of the hourly wage for the assignment. We repeated this process 1,000 times to collect a set of hourly wage records. The collected 1,000 hourly wage records were used to draw a histogram using `D3.js` (Figure 2c).

5 METHOD

To investigate requesters' wage setting behavior using different interfaces (*Control*, *Point Estimate*, and *Distribution*), we conducted a 3 (Interface conditions) $\times$ 3 (Trials) mixed design study. Interface condition was a between-subjects factor. Trial was a within-subjects factor; each participant used the interface and adjusted piece-rate reward for three times.

5.1 Participant Recruitment

We recruited participants by emailing local university students. We chose not to recruit experienced requesters to minimize the variability in crowdsourcing experience; we did not want people's prior experience in using crowdsourcing platforms to affect the result of our wage setting study. To exclude experienced ones but to include those who know about crowdsourcing, we asked our participant candidates to answer a pre-study questionnaire, which asked about their knowledge and experience related to crowdsourcing. If the respondents either requested or worked on crowdsourcing platforms, we excluded them from participating in the study. The questionnaire also collected basic demographic information (*e.g.*, age, gender). In recruiting the participants, we did not disclose that the study's intent was to observe if the participants set fair wages to not bias them.

5.2 Task Requesting Scenario and Compensation

We designed a scenario which resembled the situation when a requester posts work on AMT. As a requester, each participant was asked to make a series of decisions on how many image classification tasks to bundle in one work assignment and how much to pay for successfully completing an assignment. Our goal was to observe the evolution of participants' wage-setting behavior over trials. More specifically, our intention was to (i) see if the wage increase from *trial 1* to *trial 2* when we present wage estimates, and (ii) explore the change in participants' behaviors between *trial 2* and *trial 3* when the interface presented similar information (*e.g.*, do participants keep the wage stable?). We instructed the participants to imagine that the three trials occurred over three weeks so that the scenario resembles a realistic situation (*i.e.*, each week the participant had a quota of image classification tasks to post to AMT), but all three trials were done in one study session.

In each trial, we asked a participant to set a reward per assignment while satisfying two constraints: (i) assign 6,000 tasks per trial and (ii) the total cost per trial must be kept within \$50. We set the task number to 6,000, which roughly equals to eight hours of work (the theoretical mean task duration is 4.75s; 6,000 tasks per week x 4.75s = 7.92 hours per week). This was done so that we made sure the maximum work hour that our participants could set was below eight hours, abiding by the local work-related regulation, making sure that our participants' decisions not get influence by this rule. Learning from the research practice in behavioral economics [3, 40], participant compensation was determined based on the budget left after each trial. We employed this approach to observe how our participants balance personal welfare (*i.e.*, compensation they get) and fairness to workers. In addition to the base compensation of $5, participants got additional compensation of $1 for every imaginary $10 left in the budget. For example, if the participants were to maximize their compensation, they could pay workers close to $0 so that they get $20 ($5 base payment + $15 bonus from three trials) after three successive trials. We explained this compensation framework at the beginning of the study.

5.3 Experimental Design and Procedure

To examine how different interface design affects people's wage setting behaviors, we randomly split the participants into three groups and assigned them to one of the Interface conditions (*i.e.*, twenty-one people in each condition). At the beginning of a study session, we explained the task scenario and the procedure of the study. Prior to the first trial, we showed the interface with no hourly wage information (*i.e.*, equivalent to the *Control* interface) to each participant to explain the features of the user interface. We asked them to adjust the number of tasks per one assignment and per-assignment reward as a practice. After this practice, each participant advanced to perform a series of three wage setting trials. In the first trial, we did not present hourly wage information to participants in any conditions. In the second and third trials, the *Point Estimate* and *Distribution* interface displayed estimated assignment completion time and hourly wage. We collected quantitative data including the user set reward and number of tasks per one assignment in each trial, as well as how long each trial took to complete. We also asked the participants to think-aloud during the study session so that we could record their thought process in setting the reward. Upon finishing each session, we conducted a short unstructured interview asking about their experience.

6 RESULT AND ANALYSIS

Sixty-three people (35 female) participated in the study, whose age ranged between 19 and 28. The mean total compensation for each study condition was (*Control*, *Point Estimate*, *Distribution*) = ($16.0, $12.8, $13.6). For each of *Control*, *Point Estimate*, and *Distribution* conditions, tuples of mean estimated hourly wages (*trial 1*, *trial 2*, *trial 3*) were ($2.08, $1.87, $1.99), ($2.56, $3.07, $2.71), and ($2.33, $2.63, $2.75). This shows that mean wages increased in *Point Estimate* and *Distribution* conditions form *trial 1* to *trial 2*, while it decreases in *Control*.

An exploratory data analysis of hourly wage data with a Mauchly's test suggested that the data violated the sphericity assumption (W=0.85, p=.0097) and using ANOVA for analysis is inappropriate—a method oft-used for analyzing data from a factorial experiment. Instead, we decided to use Bayesian Growth Curve Model (BGCM) instead. BGCM also allowed us to observe the posterior distributions of estimated parameters (*e.g.*, change in hourly wage), allowing us to draw more insights from data [23]. We followed Oravecz *et al.*'s approach [30] to fit wage estimates to BGCM. See more in Appendix.

6.1 Wage Evolution

The observed wages increased from *trial 1* to *trial 2* in the *Point Estimate* condition (from $2.56/h to $3.07/h) and *Distribution* conditions

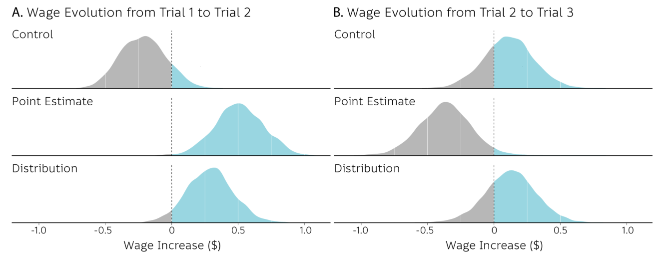

Figure 3: Posterior probability distributions of wage gradients from *trial 1* to *trial 2* and *trial 2* to trial 3.

($2.33/h to $2.63/h), while they decreased in the *Control* condition ($2.06/h to $1.86/h). From *trial 2* to *trial 3*, wage decreased in the *Point Estimate* condition ($3.06/h to $2.69/h). But changes in the wages in the *Control* and *Distribution* conditions were not remarkable (*Control*: $1.87/h to $1.99/h; *Distribution*: $2.62/h to $2.74/h).

Increase in the wage from *trial 1* to *trial 2* in the *Point Estimate* and *Distribution* conditions were significant. Figure 3 shows the posterior distributions of the wage slope between *trial 1* and *trial 2*, as well as *trial 2* and *trial 3*. Figure 3a shows that 99% of the area under the posterior distribution curve is above 0 for *Point Estimate* and 95% of the area is above 0 for *Distribution*. In the *Control* condition, 87% of the area is below 0. Thus, *people are likely to increase hourly wage when they are presented with estimated wage information* (whether it is a point estimate or a distribution) and *people tend to decrease the wage if we do not present the hourly wage information.*

While the change in the wage between *trial 2* and *trial 3* are small in the *Control* and *Distribution* conditions, people tend to decrease the wage in the *Point Estimate* condition. Figure 3b shows that 73% of the area is above 0 for the *Control* condition. Likewise, 74% of the area under the curve is above 0 for the *Distribution* condition, suggesting the absence of significant changes. In the *Point Estimate* condition, 98% of the area under the posterior distribution curve is below 0, indicating that people tend to drop the wage from *trial 2* to *trial 3* after increasing the wage.

7 DISCUSSION

Our result suggests that presenting wage information indeed affects how people set crowdsourcing rewards, and not showing wage information decreases the reward. The result depicts our participants' negotiation between commitment to pay fair wages to workers and maximizing their profit in the presence of the wage information. This pro-worker behavior is also seen in the quotes of the participants. Of the forty-two participants who were presented with the hourly wage information, fourteen participants mentioned that the hourly wage information was useful in deciding the reward for workers. One participant said *"I felt like I had done something really wrong when I saw the amount of money I had paid to workers in week 1"* (P6, *Point Estimate*). Another participant said *"with the estimated hourly wage, I could easily make the rational choice of how much I pay workers"* (P13, *Distribution*).

Our result supports Sen's view on sympathy and commitment affecting people's economic choices, and also aligns with the theory of behavioral economics. For example, Thaler and Sunstein noted that good information and prompt feedback are key factors that enable people to make good decisions [39].

However, the wage feedback may not be sufficient to keep the wage high. In the third trial, the participants in the *Point Estimate* condition decreased the wage. Some participants likely made this decision to maximize their compensation because they increased the wage a lot in *trial 2*. This makes us think, *"would the effect of presenting wage information last in long term?"* Because our study was a controlled, single-session study, we could not answer this question. Also, it is not particularly clear why we observed the drop in the hourly wage only in the *Point Estimate* condition, but not in the *Distribution* condition. We speculate that the *Distribution* interface helped the participants to make a more realistic decision that balanced their compensation and wage (thus not increasing the wage as much as the participants in the *Point Estimate* condition from *trial 1* to *trial 2*).

Although the participants indeed set higher wages in the treatment conditions, the expected wage only went up to around \$3. This may partly be the result of our study setting that introduced the ceiling of hourly wage (which was necessary to study how the participants negotiate the wage and their compensation). But it also suggests that people are not willing to spend much to compensate the workers if they are limited with their budget. This leads to a question, *"would participants pay higher wage if their budgets were ampler?"* Thus, future work could investigate the relationship between the requester budget and the hourly wage. We also suggest future work to investigate if anchoring techniques could be useful to nudge participants to set higher rewards. For example, showing a message like *"people usually pay \$5/h for this type of task"* may have nudged our participants to pay more (we intentionally chose not to show such messages to suppress the confounding effect).

Interestingly, two participants in the *Distribution* condition mentioned that they adjusted the number of tasks per worker based on the wage distribution. Although we did not explain this to our participants, when the number of tasks per worker is increased, the distribution becomes tighter; a bigger number of samples for task completion time records yield tighter assignment completion time distribution due to the central limit theorem. This in turn makes the expected hourly wage distribution tighter. One participant said *"I adjusted the values so that the distribution became keen because that means more workers are equally treated well"* (P21, *Distribution*). Future work could investigate how participants behave if we explicitly mention this behavior of the hourly wage distribution.

8 LIMITATIONS

Our task completion time and wage models are focused on the image classification task. It is not clear if we can model other types of tasks well. But this has a small impact on the core result of the paper that shows presenting hourly wage information changes how people set crowdsourcing rewards. Our interfaces did not present anticipated quality of work, which prevented the participants from balancing the worker reward and the expected data quality. But the current interface prototype is similar to the existing interface design of AMT. Our study is a lab study, and so its external validity is weaker than alternative study methods like a real-life field study with a deployed system. But this was done to control multiple factors (*e.g.*, budget, participant demographics). Our study was conducted with student participants with no experience as AMT requesters. Thus, we do not claim that our findings extend to how experienced requesters would behave. Our study was not meant to examine how people's behaviors change over a long period of time like months. A longitudinal field study is a natural extension of the current work.

9 CONCLUSION

We evaluated whether presenting the wage information affects requesters' reward setting behaviors. We conducted a controlled study with 63 participants recruited from a local university. In the study, we presented hourly wage information to participants in two treatment groups using two interfaces that we designed—*Point Estimate* and *Distribution*. We evaluated how participants' wage setting behavior changed over trials using Bayesian Growth Curve Model. We observed that hourly wage increased from \$2.56/h to \$2.69/h and \$2.33/h to \$2.74/h when the participants used *Point Estimate* interface and *Distribution* interface, while the wage dropped from \$2.06/h to \$1.99/h when we did not show expected hourly wage information in the *Control* condition.

ACKNOWLEDGMENTS

This research / project is supported by the Ministry of Education, Singapore under its Academic Research Fund Tier 2 (Project ID: T2EP20220-0016). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Ministry of Education, Singapore.

REFERENCES

[1] Janine Berg. 2016. Income security in the on-demand economy: findings and policy lessons from a survey of crowdworkers. (2016).
[2] Robin Brewer, Meredith Ringel Morris, and Anne Marie Piper. 2016. "Why Would Anybody Do This?": Understanding Older Adults' Motivations and Challenges in Crowd Work. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2246–2257. https://doi.org/10.1145/2858036.2858198
[3] Douglas D Davis and Charles A Holt. 1993. *Experimental economics*. Princeton university press.
[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 (WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 238–247. https://doi.org/10.1145/2736277.2741685
[5] Francis Ysidro Edgeworth. 1881. *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. Vol. 10. Kegan Paul.
[6] Diana Farrell and Fiona Greig. 2016. Paychecks, paydays, and the online platform economy: Big data on income volatility. *JP Morgan Chase Institute* (2016).
[7] Michael Fernandes, Logan Walls, Sean Munson, Jessica Hullman, and Matthew Kay. 2018. Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 144:1—-144:12. https://doi.org/10.1145/3173574.3173718
[8] Jon Froehlich, Leah Findlater, Marilyn Ostergren, Solai Ramanathan, Josh Peterson, Inness Wragg, Eric Larson, Fabia Fu, Mazhengmin Bai, Shwetak Patel, and James A Landay. 2012. The Design and Evaluation of Prototype Eco-feedback Displays for Fixture-level Water Usage Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2367–2376. https://doi.org/10.1145/2207676.2208397
[9] Mark Graham, Isis Hjorth, and Vili Lehdonvirta. 2017. Digital labour and development: impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European Review of Labour and Research* 23, 2 (mar 2017), 135–162. https://doi.org/10.1177/1024258916687250

[10] K. Hara, A. Adams, K. Milland, S. Savage, C. Callison-Burch, and J.P. Bigham. 2018. A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2018-April. https://doi.org/10.1145/3173574.3174023

[11] Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Benjamin V Hanrahan, Jeffrey. P. Bigham, and Chris Callison-Burch. 2019. Worker Demographics and Earnings on Amazon Mechanical Turk: An Exploratory Analysis. In *CHI Late Breaking Work*. ACM. https://doi.org/10.1145/3290607.3312970

[12] Kotaro Hara and Jeffrey. P. J.P. Bigham. 2017. Introducing People with ASD to Crowd Work. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2017*. https://doi.org/10.1145/3132525.3132544

[13] Paul Hitlin. 2016. Research in the crowdsourcing age, a case study. *Pew Research Center* (2016).

[14] John Joseph Horton and Lydia B Chilton. 2010. The Labor Economics of Paid Crowdsourcing. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC '10)*. ACM, New York, NY, USA, 209–218. https://doi.org/10.1145/1807342.1807376

[15] Panagiotis G Ipeirotis. 2010. Analyzing the Amazon Mechanical Turk Marketplace. *XRDS* 17, 2 (dec 2010), 16–21. https://doi.org/10.1145/1869086.1869094

[16] Panagiotis G Ipeirotis. 2010. Mechanical Turk, low wages, and the market for lemons. *A Computer Scientist in a Business School* 27 (2010).

[17] Lilly C Irani and M Six Silberman. 2013. Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 611–620. https://doi.org/10.1145/2470654.2470742

[18] Lilly C Irani and M Six Silberman. 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 (CHI '16)*. ACM, New York, NY, USA, 4573–4586. https://doi.org/10.1145/2858036.2858592

[19] Eric Jones, Travis Oliphant, Pearu Peterson, and Others. [n.d.]. {SciPy}: Open source scientific tools for {Python}. http://www.scipy.org/

[20] Miranda Katz. 2017. Amazon Mechanical Turk Workers Have Had Enough.

[21] Nicolas Kaufmann, Thimo Schulze, and Daniel Veit. 2011. More than fun and money. worker motivation in crowdsourcing–a study on mechanical turk. In *Proc. of AMCIS 2011*.

[22] Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. 2016. When (Ish) is My Bus?: User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5092–5103. https://doi.org/10.1145/2858036.2858558

[23] Matthew Kay, Gregory L Nelson, and Eric B Hekler. 2016. Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4521–4532. https://doi.org/10.1145/2858036.2858465

[24] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The Future of Crowd Work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 1301–1318. https://doi.org/10.1145/2441776.2441923

[25] Siou Chew Kuek, Cecilia Paradi-Guilford, Toks Fayomi, Saori Imaizumi, Panos Ipeirotis, Patricia Pina, Manpreet Singh, and Others. 2015. *The global opportunity in online outsourcing*. Technical Report. The World Bank.

[26] Anna Lioznova, Alexey Drutsa, Vladimir Kukushkin, and Anastasia Bezzubtseva. 2020. Prediction of Hourly Earnings and Completion Time on a Crowdsourcing Platform. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Virtual Event, CA, USA) *(KDD '20)*. Association for Computing Machinery, New York, NY, USA, 3172–3182. https://doi.org/10.1145/3394486.3403369

[27] Philippe Marcadent, Janine Berg, Mariya Aleksynska, Valerio De Stefano, Martine Humblet, Christina Behrendt, Susan Hayter, Christopher Land-Kazlauskas, Angelika Muller, Niall O'Higgins, Friederike Eberlein, Simone Marino, Bao Chau Le, and Calum Carson. 2016. Non-Standard Employment Around The World. (2016).

[28] 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 and Social Computing (CSCW '14)*. ACM, New York, NY, USA, 224–235. https://doi.org/10.1145/2531602.2531663

[29] Winter Mason and Duncan J Watts. 2009. Financial Incentives and the "Performance of Crowds". In *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '09)*. ACM, New York, NY, USA, 77–85. https://doi.org/10.1145/1600150.1600175

[30] Zita Oravecz and Chelsea Muth. 2018. Fitting growth curve models in the Bayesian framework. *Psychonomic Bulletin & Review* 25, 1 (2018), 235–255. https://doi.org/10.3758/s13423-017-1281-0

[31] Jenny Preece, Yvonne Rogers, and Helen Sharp. 2001. *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, Inc., New York, NY, USA.

[32] Alexander J Quinn and Benjamin B Bederson. 2011. Human Computation: A Survey and Taxonomy of a Growing Field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1403–1412. https://doi.org/10.1145/1978942.1979148

[33] Lucas Rosenblatt, Patrick Carrington, Kotaro Hara, and Jeffrey P Bigham. 2018. Vocal Programming for People with Upper-Body Motor Impairments. In *Proceedings of the Internet of Accessible Things (W4A '18)*. ACM, New York, NY, USA, 30:1—-30:10. https://doi.org/10.1145/3192714.3192821

[34] Susumu Saito, Chun-Wei Chiang, Saiph Savage, Teppei Nakano, Tetsunori Kobayashi, and Jeffrey Bigham. 2019. TurkScanner: Predicting the Hourly Wage of Microtasks. *CoRR* abs/1903.0 (2019).

[35] Niloufar Salehi, Lilly C Irani, Michael S Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, and Clickhappier. 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 (CHI '15)*. ACM, New York, NY, USA, 1621–1630. https://doi.org/10.1145/2702123.2702508

[36] Amartya K Sen. 1977. Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy & Public Affairs* 6, 4 (1977), 317–344. http://www.jstor.org/stable/2264946

[37] M S Silberman, B Tomlinson, R LaPlante, J Ross, L Irani, and A Zaldivar. 2018. Responsible Research with Crowds: Pay Crowdworkers at Least Minimum Wage. *Commun. ACM* 61, 3 (feb 2018), 39–41. https://doi.org/10.1145/3180492

[38] Saiganesh Swaminathan, Kotaro Hara, J.P. Jeffrey P Bigham, Hara Kotaro Swaminathan Saiganesh, and J.P. Jeffrey P Bigham. 2017. The Crowd Work Accessibility Problem. In *Proceedings of the Web for All Conference (W4A '17)*. ACM, New York, NY, USA. https://doi.org/10.1145/3058555.3058569

[39] Richard H Thaler and Cass R Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

[40] The Writing Group for CERTAIN-CHOICES, Vlad V Simianu, Margaret A Grounds, Susan L Joslyn, Jared E LeClerc, Anne P Ehlers, Nidhi Agrawal, Rafael Alfonso-Cristancho, Abraham D Flaxman, and David R Flum. 2016. Understanding clinical and non-clinical decisions under uncertainty: a scenario-based survey. *BMC Medical Informatics and Decision Making* 16, 1 (dec 2016), 153. https://doi.org/10.1186/s12911-016-0391-3

[41] Mark E Whiting, Grant Hugh, and Michael S Bernstein. 2019. Fair Work: Crowd Work Minimum Wage with One Line of Code. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. 197–206.