Bias in Online Freelance Marketplaces: Evidence from TaskRabbit

Aniko Hannak† Claudia Wagner* David Garcia‡ Markus Strohmaier* Christo Wilson†

†Northeastern University *GESIS Leibniz Institute for the Social Sciences
‡Chair of Systems Design, ETH Zurich, Zuirch, Switzerland

ABSTRACT

In this paper, we study whether TaskRabbit, a prominent online freelance marketplace, is impacted by racial and gender bias. We collect all worker profiles from TaskRabbit to gather information about workers' gender, race, customer reviews, ratings, and positions in search rankings. We find that gender and race are significantly correlated with worker evaluations, as well as the workers' rank in search results. We hope that our study fuels more research on the presence and implications of discrimination in online environments.

1. INTRODUCTION

Online freelance marketplaces such as Upwork, Freelancer, Care.com, and TopCoder have grown quickly in recent years. In 2014, it was estimated that 25% of the total workforce in the US was involved in some form of freelancing, and this number is predicted to grow to 40% by 2020 [24, 21].

One potential benefit of online freelance marketplaces is the promise of *equality*. Many studies have uncovered discrimination in traditional labour markets [6, 13, 4], where biases can limit the opportunities available to workers. In contrast, online platforms can act as neutral intermediaries that preclude human biases. For example, when a customer requests a personal assistant from Fancy Hands, an algorithm selects the worker that will complete the task, not the customer.

However, it is unclear whether the goal of labor equality is being achieved in freelance marketplaces. Many platforms (*e.g.* TaskRabbit, Fiverr, Care.com, TopCoder, etc.) are still designed around a "traditional" workflow, where customers search for workers and browse their personal profiles before making hiring decisions. Profiles often contain the workers full name and a headshot, making it simple for biased customers to discriminate. Furthermore, many freelancing websites allow customers to rate and review workers. This opens the door to negative social influence by making (potentially biased) collective preferences transparent to future customers. Finally, freelancing sites may use rating and review data to power recommendation and search systems. If this data is biased, it may result in algorithmic systems that reinforce real-world hiring inequalities.

In this study, our goal is to examine bias and discrimination on online freelancing marketplaces with respect to gender and race. Specifically:

1. How do gender, race, and other demographics influence the social feedback workers receive?
2. Do workers' demographics correlate with their position in search results?

These questions are both relevant, as they directly impact workers' job opportunities, and thus their ability to earn a livelihood from freelancing sites.

As a first step toward answering these questions, we present a case study on TaskRabbit, one of the most prominent online freelancing marketplaces. Our crawl includes 3,707 worker profiles covering a span of 5 years. These profiles include the tasks workers are willing to complete, and the ratings and reviews they have received from customers. Since workers on TaskRabbit do not self-report gender or race[1], we infer these variables by labeling their profile images. Additionally, we also recorded each workers' rank in search results for a set of different queries and combinations of search filters.

To analyze our dataset, we use standard regression techniques that control for each worker's attributes. Our analysis reveals that gender and race have a significant correlation with the amount and the nature of social feedback workers receive. We find that women receive significantly fewer reviews, especially White women. We also find evidence for racial bias: Black workers receive worse ratings than Asian and White workers, especially Black men. Most problematically, we find algorithmic bias in search results: gender and race have significant negative correlations with search rank.

Ultimately, our findings illustrate that real-world biases can manifest in online labor markets and significantly impact the visibility of some workers. This may cause negative outcomes for workers in the form of reduced job opportunities and income. We concur with the recommendations of other researchers [14, 39, 35],

[1]We refer to this variable as "race" rather than "ethnicity" since it is only based on people's skin color.

that online labor markets should be proactive about identifying and mitigating biases on their platforms.

2. RELATED WORK

In this section, we set the stage for our study by presenting related work.

2.1 Measuring Freelance Marketplaces

Researchers have begun empirically investigating online freelancing marketplaces. Studies have used qualitative surveys to examine workers on services like Gigwalk [36], TaskRabbit [36, 37], and Uber [26]. Zyskowski *et al.* specifically examine the benefits and challenges of online freelance work for disabled workers [41]. Others have presented quantitative results from observational studies of workers [30, 8].

2.2 Discrimination

Real-world labor discrimination is an important and difficult problem that has been studied for many years [38]. Some researchers approach the problem from the perception side, by conducting surveys [4] or performing controlled experiments [6, 13]. Other studies measure the consequences of labor discrimination using observational data sets [1, 2].

Although we are unaware of any studies that examine labor discrimination on online freelance marketplaces, studies have found racial and gender discrimination in other online contexts. For example, Google's systems served ads that disparaged African Americans [35] and withheld ads for high-paying jobs from women [12]. Similarly, two studies have found that female and Black sellers on eBay earn less that male and White sellers [3, 23]. Edelman *et al.* used field experiments to reveal that hosts on Airbnb are less likely to rent properties to racial minorities [14]. Finally, Wagner *et al.* found that biased language was used to describe women in Wikipedia articles [40].

The study that is most closely related to ours is by Thebault *et al.* [37]. In this work, the authors surveyed workers on TaskRabbit from the Chicago metropolitan area, and found that they were less likely to accept requests from customer in the socioeconomically disadvantaged South Side area. In contrast, our study examines discrimination by customers against workers, rather than by workers against customers.

Mechanisms of Discrimination. Our study is motivated by the observation that the design of websites may exacerbate preexisting social biases. This may occur through the design of pricing mechanisms [15], selective revelation of user information [29], or the form in which information is disclosed [5, 7, 11, 17].

In this study, we examine a website that present workers in ranked lists in response to queries from customers. Prior work has shown that the items at the top of search rankings are far more likely to be clicked on by users [31, 10]. When the ranked items are humans, the ranking algorithm can viewed as creating status differentiation. This opens the door for the reinforcement of social biases, if the ranking algorithm itself is afflicted by bias.

2.3 Algorithm Auditing

Recently, researchers have begun looking at the potential harms (such as gender and racial discrimination) posed by opaque, algorithmic systems. The burgeoning field of "algorithm auditing" [32] aims to produce tools and methodologies that enable researchers and regulators to examine black-box systems, and ultimately understand their impact on users. Successful prior audits have looked at personalization on search engines [19, 22], localization of online maps [33], social network newsfeeds [16], online price discrimination [20, 27, 28], dynamic pricing in e-commerce [9], and the targeting of online advertisements [18, 25].

3. BACKGROUND

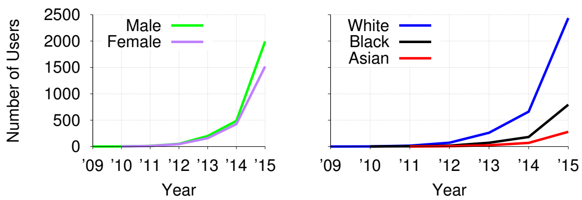
In this section, we introduce TaskRabbit. TaskRabbit, founded in 2008, is an online marketplace that allows customers to outsource small, household tasks such as cleaning and running errands to workers. TaskRabbit focuses on `physical` tasks [36], and as of December 2015, it was available in 30 US cities.

Worker's Perspective. To become a "tasker", a worker must go through three steps. First, they must sign up and construct a personal profile that includes a profile image and demographic information. Second, the worker must pass a criminal background check. Third, the worker must attend an in-person orientation at a TaskRabbit regional center [34].

Once these steps are complete, the worker may begin advertising that they are available to complete tasks. TaskRabbit predefines the task categories that are available (*e.g.* "cleaning" and "moving"), but workers are free to choose 1) which categories they are willing to perform, 2) when they are willing to perform them, and 3) their expected hourly wage for each category.

Workers are required to fill out detailed profiles with demographics, a profile image, and information about past work experiences. We observe that almost all workers have clear headshots on their profiles. However, even without these headshots, customers will still learn workers' gender and race when they physically arrive to complete tasks. Workers' profiles also list their reviews, the percentage of positive ratings they received, and the history of tasks they have completed.

Customer's Perspective. When a customer wants to hire a "tasker", they choose a task category, give their address, and specify dates and times when they would like the task to be performed. Once the customer has



(a) By Gender

(b) By Race

Figure 1: Member growth over time on TaskRabbit.

input their constraints, they are presented with a ranked list of workers who are able to perform the task. The list shows the workers' profile images, expected wages, and reviews. After a customer has hired a tasker, they may write a review on that worker's profile and rate them with a "thumbs up" or "thumbs down".

4. DATA COLLECTION

We now present our data collection and labeling methodology. Additionally, we give a high-level overview of our dataset, including worker demographics.

4.1 Crawling

To investigate bias and discrimination, we need to collect 1) demographic data about workers, 2) ratings and reviews of workers, and 3) workers' rank in search results. To gather this data, we perform extensive crawls of TaskRabbit in November 2015. We use Selenium to implement our crawlers. We were careful to impose minimal load on TaskRabbit servers during the crawl, and not to effect the workers in any way. Although the site has a Terms of Service that prohibits crawling, we believe that algorithm audits are necessary to ensure civil rights in the digital age.

At the time of our crawls, TaskRabbit provided site maps with links to the profiles of all workers in all 30 US cities that were covered by the service. Our crawler gathered all 3,707 worker profiles, including profile pictures, reviews, and ratings. Furthermore, we used our crawler to execute search queries across all task categories in the 10 largest cities that TaskRabbit is available in, to collect workers' ranks in search results.

4.2 Extracted Features

Based on the data from our crawls, we are able to extract the following information about workers:

1. *Profile metadata:* Workers' profiles include their location, languages spoken, a freetext "About" box, and links to Facebook and Google+ profiles. However, not all workers provide all of this information.

2. *Demographic information:* Workers do not self-identify their gender and race. Instead, we asked workers on Amazon Mechanical Turk to label the gender and race of workers based on their profile images. Each profile image was labeled by two workers, and in case of disagreement we evaluated the image ourselves. We find disagreement in less than 10% of cases.

3. *Activity and feedback:* For each worker, we record the date they joined TaskRabbit, the tasks they have completed in the past, when they last logged-in, and the free-text reviews and numeric ratings they have received. Workers who have 98% positive reviews and high activity in a 30 day period are marked as "Elite", which we also record.

4. *Rank:* We record the rank of workers in response to different search queries. We ran queries for all task categories in the 10 largest cities for dates one week in the future relative to the crawl date.

Overall, we identified gender and race for 88% of workers in our dataset. 42% of workers are women and 58% are men, while 73% are White, 15% are Black, and 12% are Asian. Figure 1 shows the growth of the worker population, broken down by gender and race.

5. RESULTS

We now explore race and gender bias on TaskRabbit.

5.1 Review and Rating Bias

To what extent are gender, race, and other demographic variables correlated with the reviews and ratings workers receive? This is an important question, because social feedback may influence customers' hiring decisions, especially in online scenarios where in-person evaluation is impossible before hiring.

To ensure that the effects of gender and race on social feedback are not simply due to other variables correlated with gender/race, we control for a number of factors having to do with 1) demographic information and 2) workers' experience on TaskRabbit (*e.g.* number of completed tasks). Of course, we cannot exclude the possibility that unobserved confounding variables exist, but we do control for all observable cues in our models.

5.1.1 Review Bias on TaskRabbit

Table 1a depicts the results of a negative binomial regression model using the number of reviews as dependent variable and gender and race as independent variables. The first column presents a model without interactions, while the second includes interactions between race and gender. We control for other factors, such as being an elite worker, how long the worker has been a member of TaskRabbit, the last time the worker was online (*i.e.* activity level), their average rating score, and how many tasks they have completed in the past. The "Member Since" variable of a worker is encoded as the difference

	# of Reviews (w/o Interactions)	# of Reviews (w/ Interactions)
(Intercept)	$\mathbf{-2.601}^{***}$	$\mathbf{-2.593}^{***}$
Completed Tasks	$\mathbf{0.009}^{***}$	$\mathbf{0.009}^{***}$
Elite	$\mathbf{0.368}^{***}$	$\mathbf{0.371}^{***}$
Member Since	$\mathbf{-0.308}^{***}$	$\mathbf{-0.308}^{***}$
Recent Activity	$\mathbf{0.005}^{***}$	$\mathbf{0.005}^{***}$
Rating Score	$\mathbf{0.049}^{***}$	$\mathbf{0.049}^{***}$
Female	$\mathbf{-0.087}^{***}$	$\mathbf{-0.105}^{***}$
Asian	0.092	$\mathbf{-0.145}^{**}$
Black	-0.051	0.037
Asian Women		0.127
Black Women		0.033
Observations	3,512	3,512
Log Likelihood	−11,758	−11,757

(a) Negative binomial regression using number of reviews as the dependent variable.

	Rating Score (w/o Interactions)	Rating Score (w/ Interactions)
Completed Tasks	$\mathbf{0.002}^{*}$	$\mathbf{-0.002}^{*}$
Elite	$\mathbf{0.585}^{***}$	$\mathbf{0.587}^{***}$
Member Since	$\mathbf{-0.092}^{*}$	$\mathbf{-0.100}^{*}$
Number of Reviews	0.002	0.002
Recent Activity	$\mathbf{0.017}^{***}$	$\mathbf{0.017}^{***}$
Female	−0.041	−0.08
Asian	−0.068	−0.149
Black	$\mathbf{-0.306}^{***}$	$\mathbf{-0.347}^{***}$
Asian Women		0.206
Black Women		0.092
Observations	3,513	3,513
Log Likelihood	−5,660	−-5,658.14

(b) Ordinal regression using ratings as the dependent variable.

Table 1: Variables and their relations with reviews and ratings on TaskRabbit. *Note:* $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

in years from 2015 (*i.e.* 2014 is −1, 2013 is −2, etc.). "Recent Activity" is encoded as the difference in days from the day we collected the data.

First, we examine the model without interactions. Table 1a reveals that all factors besides race have significant statistical relationships with the number of reviews a worker receives. Join date has a significant negative coefficient, which means that workers who joined recently are less likely to have received many reviews. Conversely, recent activity and the total number of completed tasks have a significant positive correlation with the number of reviews. These results are intuitive: long-term workers who are very active accrue more reviews than new or infrequent workers.

We also find that being female is associated with fewer reviews: White women receive 10% fewer reviews than White men ($IRR = 0.90$). The mean (median) number of reviews for women is 33 (11), while it is 59 (15) for men.

Next, we examine the model with interactions. In this model, the gender-coefficient captures the effect of gender for White workers, while the race-coefficient captures the effect of race on the number of reviews for men. Table 1a shows that being female given that a worker is White is associated with fewer reviews: White women receive 10% fewer reviews than White men ($IRR = 0.90$). For all three races we observe that women receive less reviews on average: for example, the mean (median) number of reviews Black women receive is 35 (12), while Black men get 65 (16) reviews.

We do not observe any significant main effects for race, but the interaction model shows that Asian men receive 13% fewer reviews than White men (IRR=0.87).

5.1.2 Ratings Bias on TaskRabbit

Ratings are another form of social feedback on TaskRabbit. Table 1b shows the results of an ordinal model using ratings as outcome variable on TaskRabbit. In the no interaction model, we observe that being Black has a significant statistical relationship with rating scores. However, we see no significant correlation in the case of gender. Furthermore, as shown by the interaction model, gender bias is most apparent in the case of men: the mean (median) normalized rating score for White workers is 0.98 (1), while it is 0.97 (1) for Black workers.

5.1.3 Disparities by City

Thus far, our analysis of TaskRabbit has focused on our entire dataset, which covers workers from 30 cities. To examine if our findings are consistent across cities, we built separate models per city and repeated each of the above analyses (number of reviews and rating score) on each geographic subset of workers. Unfortunately, most of these models produce no statistically significant results, since the sample sizes are very small (<209 workers). Instead, we present results from four of the largest TaskRabbit cities in Tables 3 and 4 in the Appendix.

We find that being female is negatively correlated with the number of reviews in every city, which aligns with our overall findings. However, we caution that only two of these correlations are statistically significant (in San Francisco and Chicago). Furthermore, we see that being Black is associated with worse ratings across all four cities, although this correlation is only significant in New York City. Overall, the correlations that we find on a city-level with respect to gender and race are in agreement with our results on the aggregate-level.

5.2 Search Ranking Bias

Motivated by the above findings, our next analysis examines whether workers? race or gender correlate with their rank in search results. To answer this question, we ran extensive searches on TaskRabbit and recorded workers' ranks in the results. This enables us to analyze

	Search Rank (w/o Interactions)	Search Rank (w/ Interactions)
Avg. Rating	**0.003*****	**0.003*****
Completed Tasks	**0.003*****	**0.003*****
Member Since	**0.457*****	**0.51*****
Recent Activity	**0.105*****	**0.089*****
Reviews	-0.000	-0.004
Female	-0.066	$-$**0.468*****
Asian	**0.283*****	**0.194***
Black	$-$**0.076***	$-$**0.428*****
Asian Female		**0.364***
Black Female		**1.3*****
Observations	12,663	9,132
Log Likelihood	$-45{,}947$	$-33{,}128$

Table 2: Ordinal regression using search result rank as the dependent variable for TaskRabbit. *Note:* $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

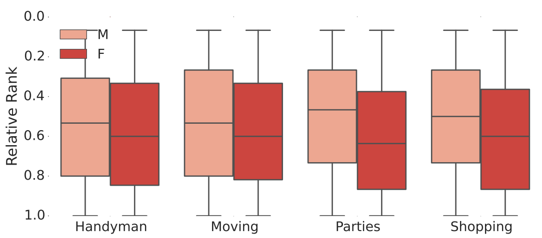


Figure 2: Search rank distributions for four task categories on TaskRabbit by gender. *Note that zero is the highest rank on the page, i.e. the first result.*

correlations between workers' rank in search results and other variables. For the purposes of our discussion, "high" ranks are the desirable positions at the top of search results, while "low" ranks are towards the bottom.

5.2.1 *Search Ranking Bias*

Table 2 shows the results of an ordinal regression model using workers' rank in search results as the dependent variable. As before, we have separate models without and with interaction effects. We observe that the number of completed tasks, membership length, and recent activity are positively correlated with rank. Additionally, ratings have a weak positive correlation, while reviews have a weak negative correlation with rank, indicating that workers with positive ratings rank higher than workers who simply have large quantities of feedback.

With respect to race, we observe that Black workers tend to be shown at lower ranks relative to White workers, while Asian workers tend be shown at significantly higher ranks. Overall, we do not observe a significant correlations with gender.

However, the results in Table 2 become more nuanced once we examine the interactions of race and gender. We observe that being a White women or a Black man has a significant negative correlation with rank. Conversely, being a Black woman has a significant positive correlation with rank. Finally, Asian workers tend to rank highly regardless of gender.

5.2.2 *Search Ranking by Task Category*

Finally, we examine rankings within individual task categories, since task categories could function as confounding factors. Figure 2 plots the search rank distribution based on gender in four different categories on TaskRabbit. Note that rank zero is the result at the top of the search results. Each bar captures the 0^{th}, 25^{th}, 50^{th}, 75^{th}, and 100^{th} percentiles. We observe that women are more likely to appear at lower ranks across all four categories, with the biggest gap in the "Parties" category and smallest in "Shopping".

6. CONCLUDING DISCUSSION

In this work we collected and analyzed data from the online freelance marketplace TaskRabbit and quantified race- and gender-based biases. Our controlled regression models reveal that social feedback on the site is impacted by gender and racial biases. Specifically, we find:

- Women, especially White women, receive 10% fewer reviews than men with equivalent work experience.
- Black workers, especially men, receive significantly lower feedback scores than other workers with similar attributes.
- TaskRabbit's algorithm produces results that are significantly correlated with race and gender, although the specific groups that are ranked lower change from city-to-city.

It is unclear why TaskRabbit's search algorithm exhibits bias. *We find no evidence that the algorithm was designed to rank based on demographic features*, and we consider this to be unlikely. Instead, we believe that the algorithm takes customer behavior into account (*e.g.* ratings, reviews, and even clicks on profiles). Unfortunately, as we have shown, customer feedback on TaskRabbit is biased, which may cause the search algorithm to exhibit bias.

Unfortunately, simply getting rid of social feedback is not an option, since customers rely on reviews when shopping online. Given that feedback must be presented to customers, marketplace proprietors should take steps to mitigate inherent biases in the data.

Our case study on TaskRabbit leaves open several directions for future work. A longitudinal observational study of worker behavior could tell us whether adverse working conditions for women and minorities cause them to drop-out of the freelancing workforce at greater rates than men. Another critical question is the precise impact of social feedback on customers' hiring decisions, which could be answered through in-person experiments.

7. REFERENCES

[1] Aleo, M., and Svirsky, P. Foreclosure fallout: The banking industry's attack on disparate impact race discrimination claims under the fair housing act and the equal credit opportunity act. *BU Pub. Int. LJ 18* (2008).

[2] Altonji, J. G., and Blank, R. M. Chapter 48 race and gender in the labor market. In *Handbook of Labor Economics*, vol. 3, Part C. Elsevier, 1999, pp. 3143 – 3259.

[3] Ayres, I., Banaji, M., and Jolls, C. Race effects on eBay. *The RAND Journal of Economics 46*, 4 (2015), 891–917.

[4] Bertrand, M., and Mullainathan, S. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American Economic Review 94*, 4 (2004), 991–1013.

[5] Bolton, G., Greiner, B., and Ockenfels, A. Engineering trust: Reciprocity in the production of reputation information. *Management Science 59*, 2 (2013), 265–285.

[6] Carlsson, M., and Rooth, D.-O. Evidence of ethnic discrimination in the swedish labor market using experimental data. *Labour Economics 14*, 4 (2007), 716 – 729. European Association of Labour Economists 18th annual conference CERGE-EI, Prague, Czech Republic 21-23 September 2006.

[7] Che, Y. K., and Horner, J. Optimal Design for Social Learning. Levine's Bibliography 786969000000001075, UCLA Department of Economics, Sept. 2015.

[8] Chen, L., Mislove, A., and Wilson, C. Peeking beneath the hood of uber. In *Proceedings of the 2015 ACM Conference on Internet Measurement* (2015).

[9] Chen, L., Mislove, A., and Wilson, C. An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace. In *Proceedings of the 25th International World Wide Web Conference* (May 2016).

[10] Craswell, N., Zoeter, O., Taylor, M., and Ramsey, B. An experimental comparison of click position-bias models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* (2008).

[11] Dai, W., Jin, G. Z., Lee, J., and Luca, M. Optimal aggregation of consumer ratings: An application to yelp.com. Working Paper 18567, National Bureau of Economic Research, November 2012.

[12] Datta, A., Tschantz, M. C., and Datta, A. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. In *In Proceedings of the 15th Privacy Enhancing Technologies Symposium* (2015).

[13] Dovidio, J. F., and Gaertner, S. L. Aversive racism and selection decisions: 1989 and 1999. *Psychological Science 11*, 4 (2000), 315–319.

[14] Edelman, B. G., Luca, M., and Svirsky, D. Racial discrimination in the sharing economy: Evidence from a field experiment, 2015. http://ssrn.com/abstract=2701902.

[15] Einav, L., Farronato, C., Levin, J. D., and Sundaresan, N. Sales mechanisms in online markets: What happened to internet auctions? Working Paper 19021, National Bureau of Economic Research, May 2013.

[16] Eslami, M., Aleyasen, A., Karahalios, K., Hamilton, K., and Sandvig, C. Feedvis: A path for exploring news feed curation algorithms. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing* (2015).

[17] Fradkin, A., Grewal, E., Holtz, D., and Pearson, M. Bias and reciprocity in online reviews: Evidence from field experiments on airbnb. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation* (2015).

[18] Guha, S., Cheng, B., and Francis, P. Challenges in measuring online advertising systems. In *Proceedings of the 2010 ACM Conference on Internet Measurement* (2010).

[19] Hannak, A., Sapieżyński, P., Kakhki, A. M., Krishnamurthy, B., Lazer, D., Mislove, A., and Wilson, C. Measuring Personalization of Web Search. In *Proceedings of the 22nd International World Wide Web Conference* (May 2013).

[20] Hannak, A., Soeller, G., Lazer, D., Mislove, A., and Wilson, C. Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 ACM Conference on Internet Measurement* (2014).

[21] Intuit. Intuit 2020 report. http://http-download.intuit.com/http.intuit/CMO/intuit/futureofsmallbusiness/intuit_2020_report.pdf.

[22] Kliman-Silver, C., Hannak, A., Lazer, D., Wilson, C., and Mislove, A. Location, location, location: The impact of geolocation on web search personalization. In *Proceedings of the 2015 ACM Conference on Internet Measurement* (2015).

[23] Kricheli-Katz, T., and Regev, T. How many cents on the dollar? women and men in product markets. *Science Advances 2*, 2 (2016).

[24] Laumeister, G. *The Next Big Thing In E-Commerce: Online Labor Marketplaces*. Forbes, 2014. http://www.forbes.com/sites/groupthink/2014/10/21/the-next-big-thing-in-e-commerce-online-labor-marketplaces/.

[25] Lécuyer, M., Ducoffe, G., Lan, F., Papancea, A., Petsios, T., Spahn, R., Chaintreau, A., and Geambasu, R. Xray: Enhancing the web's transparency with differential correlation. In *Proceedings of the 23rd USENIX Conference on Security Symposium* (2014).

[26] Lee, M. K., Kusbit, D., Metsky, E., and Dabbish, L. Working with machines: The impact of algorithmic and data-driven management on human workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015).

[27] Mikians, J., Gyarmati, L., Erramilli, V., and Laoutaris, N. Detecting price and search discrimination on the internet. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks* (2012).

[28] Mikians, J., Gyarmati, L., Erramilli, V., and Laoutaris, N. Crowd-assisted search for price discrimination in e-commerce: First results. In *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies* (2013).

[29] Morton, F., Zettelmeyer, F., and Silva-Risso, J. Consumer information and discrimination: Does the internet affect the pricing of new cars to women and minorities? *Quantitative Marketing and Economics 1*, 1 (2003), 65–92.

[30] Musthag, M., and Ganesan, D. Labor dynamics in a mobile micro-task market. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013).

[31] Richardson, M. Predicting clicks: Estimating the click-through rate for new ads. In *In Proceedings of the 16th International World Wide Web Conference* (2007).

[32] Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. Auditing algorithms: Research methods for detecting discrimination on internet platforms. In *Proceedings of "Data and Discrimination: Converting Critical Concerns into Productive Inquiry", a preconference at the 64th Annual Meeting of the International Communication Association* (2014).

[33] Soeller, G., Karahalios, K., Sandvig, C., and Wilson, C. MapWatch: Detecting and Monitoring International Border Personalization on Online Maps. In *Proceedings of the 25th International World Wide Web Conference* (May 2016).

[34] Support, T. Are your taskers screened and background checked? TaskRabbit, May 2016. https://support.taskrabbit.com/hc/en-us/articles/204411630-Are-your-Taskers-screened-and-background-checked-.

[35] Sweeney, L. Discrimination in online ad delivery, 2013. http://ssrn.com/abstract=2208240.

[36] Teodoro, R., Ozturk, P., Naaman, M., Mason, W., and Lindqvist, J. The motivations and experiences of the on-demand mobile workforce. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and; Social Computing* (2014).

[37] Thebault-Spieker, J., Terveen, L. G., and Hecht, B. Avoiding the south side and the suburbs: The geography of mobile crowdsourcing markets. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing* (2015).

[38] Todisco, M. Share and share alike? considering racial discrimination in the nascent room-sharing economy. *Stanford Law Review Online 67* (2015), 121–129.

[39] Wagner, C., Garcia, D., Jadidi, M., and Strohmaier, M. It's a man's wikipedia? assessing gender inequality in an online encyclopedia. *CoRR abs/1501.06307* (2015).

[40] Wagner, C., Garcia, D., Jadidi, M., and Strohmaier, M. It's a man's wikipedia? assessing gender inequality in an online encyclopedia. In *In Proceedings of the 9th Annual International AAAI Conference on Web and Social Media* (2015).

[41] Zyskowski, K., Morris, M. R., Bigham, J. P., Gray, M. L., and Kane, S. K. Accessible crowdwork?: Understanding the value in and challenge of microtask employment for people with disabilities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing* (2015).

Appendix

The tables in this section provide additional analysis of our dataset. Tables 3–4 examine reviews and ratings for workers on TaskRabbit in four different US cities.

	NYC		SF		LA		Chicago	
	w/o Int.	w/ Int.	w/o Int.	w/ Int.	w/o Int.	w/ Int.	w/o Int.	w/ Int.
Intercept	**-2.892*****	**-2.888*****	**-2.033*****	**-0.041*****	**-2.599*****	**-2.596*****	**-3.475*****	**-3.404*****
Completed Tasks	**0.01*****	**0.01*****	**0.006*****	**0.006*****	**0.012*****	**0.012*****	**0.016*****	**0.016*****
Elite	**0.372****	**0.375****	**0.438*****	**0.436*****	0.232	0.222	0.384	0.405
Member Since	**-0.321*****	**-0.322*****	**-0.303*****	**-0.303*****	**-0.286*****	**-0.28*****	**-0.277****	**-0.287****
Recent Activity	**0.008***	**0.009***	0.003	0.003	0.001	0.001	0.004	0.002
Rating Score	**0.051*****	**0.05*****	**0.047*****	**0.047*****	**0.047*****	**0.047*****	**0.055*****	**0.055*****
Female	-0.073	-0.069	**-0.127***	-0.109	-0.017	-0.049	-0.186	**-0.31***
Asian	0.126	0.004	**-0.245****	-0.201	-0.105	-0.043	**-0.632****	**-1.379*****
Black	**0.137***	**0.166***	0.01	0.04	0.057	-0.042	0.159	0.082
Asian Female		0.256		-0.1		-0.199		**1.189****
Black Female		-0.074		-0.065		0.204		0.163
Observations	1194	1194	845	845	582	582	211	211
Log Likelihood	-3587.8	-3587	-3375	-3374.8	-1777.1	-1776.6	-609.56	-608.08

Table 3: Negative binomial regression on TaskRabbit using number of reviews as the dependent variable. We show results without and with interactions for four different cities. *Note:* $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

	NYC		SF		LA		Chicago	
	w/o Int.	w/ Int.	w/o Int.	w/ Int.	w/o Int.	w/ Int.	w/o Int.	w/ Int.
Completed Tasks	-0.005	-0.005	0	0	-0.006	-0.006	-0.017	-0.017
Elite	**0.683***	**0.683***	0.464	0.46	0.64	0.477	0.318	0.32
Member Since	-0.148	-0.147	0.107		-0.134	-0.142	-0.532	-0.536
Number of Reviews	0.006	0.006	0	0	0.007	0.008	0.02	0.02
Recent Activity	**0.033*****	**0.033*****	-0.002	-0.002	**0.019***	**0.019***	**0.074*****	**0.074*****
Female	-0.069	-0.189	-0.004	-0.01	-0.132	-0.163	0.331	0.312
Asian	-0.211	-0.314	0.111	-0.013	-0.468	-0.631	**2.395****	**2.719***
Black	**-0.292***	**-0.41****	-0.301	-0.0164	-0.07	-0.062	-0.561	-0.621
Asian Female		0.237		0.371		0.495		-0.663
Black Female		0.284		-0.289		-0.006		0.118
Observations	1194	1194	845	845	611	611	211	211
Log Likelihood	-1858.36	-1858.61	-1448.24	-1447.58	-934.73	-934.44	-293.24	-293.12

Table 4: Ordinal regression on TaskRabbit using ratings as the dependent variable. We show results without and with interactions for four different cities. *Note:* $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$