How Much Do Platform Workers Value Reviews? An Experimental Method

David Holtz
Haas School of Business, University of California, Berkeley and MIT Initiative on the Digital Economy
Berkeley, California, USA
dholtz@haas.berkeley.edu

Liane Scult
Microsoft Corporation
Redmond, Washington, USA
lidona@microsoft.com

Siddharth Suri
Microsoft Research
Redmond, Washington, USA
suri@microsoft.com

ABSTRACT

Previous qualitative work has documented that platform workers place an immense importance on their reputation due to the use of algorithmic management by online labor platforms. We provide a general experimental method, which can be used across platforms and time, for numerically quantifying the intensity with which platform workers experience reputation system-based algorithmic management. Our method works via an experiment where workers choose between a monetary bonus or a positive review. We demonstrate this method by measuring the value that freelancers assigned to positive feedback on Upwork in June 2020. The median freelancer in our sample valued a single positive review at ∼$49 USD. We also find that less experienced freelancers valued a positive review more highly than those with more experience. Qualitative data collected during the experiment indicates that many freelancers considered issues related to reputation system-based algorithmic management while choosing between the monetary reward and the positive review.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; • **General and reference** → **Experimentation**; • **Information systems** → **Reputation systems**; **Reputation systems**.

KEYWORDS

crowdsourcing, behavioral experiment, algorithmic management, upwork, online labor market, reputation

ACM Reference Format:
David Holtz, Liane Scult, and Siddharth Suri. 2022. How Much Do Platform Workers Value Reviews? An Experimental Method. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3491102.3501900

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

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3501900

1 INTRODUCTION

The last twenty years has seen the rise of online labor platforms focusing on types of work as varied as microtasks (e.g. MTurk), transportation (e.g. Uber and Lyft), food delivery (e.g. DoorDash), and freelance work (e.g. Upwork and Freelancer.com). These platforms generally operate as two-sided markets, in which the platform manages the relationship between the requester, who creates the task/job, and the worker, who completes the task/job.

Given the distributed nature of these markets, and in order to reduce costs, many online platforms engage in "algorithmic management" [21, 33], i.e., the practice of controlling the behavior of workers through the strategic use of information symmetry, reputation systems, behavioral nudges, and opaque ranking and matching algorithms [15].[1] In recent years, a fast-growing and substantial literature has emerged that focuses on the issues associated with algorithmic management [21, 28, 33, 37].

Online reputation systems are a key component of the algorithmic management systems currently used by many online labor platforms to exert control over their workforce. Ratings and reviews left by requesters are aggregated into reputation scores. This reputational information is often displayed to would-be requesters, used as an input by platform search ranking algorithms, and/or used as an input by the algorithms that match workers to requesters and tasks. All of these mechanisms have an effect on the way that tasks are distributed across the population of workers, and consequently impact how much a worker can earn. The flow of feedback data from previous clients into user interfaces, search algorithms, and ranking algorithms, which in turn affects which freelancers get matched to which clients (if any at all) is perhaps one of the most salient types of algorithmic management.

There are a number of qualitative studies that specifically report on the importance platform laborers place on managing their reputations. For example, in the context of Upwork, clients (who create the tasks) provide feedback in the form of ratings and reviews to the freelancers (who complete the tasks) they hired. This feedback information is not only displayed to would-be future clients, but is likely also used as an input to Upwork's search, ranking, and

[1] Jarrahi et al. [19] have recently advanced the concept of "platformic management," which is distinct from algorithmic management and argues that platform algorithms are just one part of an overall managerial structure that includes a platform's features, policies, and norms of use. We find this concept compelling, however, refer to algorithmic management through the paper as it is, at the time of writing, a more commonly cited concept. Although reputation systems are typically not themselves algorithmic in nature, much of the literature on algorithmic management considers online reputation systems a tool for algorithmic management.

matching algorithms [19]. Sutherland et al. [34] show that maintaining a strong reputation in the face of the platform's opaque reputation systems is a key source of precarity among Upwork freelancers. In the context of ridesharing platforms such as Uber and Lyft, passengers provide ratings of their drivers, which the platform then uses to assign future rides to drivers. However, drivers are left to guess how to obtain and maintain a high-quality reputation, because there is no consistent basis by which riders are evaluating their performance [33]. Griesbach et al. [15] reports that food delivery workers on both DoorDash and Postmates were troubled by platform policies under which they would be deactivated or removed from the platforms if their reputation score fell below certain thresholds. Across of all these platforms, a worker's reputation has a direct impact on whether they can work, with whom they're matched, and on their future earnings.

Although existing research has established that online labor market workers place a great deal of importance on online reputation, the fact that almost all of this work is qualitative makes it difficult to understand how the importance of online reputation to workers varies from platform to platform, or how its importance evolves over time in response to changes to online market conditions, offline market conditions, and/or platform design. Our key contribution in this work is building on and extending the research on algorithmic management, and in particular, the research on algorithmic management through online reputation systems, by developing a method that can produce numeric measurements of the value platform laborers assign to online reputation. This valuation is a measure of the intensity with which workers experience reputation system-based algorithmic control. As an illustration of the proposed method, in this paper we measure the intensity with which workers experience reputation system-based algorithmic management on a single platform at a single moment in time. However, obtaining similar numeric measurements on a consistent basis across different platforms would allow researchers to compare how the intensity with which platform workers experience reputation system-based algorithmic management varies both across platforms and across time. Doing so would make it possible to identify which platform features, market conditions, and/or external factors mitigate or exacerbate the extent to which platform workers experience reputation system-based algorithmic control.

Willingness to accept (WTA) is defined as the monetary compensation that would be required in order for a person to forfeit a particular item. Our measure of how much workers value online reputation is the WTA that they assign to a single positive review. In order to estimate workers' WTA for a positive review, we ask: how much, in terms of US dollars, do workers on the platform need to be paid in order to forgo a single 5-star rating with positive text? To answer this question, we conduct a behavioral experiment in which, after completing a recruitment task, platform workers are given the non-hypothetical choice between a positive review and a monetary reward randomly chosen from a predetermined set of values. Analysis of the resulting data allows us to trace out the demand curve and estimate the median valuation for a single positive review.[2] In contrast to qualitative methods, this approach can be used to compare cross-platform and longitudinal estimates of the intensity with which platform workers experience reputation system-based algorithmic management.

As a second contribution, we demonstrate our method by measuring the value of reputation on Upwork, an online labor market for macrotasks, as of June 2020. We analyze data from a sample of 520 Upwork freelancers who were given the non-hypothetical choice between a 5-star rating with a positive textual review, or a monetary bonus that was randomly chosen from a set of values ranging from $25 USD to $175 USD. We estimate that the median valuation of a 5-star rating and review among Upwork freelancers in our sample, which is a stratified sample consisting of customer support workers (average hourly rate = $12.70 USD) and graphic designers (average hourly rate = $23.30 USD), is $49 USD. Surprisingly, even when offered a $175 USD bonus, 32% of freelancers chose to receive a 5-star rating instead. We further show that the value that Upwork freelancers assigned to a positive review was moderated by the amount of experience they had on the platform; the median valuation of a positive review among experienced freelancers was $30.65 USD lower than that of less experienced freelancers.

We complement our quantitative analyses, which precisely measure the value that Upwork freelancers in our sample assigned to a positive review in June 2020, with qualitative analysis of the voluntary, freeform choice explanations provided by about half of the freelancers in our sample. Our qualitative analyses confirm that reputation system-based algorithmic management is a major factor that freelancers consider while navigating the trade-off between short-term profit (from the monetary bonus) and increased long-term earnings (from the positive review). Furthermore, a substantial fraction of freelancers who chose the monetary bonus did so due to COVID-19-related financial hardship, highlighting the way in which the value that online laborers assign to online reputation and the intensity with which they perceive reputation system-based algorithmic control may fluctuate over time.

2 RELATED WORK

In this section, we consider prior work on algorithmic management and reputation systems, numerically quantifying algorithmic management, online choice experiments, the impact of online reputation on earnings, and incentivizing online reviews with monetary rewards.

2.1 Algorithmic Management and Reputation Systems

The main motivation for this paper is a large body of qualitative research that focuses on "algorithmic management" (a term originally coined by Lee et al. [21] in the context of ridesharing), and in particular, the role of online reputation systems in algorithmic management. This body of work spans multiple online labor platforms and types of labor. Lee et al. [21], Robinson [32], and Rosenblat and Stark [33] all document the effects of algorithmic management on ridesharing drivers. Most relevant to this work, these papers discuss the opacity of the platforms' rating algorithms, the challenges associated with earning and maintaining a high rating, and the possibility of being removed from the platforms due to low ratings. Research by Carlos Alvarez de la Vega et al. [7], Sutherland et al.

[2]The demand curve for an item measures the percentage of people that would purchase that particular good at every possible price.

[34], and Kinder et al. [20] documents the ways in which reputation systems are used as an algorithmic management tool on macrotask platforms like Upwork; they find that Upwork freelancers prioritize building up reputation when new to the platform, that maintaining a strong reputation in the face of opaque reputation systems is a key source of precarity among Upwork freelancers, and that even slight differences in Upwork freelancers' rating scores can be the difference between many and few job prospects. Kinder et al. [20] also analyze how Upwork freelancers use their agency to circumvent power asymmetries imposed by algorithmic control. Griesbach et al. [15] documents the anxiety food delivery workers on platforms such as DoorDash and Postmates experience due to platform policies under which they may be deactivated or removed due to poor reputation scores, and Wood et al. [37] shows that platform workers in Sub-Saharan Africa and Southeast Asia who don't have strong reputations struggle to find work, indicating that findings in this literature extend to online labor markets worldwide. We build on and extend these works by quantitatively measuring the intensity with which platform workers experience the reputation system-based aspects of algorithmic control. Measuring this intensity quantitatively, as opposed to qualitatively, will allow researchers to make more direct comparisons across platforms and across time, and subsequently, better identify what platform features, market conditions, and external factors mitigate or exacerbate the intensity with which platform workers experience reputation system-based algorithmic control.

2.2 Numerically Quantifying Algorithmic Management

There is also an emerging body of research that aims to numerically quantify aspects of algorithmic management that have been documented in other, more qualitative research. Gray and Suri [14], Wood et al. [37], Jarrahi et al. [19], and Lee et al. [21] all discuss the largely illusory nature of the flexibility promised by online labor, due to the rigidity of the algorithms used to manage the workers, and the fact there is often a "feast or famine" amount of available work to be done. Building on this work, Yin et al. [38] measure how much workers value tasks with temporal flexibility built in. They show that MTurk workers value tasks with a flexible deadline (i.e., anytime in the next 24-hours) about $1 USD more per hour than the same tasks with an inflexible deadline (i.e., do them immediately). There is also prior qualitative work that observes that workers devote considerable time and effort to finding high quality work on online platforms [14, 25], often because the platform search and matching algorithms do not fully serve the workers. Toxtli et al. [35] measure the overhead workers experience while working on MTurk and determine that it effectively decreases workers' hourly wages from $3.76 USD to $2.83. We view our paper as a natural continuation of a research program that numerically quantifies the effects of algorithmic management on platform workers. As researchers develop methods to numerically quantify more components of algorithmic management systems, it becomes possible to compare which components are most intense or salient in the eyes of platform workers, e.g., do workers find it more "costly" to lose flexibility, or to have their ability to work governed by a reputation system? And what is the magnitude of the difference between the intensity with which platform workers experience these two different components of an algorithmic control apparatus?

2.3 Online Choice Experiments

This work also contributes to a growing research literature that uses online choice experiments to characterize the demand for a number of different goods, including free digital goods such as Facebook, Google search, and Wikipedia [1, 5], social connections [4], flexible work arrangements [26] and non-work time [27]. Among these papers, our work is most closely related to the work of Brynjolfsson et al. [5] and Benzell and Collis [4], who use single-binary discrete-choice experiments [8] to estimate the value that consumers assign to free digital goods (in the case of Brynjolfsson et al. [5]) and social connections (in the case of Benzell and Collis [4]).[3] We build on this work by proposing the use of online choice experiments to measure the value of a different type of good (a positive review on an online labor platform), and by extension, characterize the intensity with which platform workers experience reputation system-based algorithmic management.

2.4 The Effect of Reputation on Long-term Earnings

There is also a large body of research that aim to measure the effects of positive (and negative) reputation on the long-term earnings of sellers in online markets and online labor market workers. In the context of eBay, for instance, Resnick et al. [31] find that buyers are willing to pay 8.1% more when purchasing from a high reputation seller, and Cabral and Hortacsu [6] find that when sellers received a negative review, their sales growth rate dropped from +5% to -8%. In more traditional eCommerce settings, Park et al. [30] document that the valence of a product's first review (negative or positive) has a significant effect on the valence and volume of future reviews, while Vana and Lambrecht [36] find that individual online reviews significantly influence consumers' purchase likelihood. Finally, Pallais [29] conducts an experiment on oDesk (a predecessor of Upwork) and finds that hiring workers and providing a "coarse evaluation" had a positive impact on the longer-term employment outcomes of inexperienced workers, whereas hiring workers and providing a "detailed evaluation" had a positive impact on the longer-term employment outcomes of experienced workers. While all of these papers focus on measuring the tangible impact of online reviews on future reviews and/or future sales, we measure the subjective value that online workers assign to a single positive review, which need not be equal to the actual monetary value of a positive review. For instance, above and beyond the actual monetary value of a positive review, workers may assign some monetary value to the positive feelings they experience when receiving positive feedback, or to avoiding the negative feelings they experience when receiving negative feedback.

2.5 Incentivized Review Programs

Finally, in the absence of appropriate incentives, fewer online market participants leave ratings and reviews than is optimal from a market efficiency perspective [2]. In light of this fact, there is a

[3] A single-binary discrete-choice experiment is one in which participants are asked to make a choice between two options.

growing body of work studying the viability of coupon or rebate mechanisms to solicit reviews in online marketplaces. Existing research on this topic proposes that either the platform itself [13] or individual sellers [6, 22–24] offer a coupon or rebate in exchange for an honest review. Li et al. [23] find that 60.82% of Taobao sellers participated in such a program at least once over the course of six months, and that those that participated in the program increased sales by 36%. While these papers investigate mechanisms in which shoppers are offered monetary rewards (often by the sellers themselves) in exchange for a review of unknown quality, in this work we measure how much money platform workers are willing to forgo in exchange for a guaranteed positive review. The fact that we report a high median WTA for positive feedback among the Upwork freelancers in our sample suggests that platform workers would indeed be willing to pay a large amount in exchange for a single positive review. In light of this fact, it is tempting to conclude that such a coupon or rebate program may be well-suited to online labor markets such as Upwork. However, given the fact that many platform workers, particularly those that are inexperienced, are subject to intense economic precarity [14], we believe that such a program has the potential to do more harm than good in our research setting.

3 EXPERIMENT DESIGN

In order to demonstrate the potential of our method, we conducted an experiment on Upwork in June 2020. At a high level, we first used a brief task to recruit Upwork freelancers, and then used the set of 520 workers that accepted and completed the recruitment task as the sample (as in Ho et al. [16]) for a deterministically binding and incentive compatible single-binary discrete-choice experiment [5, 8]. In the experiment, freelancers were given the non-hypothetical choice between a 5-star rating with positive text, or a monetary bonus. The primary research question our study aims to answer is:

> **RQ1:** What is the monetary value that Upwork freelancers assigned to a single positive review in June 2020?

Our experiment design went through Institutional Review Board Review, and our experiment design and analysis plan was preregistered at https://osf.io/4ectd. A more detailed description of the experiment design follows.

3.1 Subject Recruitment

Two identical recruitment tasks were created on Upwork, one in the category of customer support and one in the category of graphic design. Out of all of the categories of work offered on Upwork, we specifically chose graphic design and customer support because the labor markets for the two types of work vary in a number of ways that we hypothesized may affect the perceived value of reputation for workers. First, there is more horizontal differentiation in the market for graphic design work than in the market for customer support work, i.e., a client is more likely to determine that a given graphic designer is a poor fit for a job due to factors that are a matter of taste, such as the style of the designer's work. Second, barriers to entry are likely higher in the market for graphic design work than in the market for customer support work. Graphic designers typically need to obtain expensive software and hardware (e.g., drawing tablets and Adobe products), and have often devoted many hours to formal and/or informal training, whereas the physical and human capital prerequisites to working in customer support are lower. Graphic designers are also typically able to charge more on Upwork than customer support workers: in our sample, the average graphic designer charged $23.30 USD per hour, whereas the average customer support worker charged $12.70 USD per hour. Finally, we expected that the sets of workers offering services in the two categories might differ on a number of demographic dimensions; for instance, previous work has shown that Upwork freelancers in different parts of the world tend to work in different categories [18].

Both recruitment tasks had the title "[5-10 minute task, help w/ research about Upwork] Tools you use doing {category} work," where {category} was filled in with either "Customer support" or "Graphic design" depending on the task. Both tasks had the following description:

> Please write a brief paragraph (3-4 sentences) describing the software and/or hardware tools that you use while doing {category} work. Answer this question however you see fit, there are no right or wrong answers. You can send your response as a word document, a PDF, a link to a collaborative document (like a google doc), or even as text in Upwork chat. Please do not spend more than 5-10 minutes on this task, it is not meant to take up much of your time. Please do not accept this task unless you market yourself as a freelancer who does {category}.

The posted price for both tasks was $10 USD. Under the assumption that freelancers took no more than the recommended 10 minutes to complete our tasks, the pay rate for the tasks comes out to at least $60 USD per hour, which is well above the average hourly rate that graphic designers and customer support freelancers in our sample charged ($23.30 USD per hour and $12.70 USD per hour, respectively).

Our goal was to construct a stratified sample of about 500 workers that was split evenly between customer support and graphic design, by hiring approximately 250 freelancers for each task. We first invited the top 2,000 freelancers recommended by Upwork's search algorithm for each task to apply for the job. We then extended offers to freelancers who applied for each task in the order the applications were received. Because more than 2,000 invitations ended up being required in order to reach the desired sample size for both the graphic design and customer support tasks, we continued inviting freelancers to both jobs (in the order recommended by Upwork's search algorithm) until both jobs had approximately 300 applications. It is not surprising that more than 2,000 invitations were required to construct our sample, given that it has been observed in other online labor markets that a small minority of workers do the vast majority of the work [9].

Not all freelancers who applied for our tasks ended up accepting our job offer. There were also a small number of cases in which the hiring process for particular freelancers was delayed due to legal compliance issues; when this happened, we extended an offer to another freelancer to ensure that our final sample for both tasks would be at least 250 freelancers. Overall, our hiring process resulted in us making offers to 267 customer support freelancers and

Table 1: Summary statistics for hired freelancers in the experiment.

Demographic	Customer support	Graphic design
Highest education		
Post-Grad	40 (15%)	58 (22%)
College	188 (73%)	160 (61%)
High School	7 (2.7%)	5 (1.9%)
Other	14 (5.4%)	25 (9.6%)
Unknown	10 (3.9%)	13 (5.0%)
Country		
Philippines	98 (38%)	21 (8.0%)
Ukraine	2 (0.8%)	32 (12%)
Pakistan	12 (4.6%)	26 (10.0%)
United States	10 (3.9%)	24 (9.2%)
India	21 (8.1%)	13 (5.0%)
Bangladesh	6 (2.3%)	19 (7.3%)
Russia	0 (0%)	13 (5.0%)
Egypt	7 (2.7%)	12 (4.6%)
Serbia	6 (2.3%)	12 (4.6%)
Canada	4 (1.5%)	5 (1.9%)
Other	93 (36%)	84 (32%)

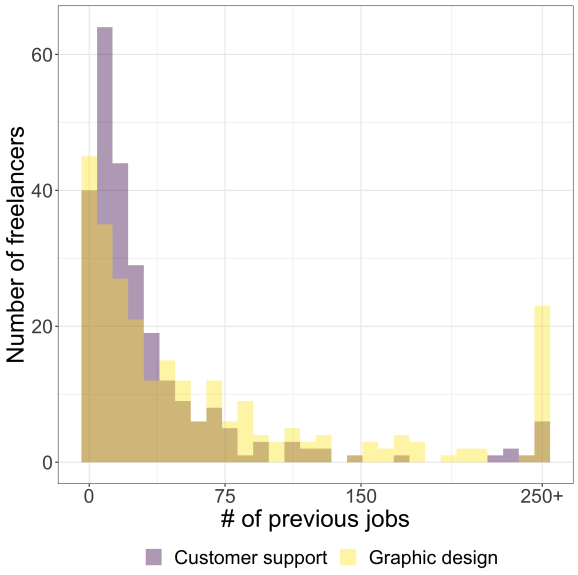

Figure 1: The distributions of number of previous jobs completed for both the graphic designers and customer support workers in our sample.

hiring 259, and making offers to 287 graphic design freelancers and hiring 261.

3.2 Subject Characteristics

At the time that freelancers in both categories were offered the job, the first author manually recorded multiple different pieces of information that were publicly visible on the freelancer's profile, including their first name, their country of residence, their highest listed level of education,[4] whether or not they had a 'Top Rated' or 'Rising Star' badge, the number of jobs they had completed on Upwork, the number of hours they had worked on Upwork, the total amount of money they had earned on Upwork, their Job Success Score (JSS), and their standard hourly wage. Using the Penn World Table [10], we also recorded the price level of household consumption as of 2019 in the freelancer's country of residence, which we use as a measure of the cost of living.[5]

Of the freelancers that were hired, 60.6% of the customer support freelancers had a Top Rated badge, and 52.9% of graphic design freelancers had a Top Rated badge. Table 1 shows the distributions of education level and geographic location for freelancers in both categories. Among both customer support workers and graphic designers, most freelancers had attended some sort of college. The geographic distributions of the two types of freelancers are also quite different: whereas customer support workers are somewhat concentrated in the Philippines (38%), graphic design workers are more evenly distributed across the world, with the highest concentration in Pakistan (10%). The average freelancers in our sample resided in a country in which the cost of living was 67.4% as high as in the United States. In other words, the average freelancer in our sample lived in a country where the US dollar goes 1.5 times as far as it does in the US. Figure 1 shows the distribution of previous work experience for workers in each category; the median number of previous jobs among hired customer support workers was 18, whereas the median number of jobs among hired graphic designers was 32.

Freelancer-level characteristics were collected in part because we wanted to test the following pre-registered hypotheses:

- **H1a:** Because of a number of factors, such as higher barriers to entry and greater levels of horizontal differentiation, graphic designers will value a positive review less than customer support workers.
- **H1b:** Because they have higher earning potential, graphic designers will value a positive review more than customer support workers.
- **H2:** Freelancers who are experienced on Upwork will value a positive review less than those who are not.
- **H3:** Freelancers who have already established a high-quality reputation on Upwork will value a positive review more highly than those who do not have a high-quality reputation to maintain.
- **H4:** Freelancers who live in countries where the cost of living is lower will be more likely to choose the monetary bonus.

Observe that H1a and H1b are opposing hypotheses. We saw the barriers to entry and higher earning potential as in tension, pulling the valuation of a review in both directions, but could not predict which would have a larger impact on freelancers' valuations.

[4]We were not able to verify that freelancers went to the institutions listed, or that they had completed the specified degree or program at any particular institution. In cases where it was unclear what level of education a freelancer had obtained, it was coded as 'Unknown.'

[5]2019 was the most recent year for which the Penn World Table had data.

3.3 Treatment Intervention

After hired freelancers had submitted their one paragraph description of the software and/or hardware they used while working, which was manually checked to confirm that the submitted text satisfied the requirements of our recruiting task, participants were sent the following message:

> Thank you, this is great work! I'm a researcher trying to understand how much freelancers on Upwork value reviews. We are offering you either a {random monetary reward} USD bonus OR a 5-star review with the following text:
>
> "{freelancer name} did a great job - they were fast, communicative, and did amazing work. I highly recommend them and would hire them again."
>
> We will give you whichever you choose, the monetary bonus or the review, but not both.

In the above, {freelancer name} was replaced with the freelancer's first name as shown on their profile, and {random monetary reward} was chosen with equal probability from the following set of values: {$25, $50, $75, $125, $175}.[6]

Freelancers who chose the 5-star review and positive text were left the rating and review as described, whereas those who chose the monetary reward did not receive any type of feedback and were instead given a bonus in the appropriate amount through Upwork's bonus payment feature. After the freelancer had received either the 5-star rating and review or the monetary bonus, they were sent a debrief statement explaining the experiment design and the purpose of the study. Many freelancers also sent free form text responses either when making their decision or after receiving the debrief statement. In these text responses, freelancers explained the factors motivating their decision, and in some cases, the factors that other freelancers might consider. Although our experiment design did not include asking freelancers for an explanation of their choice, our informed consent form indicated to participants that the number and textual content of Upwork messages sent during the experiment may be collected as part of the study.[7]

It is worth noting that our experiment design focused on 5-star ratings in particular so as to minimize both potential harm to participants and any unintended effects for Upwork as a platform. Prior work has shown that rampant reputation inflation on Upwork makes it such that leaving ratings of 4-stars or lower can harm a freelancer's reputation, and consequently, their ability to earn money [11]. Furthermore, because of the same reputation inflation dynamics, we did not expect there to be any impacts for Upwork, either positive or negative, that stemmed from the addition of a few hundred 5-star ratings to the platform's reputation system, which already included millions of ratings at the time of our study. Finally, given that the freelancers in our study satisfied the requirements of the recruitment task for which they were hired, we considered it appropriate to leave a review indicating that they had done satisfactory work.

[6]Prior to the experiment reported in this paper, a small pilot study was conducted with a different set of monetary values: {$50, $100, $150, $200, $250}. In our pilot study, we observed that for the highest monetary values, nearly all freelancers chose the monetary reward. As a result, we shifted our grid of values to cover lower monetary amounts, so as to better estimate the underlying demand curve.

[7]Our informed consent form also indicated to participants that they could opt out of the study and/or contact the authors to have their data removed.

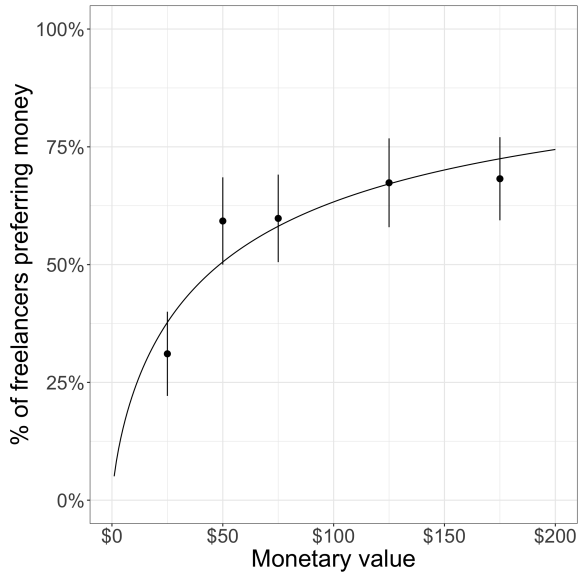

Figure 2: The estimated demand curve for a positive review across our entire sample. Points depict the proportion of freelancers choosing a monetary bonus at each discrete dollar amount, and error bars depict 95% confidence intervals. The curve is produced by estimating parameter values for the model in Equation 1.

4 RESULTS

4.1 Demand Curve Estimation

Figure 2 shows the overall results of the experiment. The five points depict the percentage of freelancers that chose the bonus over the positive review at each different monetary amount, and error bars depict 95% confidence intervals. When given the non-hypothetical choice between a positive review and a $25 USD bonus, 31.07% (±8.98%) of freelancers chose the monetary bonus, whereas when given the non-hypothetical choice between a positive review and a $175 USD bonus, 68.22% (±8.86%) of freelancers chose the monetary bonus. In other words, even when offered an extremely large sum of money, representing between 7 and 15 hours of work, roughly one-third of the freelancers still preferred to receive a single positive review.

The curve shown in Figure 2 represents the best-fit line obtained by estimating the parameters of a logit model with functional form

$$R = \alpha + \beta \log(M_i) + \epsilon_i, \tag{1}$$

where R is a binary variable that is set to 1 in cases where freelancer i chose the review and set to 0 otherwise, α is the intercept term, M is the amount of money offered to freelancer i, β is a parameter that characterizes how the probability of choosing the review

varies as a function of M_i, and ϵ_i is the residual.[8,9] Because the amount of money offered to each freelancer was randomly selected, the distribution of both observable and unobservable freelancer characteristics should in expectation be the same in each treatment arm. As a result, we do not need to control for freelancer characteristics to obtain an unbiased estimate of the relationship between monetary value offered and probability of choosing the monetary reward. The parameter estimates obtained for this model are found in column (1) of Table 2. As expected, we find a statistically significant and positive relationship between the amount of money offered to the freelancer and the probability that they choose the positive review (multiple comparisons adjusted p-value $< 10^{-5}$). In order to correct for multiple hypothesis testing, throughout the paper we report adjusted p-values calculated using the procedure developed in Hommel [17].

Recall that willingness to accept (WTA) is defined as the monetary compensation that would be required in order for a person to forfeit a particular good, in our case, a positive review. Having estimated the best-fit curve, we are able to estimate the median WTA for a 5-star rating and positive review across our sample of freelancers by calculating the value of M^* at which we predict that a representative freelancer would be indifferent between a review and a monetary bonus, i.e.,

$$\frac{\exp(\alpha + \beta \log(M^*))}{1 + \exp(\alpha + \beta \log(M^*))} = 0.5. \tag{2}$$

Thus, $M^* = e^{-\frac{\alpha}{\beta}}$. Across our entire experimental sample, we estimate that the median WTA was \$48.54 USD (95% CI = [\$33.20,\$58.06]). Throughout the paper, we calculate 95% confidence intervals for WTA estimates using a standard bootstrapping procedure ($n = 1,000$). Given that the average hourly rate among the customer support freelancers in our sample was \$12.70 USD and the average hourly rate among graphic designers in our sample was \$23.30, this result implies that the median freelancers in the two categories of work considered the value of a single positive review to be equivalent to 3.86 and 2.10 hours of work, respectively.

4.1.1 Heterogeneity in Valuations. In order to understand how the demand curve and median WTA for a positive review vary across the population, we estimate the parameters of a modified version of Equation 1 with functional form

$$R = \alpha + \beta \log(M_i) + \delta X_i + \epsilon_i, \tag{3}$$

where X_i is the value of some freelancer-level covariate, e.g., their category of work or experience level. For a given choice of covariate, if our estimate of δ is statistically significant, that indicates that in our sample there is heterogeneity with respect to that covariate in the valuation of a positive review.[10] Once again, we do not need to control for freelancer characteristics to obtain an unbiased estimated of the relationship between monetary value offered and probability of choosing the monetary reward, as the amount of money offered to each freelancer was randomly selected.

Figure 3: The estimated demand curves for a positive review across subpopulations with above- or below-median experience on Upwork. Points depict the proportion of freelancers choosing a monetary bonus at each discrete dollar amount, and error bars depict 95% confidence intervals. The curves are produced by estimating parameter values for the model in Equation 3.

Figure 3 shows the demand curves we estimate using this approach for freelancers that had less than and greater than the median amount of prior experience on Upwork, as measured through the number of jobs they had previously completed. We detect statistically significant heterogeneity in the valuation of a positive review with respect to the amount of experience a freelancer has on Upwork (multiple comparisons adjusted p-value = 0.049). The median WTA for freelancers with less than the median number of previous jobs completed was \$65.89 (95% CI = [\$47.80, \$84.94]), whereas the median WTA for freelancers with greater than the median number of previous jobs completed was \$35.24 (95% CI = [\$15.79, \$47.09]). We estimate that the difference between the median WTA for the two subgroups was \$30.65 (95% CI = [\$6.49, \$68.26]). This result is consistent with our pre-registered hypothesis (*H2*) that less experienced freelancers will value a positive review more highly, since they have more to gain by establishing a strong reputation, and is supported by textual responses provided by the freelancers in our study, which we analyze in the next section.

Using the approach described above, we also tested our pre-registered hypotheses that there would be heterogeneity in the demand curve and median WTA with respect to the type of work a freelancer specialized in (*H1a* and *H1b*), the quality of a freelancer's reputation (*H3*), and the cost of living in the country where the freelancer resided (*H4*). We were unable to detect statistically significant

[8] A model that does not log transform M_i produced qualitatively similar results, but provided a slightly worse fit to the data.
[9] This functional form is chosen because it corresponds to the random utility model, which is the standard framework to estimate utilities in choice modeling [5].
[10] This exact model specification deviates slightly from what was specified in our pre-registration, but was chosen instead due to statistical power constraints.

Table 2: Parameter estimates obtained from the overall model (column (1)) and from the heterogeneity in valuations model (columns (2) through (5)). *p<0.1; **p<0.05; ***p<0.01

	Choose 5-star review				
	(1)	(2)	(3)	(4)	(5)
log($)	-0.755^{***}	-0.755^{***}	-0.767^{***}	-0.752^{***}	-0.753^{***}
	(0.136)	(0.136)	(0.138)	(0.137)	(0.136)
Graphic design		−0.064			
		(0.183)			
Above median jobs			-0.462^{**}		
			(0.185)		
Top rated				−0.290	
				(0.185)	
Above median cost of living					0.060
					(0.183)
Constant	2.930^{***}	2.963^{***}	3.211^{***}	3.081^{***}	2.891^{***}
	(0.586)	(0.594)	(0.603)	(0.597)	(0.598)
Observations	520	520	520	520	520
Log Likelihood	−338.982	−338.921	−335.816	−337.744	−338.928
Akaike Inf. Crit.	681.963	683.841	677.632	681.487	683.855

differences in freelancers' valuations of a positive review with respect to any of these dimensions. The parameter estimates obtained for versions of Equation 3 that test for heterogeneity across these different freelancer attributes are found in columns (2) through (5) of Table 2.

4.2 Freelancer Explanations

Our quantitative analyses shed light on how much freelancers value a review, but they cannot shed light on why freelancers value reviews so highly. Importantly, it is not clear to what extent the positive review valuations held by freelancers are indicative of the intensity with which they were experiencing reputation system-based algorithmic management. Thus, we conducted a qualitative analysis that complements our quantitative analyses and begins to uncover the factors that drive freelancers' valuations of a positive review. Our qualitative analysis will show that freelancers take into account factors related to reputation system-based algorithmic management, such as their current reputation score and how many reviews they already had, when choosing between a positive review and a monetary bonus. Freelancers also took into account their current financial situation, and were actively aware of the short term vs. long term tradeoff presented by taking either the bonus or the review.

Recall that either after choosing between the positive review and monetary bonus, or after receiving their debrief statement, many freelancers voluntarily explained their choice without prompting using the chat feature of Upwork's client software. Overall, 267 of the 520 freelancers in our sample provided some textual explanation (128 graphic designers and 139 customer support freelancers). Freelancers were notified in the informed consent form delivered at the outset of the experiment that these comments would be collected and potentially analyzed by the research team. From these explanations, the last author coded the primary reason that each freelancer provided for making their decision with codes inductively generated using the freelancers' own words.

Table 3 shows the most common explanations for those who chose the review and those that chose the bonus. For those that chose the review, the most common response (30% of those who gave an explanation) was simply stating they valued the positive review, e.g. "I need your review it's very important to me, more [than] the bonus." Although vague, this explanation is not inconsistent with these freelancers having reputation system-based algorithmic management in mind when making their decision, given that some freelancers may have been unable or unwilling to fully articulate the reasons for their choice. The next most common response (27% of those who gave an explanation) focused on the fact that receiving a positive review would help the freelancer earn more in the future, e.g. "Given the current economic situation right now, I was really tempted to accept the $125 which will definitely help me with my mom's cancer meds now. However, your 5-star review and rating will definitely help me more in the long run to get me more work. I am choosing to get your 5-star rating and review." Finally, the third most common response (18% of those who gave an explanation) was that the freelancer chose the positive review

Table 3: Of those who gave an explanation, the fraction who gave each explanation as their primary reason split by those who chose the positive review versus the monetary bonus. For brevity, explanations accounting for 10 percent or less of responses in *both* categories of work were grouped into 'Other'.

Explanation	Review	Bonus
Values positive review	30%	1%
Future earnings	27%	1%
Maintain/improve reputation	18%	1%
Hardship	1%	39%
Lots of reviews already	0%	29%
Values monetary bonus	1%	11%
Other	23%	19%

because they needed it to maintain or improve their reputation on the platform, e.g., "...if one's freelancer rating on Upwork is below a 90% success rating, one gets the opportunity to bid on VERY few Projects. So, whilst I'd love the $175, my current rating could use a boost." Both the second and third most common response, which account for 45% of freelancers who chose the 5-star review, indicate that freelancers are taking reputation system-based algorithmic management into account when making their decision. It is also worth noting that all three of these reasons were largely unique to those who chose the positive review, and were almost never reported by those who chose the monetary bonus. Taken as a whole, these results show that freelancers who chose to receive the positive review were often directly considering the value of the positive review in terms that are closely related to reputation system-based algorithmic management.

The most common reason freelancers reported for choosing the monetary bonus was financial hardship (39% of those who gave an explanation). This was most often due to the COVID-19 pandemic, e.g. "I'm gonna go with the $50 bonus, for I'm in dire need brought on by the pandemic. I would have picked the positive review had I been in a different circumstance during the study." This finding corroborates the well-documented fact that the lack of a social safety net is a major source of precarity for platform workers [14, 15, 33], and suggests that the number of freelancers choosing the positive review, and hence, the median WTA that the freelancers in our sample assigned to a positive review, would have been higher had the same experiment been run prior to the COVID-19 pandemic.

The second most common reason for choosing the monetary bonus (29% of those who gave an explanation) was that the freelancer already had many reviews and/or a strong reputation. Because of this, they felt they could afford to forgo an additional rating in exchange for the money, e.g. "People like myself, who are freelancers more that 10 years - we prefer money ;) We already have so many reviews, that I don't even pay attention if a client leaves one at the end of the completed task ;)" In other words, while these freelancers were still aware of reputation system-based algorithmic management, the intensity with which they were experiencing it was not severe enough to dissuade them from choosing the monetary bonus. Finally, the third most common reason for choosing the monetary bonus (11% of those who gave an explanation) was simply that the freelancer valued the money, e.g., "of course I value a positive feedback, but currently not over 50 USD." Here again, these reasons were largely unique to those who chose the monetary bonus, and were almost never reported by those who chose the positive review. These results suggest that those who chose the bonus either had a reason that the money was more valuable to them (e.g. hardship) or were experiencing reputation system-based algorithmic management with less intensity than other freelancers, e.g., because they already had many reviews and/or a good reputation.

Across both workers who chose the positive review and workers who chose the monetary bonus, a substantial fraction of freelancers described the trade-off between the potential future earnings that a positive review could generate and the value of receiving money immediately. For example, "[From] a freelancer standpoint, if the survey was given to a freelancer who is just starting and have no reviews yet, they would most likely value a review than a monetary value. This is because they are starting and their account has no reviews or feedback yet. These reviews will be important to attract future clients. However, if you are top rated and you have enough reviews from previous clients, a monetary value will be preferred." Thus, freelancers were not just aware of reputation system-based algorithmic management as a general concept, but were actively considering the intensity with which *they specifically* were experiencing it, and how that intensity compared to the potential benefit they could receive by choosing the monetary bonus.

5 DISCUSSION

In summary, this paper develops a behavioral experiment-based method to measure the value that platform workers assign to online reputation. Insofar as platform workers' monetary valuation of online reputation is commensurate with the importance they place on online reputation, this method allows us to estimate the intensity with which platform workers experience reputation system-based algorithmic management. We demonstrate this method by estimating the value of online reputation to Upwork freelancers in the work categories of graphic design and customer support as of June 2020. We estimate that the median amount workers in our sample were willing to accept for a single positive review was approximately $49 USD. We also observed that less experienced freelancers valued reviews more. Furthermore, qualitative responses provided by the freelancers in our study indicate that the likelihood with which freelancers chose either the monetary bonus or the positive review was determined at least in part by the intensity with which they were experiencing reputation system-based algorithmic management. Freelancers who chose the positive review often did so because they thought a positive review would increase their long-term earnings or because they needed to improve/maintain their reputation on Upwork. On the other hand, those who chose the monetary bonus often did so because they were experiencing financial hardship or thought that they could forgo an additional positive review on the basis of their strong existing reputation.

Although it is plausible that the valuations provided by the freelancers in our sample would have been higher had our experiment

been conducted prior to the onset of the COVID-19 pandemic, the median WTA that we report is still extremely high relative to the hourly rates of the freelancers in our sample, and the cost of living in the countries where they reside. Recall that our estimate of the median valuation for a single positive review among the freelancers in our sample was $49 USD, whereas the average hourly rate among the customer support freelancers in our sample was $12.70 USD and the average hourly rate among graphic designers in our sample was $23.30. This means that the median freelancers in the two categories of work considered the value of a single positive review to be equivalent to 3.86 and 2.10 hours of work, respectively. Furthermore, the average freelancer in our sample resided in a country in which the cost of living was 67.4% as high as in the United States, i.e., the average freelancer in our sample resided in a country where the US dollar goes 1.5 times as far as it does in the US. These facts, taken in tandem with the fact that approximately 32% of the freelancers who were offered a $175 USD monetary bonus still preferred a single positive review, underscore the extreme intensity with which Upwork freelancers experienced algorithmic management, and in particular, reputation system-based algorithmic management, in June 2020.

5.1 Practical Implications

Our work has multiple practical implications. First, if the method proposed in this work were used to measure how the intensity with which workers experience reputation system-based algorithmic management varies across different online labor platforms, or how the intensity with which workers on the same online labor platform (e.g., Upwork) experience reputation system-based algorithmic management fluctuates over time, both researchers and advocates for platform workers would be better positioned to understand how different platform features, market conditions, and/or external economic conditions mitigate or exacerbate the intensity of the algorithmic control that workers are subject to. With this knowledge in hand, it will be possible advocate for targeted platform interventions that will lead to significant improvements in the working conditions for online laborers. Furthermore, if online labor market "clients" (e.g., the parties hiring workers) are made aware of the high value that workers assign to their reputations, it may encourage them to review their counterparties more often, and/or develop constructive relationships with their counterparties that are more likely to result in positive evaluations.

5.2 Limitations

The experiment that we conduct to demonstrate our method is not without limitations. First, the freelancers in our study are a non-random sample of the freelancers on Upwork, and are limited to two specific categories (graphic design and customer support). It is possible that the valuation we estimate is not representative of the valuation across all graphic designers and customer support workers, and it is also possible that Upwork freelancers in other categories would have assigned drastically different valuations to a single positive review in June 2020. Second, and more importantly, our experiment's sample size ($n = 520$) prevents us from detecting heterogeneity that is not extremely large in magnitude, which may at least partially explain why we were unable to detect heterogeneity with respect to category of work, quality of reputation, and local cost of living. Unfortunately, because every single freelancer who chose the monetary bonus was paid said bonus, we were unable to scale the experiment to a larger sample. Even in its current iteration, our experiment cost tens of thousands of dollars to run. Finally, it is possible that the median WTA we estimate is affected by the text of the positive review, which freelancers in our sample were shown before making their decision. Given that the reviews we provided are short and vague, the freelancers in our sample might find them less useful than a longer, more specific review. However, if this is the case, we expect that the median WTA we estimate is *lower* than the median WTA Upwork freelancers would have for longer, more specific positive feedback.

5.3 Future Work

While our demonstration estimates the value of online reputation on a particular platform (Upwork) at a particular moment in time (June 2020), it is almost certainly the case that the median valuation of a 5-star review with positive text will vary across platforms, and across time. For instance, different platforms may display workers' reviews more prominently, put greater emphasis on reviews in their ranking and matching algorithms, or more stringently remove workers based on their review history. Over time, a given platform may also make changes to its design, and/or the intensity of competition on a given platform may increase or decrease. Furthermore, changes in the offline world may impact the extent to which laborers value work on a particular platform, or the extent to which they can prioritize maximizing their platform reputation over other factors. For instance, many of the freelancers in our experiment reported choosing the monetary bonus over the positive review due to COVID-19-related financial hardship. It is plausible that if our experiment had been conducted in June 2019, before the onset of the COVID-19 pandemic, the median valuation of a single positive review would have been even higher than the $49 USD estimate we report. Importantly, the method proposed in this paper can be used to compare cross-platform and longitudinal estimates of the value of online reputation, allowing researchers to understand how the intensity of reputation system-based algorithmic management experienced by platform workers evolves over time on a given platform and varies from platform to platform. It is also possible that the behavioral experiment-based method we present could be extended to fully characterize the intensity with which platform workers experience all aspects of algorithmic control, including aspects that have already been quantitatively measured, such as flexibility [38] and work overheads [35], and aspects that have not, such as search ranking and information asymmetries. As an example of how our method might be deployed in another online marketplace, one can imagine conducting a similar experiment on a ridesharing platform, in which drivers are hired for jobs, and then given a choice between an in-app tip or a 5-star review at the end of the ride.

One obvious obstacle to replicating our proof-of-concept across platforms and across time is its high cost. However, follow up work can and should explore the use of incentive compatible, stochastically binding versions of single-binary discrete-choice experiments

[8], in which only a randomly chosen subset of those choosing the monetary reward have their selection fulfilled, or other types of choice experiments, such as Becker-DeGroot-Marschak lotteries [3] and best-worst scaling [12]. One factor worth considering while exploring these alternative choice experiment designs is the extent to which online platform workers, many of whom may not be native English speakers, will be able to easily follow the details of complicated choice frameworks, e.g., a discrete-choice experiment in which their selection is only fulfilled stochastically.

ACKNOWLEDGMENTS

The authors thank Sean Rintel, Avinash Collis, and Daniel Rock for their assistance and feedback. We are also grateful to attendees of CODE 2020, WISE 2021, members of the MIT Sloan Social Analytics Research Group, and attendees of the Berkeley Haas MORS Macro Learning Lunch for their feedback on this work.

REFERENCES

[1] Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. The welfare effects of social media. *American Economic Review* 110, 3 (2020), 629–76.

[2] Christopher Avery, Paul Resnick, and Richard Zeckhauser. 1999. The market for evaluations. *American economic review* 89, 3 (1999), 564–584.

[3] Gordon M Becker, Morris H DeGroot, and Jacob Marschak. 1964. Measuring utility by a single-response sequential method. *Behavioral science* 9, 3 (1964), 226–232.

[4] Seth Benzell and Avinash Collis. 2020. How to Govern Facebook: A Structural Model for Taxing and Regulating Big Tech. *Available at SSRN 3619535* (2020).

[5] Erik Brynjolfsson, Avinash Collis, and Felix Eggers. 2019. Using massive online choice experiments to measure changes in well-being. *Proceedings of the National Academy of Sciences* 116, 15 (2019), 7250–7255.

[6] Luis Cabral and Ali Hortacsu. 2010. The dynamics of seller reputation: Evidence from eBay. *The Journal of Industrial Economics* 58, 1 (2010), 54–78.

[7] Juan Carlos Alvarez de la Vega, Marta E. Cecchinato, and John Rooksby. 2021. "Why lose control?" A Study of Freelancers' Experiences with Gig Economy Platforms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[8] Richard T Carson, Theodore Groves, and John A List. 2014. Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists* 1, 1/2 (2014), 171–207.

[9] Jesse Chandler, Pam Mueller, and Gabriele Paolacci. 2014. Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods* 46 (2014), 112–130. Issue 1.

[10] Robert C Feenstra, Robert Inklaar, and Marcel P Timmer. 2015. The next generation of the Penn World Table. *American economic review* 105, 10 (2015), 3150–82.

[11] Apostolos Filippas, John J. Horton, and Joseph M. Golden. 2020. Reputation Inflation. (April 2020). Under Review.

[12] Terry N Flynn, Jordan J Louviere, Tim J Peters, and Joanna Coast. 2007. Best–worst scaling: what it can do for health care research and how to do it. *Journal of health economics* 26, 1 (2007), 171–189.

[13] Andrey Fradkin, Elena Grewal, Dave Holtz, and Matthew Pearson. 2015. Bias and reciprocity in online reviews: Evidence from field experiments on airbnb. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*. 641–641.

[14] Mary L Gray and Siddharth Suri. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.

[15] Kathleen Griesbach, Adam Reich, Luke Elliott-Negri, and Ruth Milkman. 2019. Algorithmic control in platform food delivery work. *Socius* 5 (2019), 2378023119870041.

[16] 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*. 419–429.

[17] Gerhard Hommel. 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* 75, 2 (1988), 383–386.

[18] John J Horton. 2021. *The Ruble Collapse in an Online Marketplace: Some Lessons for Market Designers*. Technical Report. National Bureau of Economic Research.

[19] Mohammad Hossein Jarrahi, Will Sutherland, Sarah Nelson, and Steve Sawyer. 2020. Platformic Management, Boundary Resources for Gig Work, and Worker Autonomy. *Computer Supported Cooperative Work* (04 2020). https://doi.org/10.1007/s10606-019-09368-7

[20] Eliscia Kinder, Mohammad Hossein Jarrahi, and Will Sutherland. 2019. Gig platforms, tensions, alliances and ecosystems: An actor-network perspective. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.

[21] Min Kyung Lee, Daniel Kusbit, Evan Metsky, and Laura Dabbish. 2015. Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers. https://doi.org/10.1145/2702123.2702548

[22] Lingfang Li. 2010. Reputation, trust, and rebates: How online auction markets can improve their feedback mechanisms. *Journal of Economics & Management Strategy* 19, 2 (2010), 303–331.

[23] Lingfang Li, Steven Tadelis, and Xiaolan Zhou. 2020. Buying reputation as a signal of quality: Evidence from an online marketplace. *The RAND Journal of Economics* 51, 4 (2020), 965–988.

[24] Lingfang Li and Erte Xiao. 2014. Money talks: Rebate mechanisms in reputation system design. *Management Science* 60, 8 (2014), 2054–2072.

[25] 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*. 224–235.

[26] Alexandre Mas and Amanda Pallais. 2017. Valuing alternative work arrangements. *American Economic Review* 107, 12 (2017), 3722–59.

[27] Alexandre Mas and Amanda Pallais. 2019. Labor supply and the value of non-work time: Experimental estimates from the field. *American Economic Review: Insights* 1, 1 (2019), 111–26.

[28] Mareike Möhlmann and Lior Zalmanson. 2017. Hands on the wheel: Navigating algorithmic management and Uber drivers' autonomy.

[29] Amanda Pallais. 2014. Inefficient hiring in entry-level labor markets. *American Economic Review* 104, 11 (2014), 3565–99.

[30] Sungsik Park, Woochoel Shin, and Jinhong Xie. 2021. The fateful first consumer review. *Marketing Science* (2021).

[31] Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood. 2006. The value of reputation on eBay: A controlled experiment. *Experimental economics* 9, 2 (2006), 79–101.

[32] Hilary C Robinson. 2017. *Making a digital working class: Uber drivers in Boston, 2016-2017*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[33] Alex Rosenblat and Luke Stark. 2016. Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers. *International Journal of Communication* 10, 0 (2016). https://ijoc.org/index.php/ijoc/article/view/4892

[34] Will Sutherland, Mohammad Hossein Jarrahi, Michael Dunn, and Sarah Beth Nelson. 2020. Work precarity and gig literacies in online freelancing. *Work, Employment and Society* 34, 3 (2020), 457–475.

[35] Carlos Toxtli, Siddharth Suri, and Saiph Savage. 2021. Quantifying the Invisible Labor in Crowd Work. In *Computer Supported Cooperative Work (CSCW 2021)*.

[36] Prasad Vana and Anja Lambrecht. 2021. The effect of individual online reviews on purchase likelihood. *Marketing Science* (2021).

[37] Alex Wood, Mark Graham, Vili Lehdonvirta, and Isis Hjorth. 2018. Good Gig, Bad Big: Autonomy and Algorithmic Control in the Global Gig Economy. *Work, Employment and Society* 33 (08 2018), 095001701878561. https://doi.org/10.1177/0950017018785616

[38] Ming Yin, Siddharth Suri, and Mary L Gray. 2018. Running out of time: The impact and value of flexibility in on-demand crowdwork. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–11.