

NBER WORKING PAPER SERIES

INCENTIVES CAN REDUCE BIAS IN ONLINE REVIEWS

Ioana Marinescu
Nadav Klein
Andrew Chamberlain
Morgan Smart

Working Paper 24372
<http://www.nber.org/papers/w24372>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2018

We would like to thank Ellora Derenoncourt, James Guthrie, Nan Li, Serge de Motta Veiga, Imran Rasul, Aaron Sojourner, Sameer Srivastava, Donald Sull, Kyle Welch, Ashley Whillans and participants to seminars at the University of Chicago for their useful comments. We would also like to thank Chen Jiang for excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Ioana Marinescu, Nadav Klein, Andrew Chamberlain, and Morgan Smart. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Incentives Can Reduce Bias in Online Reviews

Ioana Marinescu, Nadav Klein, Andrew Chamberlain, and Morgan Smart

NBER Working Paper No. 24372

March 2018

JEL No. J2,J28,L14,L86

ABSTRACT

Online reviews are a powerful means of propagating the reputations of products, services, and even employers. However, existing research suggests that online reviews often suffer from selection bias—people with extreme opinions are more motivated to share them than people with moderate opinions, resulting in biased distributions of reviews. Providing incentives for reviewing has the potential to reduce this selection bias, because incentives can mitigate the motivational deficit of people who hold moderate opinions. Using data from one of the leading employer review companies, Glassdoor, we show that voluntary reviews have a different distribution from incentivized reviews. The likely bias in the distribution of voluntary reviews can affect workers' choice of employers, because it changes the ranking of industries by average employee satisfaction. Because observational data from Glassdoor are not able to provide a measure of the true distribution of employer reviews, we complement our investigation with a randomized controlled experiment on MTurk. We find that when participants' decision to review their employer is voluntary, the resulting distribution of reviews differs from the distribution of forced reviews. Moreover, providing relatively high monetary rewards or a pro-social cue as incentives for reviewing reduces this bias. We conclude that while voluntary employer reviews often suffer from selection bias, incentives can significantly reduce bias and help workers make more informed employer choices.

Ioana Marinescu
University of Pennsylvania
School of Social Policy & Practice
3701 Locust Walk
Philadelphia PA, 19104-6214
and NBER
ioma@upenn.edu

Nadav Klein
University of Chicago
1155 E 60th St.
Chicago, IL 60637
nklein07@gmail.com

Andrew Chamberlain
Glassdoor, Inc.
100 Shoreline Highway Building A
Mill Valley, CA 94941-3645
andrew.chamberlain@gmail.com

Morgan Smart
Glassdoor, Inc.
100 Shoreline Highway Building A
Mill Valley, CA 94941-3645
morgan.smart@glassdoor.com

Incentives can reduce bias in online reviews

1. Introduction

In the age of the internet, one's reputation is almost never a blank slate. Consumers can easily find reviews of most products and services in the marketplace because other consumers have gone to the trouble of posting their opinions of these products and services online. These online reviews are an important decision aid for consumers (Chatterjee 2001; Chintagunta, Gopinath, and Venkataraman 2010; Floyd et al. 2014; Luca 2016; Mayzlin et al. 2013; Moe and Trusov 2011; Senecal and Nantel 2004), and can be helpful in making economic decisions large and small.

Choosing a job is a high-stakes decision that can be influenced by online reviews. Online reviews help to fill information gaps related to employer quality and other attributes such as salary and benefits (Card et al. 2012). A recent survey of randomly selected job seekers finds that 48% of them have used Glassdoor, the largest employer review website in the United-States, to gather information about employers (Forbes 2014). However, the largely voluntary nature of online reviews means that their aggregation may not always truthfully represent the true distribution of workers' satisfaction with their employers. If workers selectively choose to post reviews depending on how they feel about their employers -- either due to strongly positive or strongly negative opinions -- the resulting distribution of reviews may suffer from selection bias. We propose that providing incentives for reviewing can be a promising solution to address such bias.

To provide incentives to review an employer, Glassdoor limits access to information on its website unless the user agrees to provide an employer review, or some other information that Glassdoor publishes, such as an anonymous salary report. This is known at Glassdoor as the "Give-to-Get" model, i.e. give information to get information. In this article, we measure how the distribution of employer reviews incentivized by this Give-to-Get (GTG) mechanism differ from those voluntarily submitted to Glassdoor.

4 Reducing Bias in Online Reviews

We find that compared to incentivized reviews, voluntary employer reviews on Glassdoor have a significantly different distribution, exhibiting relatively more extreme “one star” and “five stars” reviews. Even after controlling for observable characteristics like employee tenure, voluntary reviews are 1.4 percentage points more likely to be one star, and 4.3 percentage points more likely to be five star compared to incentivized GTG reviews. On average, voluntary reviews have a more positive perception of employers than GTG reviews. To the extent that this is indicative of a biased distribution of voluntary reviews, can this bias significantly influence workers’ choices among employers in the labor market?

To assess the importance of this form of bias for employees’ choices among employers, we rank industries by average star rating reported by employees. We find that the industry ranking based on voluntary reviews is significantly different from the ranking based on incentivized reviews. Thus, a worker may be misled and gravitate toward jobs in, for example, advertising and marketing over an otherwise similar job in consulting, due to the inflated ratings of advertising and marketing employers among voluntary reviews. By affecting the perception of employer desirability, the likely bias in voluntary online employer reviews has the potential to affect important life choices.

However, one drawback of assessing bias in observational online employer reviews is that the “true” underlying distribution of employee opinion cannot be directly observed. In the absence of data on the true distribution of employee sentiment, one cannot be certain that voluntary reviews are in fact biased relative to incentivized reviews. To address this issue, we turn to a randomized controlled experiment on Amazon’s Mechanical Turk, an online panel widely used in behavioral experiments (Burmester et al. 2011). We experimentally manipulate participants’ ability to opt out of providing an employer review, and find that voluntary reviews are negatively biased relative to forced reviews, which are the best measure of the “true” distribution of employer ratings. On average, voluntary reviews are 0.6 stars lower than forced reviews ($p < 0.05$). We then experimentally manipulate incentives to review and find that two types of incentives

5 Reducing Bias in Online Reviews

increase the response rate and also decrease bias in reviews—a relatively high monetary incentive (75% of the study payment as reward for completing a review) and a nonspecific pro-social incentive (a request to consider how one’s review will be helpful to others). Our MTurk experiment eliminates selection effects and allows us to show that the distribution of voluntary employer reviews is in fact biased relative to the distribution of forced reviews, and that incentives for reviewing can help to remedy this bias.

Our work makes two key contributions to existing research on reputation and online reviews. First, while prior studies have suspected that bias exists in online reviews based on the shape of the distribution, we use experimental evidence to rigorously document bias (Chevalier and Mayzlin 2006; Hu et al. 2007). By comparing voluntary reviews with experimentally collected forced reviews, we can directly measure bias in the distribution of observed voluntary employer ratings.

Our second key contribution is that we leverage data from both an employer reviews website and an MTurk experiment to demonstrate that incentives -- both in a controlled experiment and in a real-world business setting -- can significantly reduce bias in online reviews. An important lesson from this work is that providing incentives does not need to be expensive: offering further information in exchange for reviewing or reminding people that their reviews can help others appear to be effective tools to reduce bias. This finding is important because it is possible that providing incentives to review could *increase* bias in observed reviews by encouraging the “wrong” people to review. In practice, we do not find this type of trade-off between offering incentives and accuracy of reviews. Instead, our findings suggest that online retailers and service providers could improve the quality of reviews by incentivizing users to provide reviews.

Existing literature has mostly focused on reviews of goods and services to consumers (e.g. for services Fradkin et al. 2015). A small number of studies have analyzed worker and employer reviews in online work platforms (e.g. Pallais 2014, Benson et al. 2015, Filippas et al. 2017). By showing that employer reviews on online work platforms are valuable to both workers and

employers, Benson et al. (2015) validate the importance of employer reviews more generally. Our work complements this literature on online work platforms by focusing on online reviews of offline employers, and by studying the role of incentives in mitigating bias in online reviews.

More broadly, our work speaks to the usefulness of online data as a powerful complement to government surveys that track economic outcomes. Online data are abundant and can be cheap to collect. However, since participation on websites is typically voluntary, the resulting information may be biased due to selection, and proper survey design and interpretation is needed (Philipson, 2001, Dillman et al. 2014). Representative government surveys not subject to these types of opt-in selection biases may seem more reliable. However, there is no government survey of worker satisfaction at the employer level, which is the focus of our paper. Furthermore, in practice the response rates for many prominent U.S. government surveys have been declining over time (Meyer et al. 2015), eroding their reliability. Our findings suggest that online data based on incentivized online responses can be reliable if properly administered, and that government surveys may also benefit from greater use of participation incentives to help fight the recent decline in response rates.

2. Literature Review and Hypothesis Development

2.1. Literature Review

The largely voluntary nature of online reviews means that their aggregation may not truthfully represent the quality of the products and services they are meant to review. If people selectively choose to post reviews of some products and services but not of others, the resulting distribution of reviews may suffer from selection bias. Indeed, existing research finds that the distributions of most reviews of retail products, motion pictures, books, and medical physicians are “J-shaped,” meaning that consumers are more likely to provide positive reviews than negative reviews, and to an even greater extent more likely to provide positive reviews than moderate reviews (Chevalier and Mayzlin 2006; Hu et al. 2007; Liu 2006; Lu and Rui 2017). These skewed

7 Reducing Bias in Online Reviews

distributions stand in contrast to “bell curve” distributions obtained in randomized experiments in which participants do not have a choice but are instead instructed to provide reviews of products. In these experiments, moderate evaluations form the overwhelming majority of reviews (Hu et al., 2009). These results suggest that consumers with moderate opinions are less motivated to provide reviews than consumers with extreme opinions, and raise the possibility that online reviews may not adequately represent the true underlying quality of the products and services they evaluate, thereby limiting their usefulness as a decision aid for consumers.

Moreover, existing research on the psychological factors motivating word-of-mouth sharing highlights the ways in which non-representative distributions of reviews can occur. Studies find that people are more likely to share emotionally charged information than non-arousing information and are also likely to share information to gain social acceptance (Berger 2011; Berger and Milkman 2012; Chen 2017). Beyond personal motives for sharing information, environmental cues also affect whether people post information online. One study of factors underlying word-of-mouth communication among consumers of a variety of products and services finds that the environmental reminders and public visibility of products makes people more likely to discuss them with others (Berger and Schwartz 2011). Similar processes may operate in online employer reviews, whereby employers who advertise more or are more visible in the marketplace are also more likely to be reviewed. Considering all of these factors together makes it clear why people may selectively review certain employers and certain employment experiences while neglecting to review others. If emotional and extreme employment experiences—negative or positive—are more likely to motivate employees to review their employers, then the resulting distribution of reviews visible to potential employees is likely to be biased.

Can the bias in online employer reviews be reduced? Here we test one way to do so by measuring the effects of different incentives on the resulting distribution of reviews. One mechanism that can explain the J-shaped distributions characteristic of online reviews is a

8 Reducing Bias in Online Reviews

motivational one. Consumers who are either highly satisfied or highly dissatisfied with their product or service are more likely to be motivated to post a positive or negative review, respectively, than consumers who do not have a strong opinion. This mechanism suggests that by changing the motivation underlying the decision to provide a review, this selection bias can be mitigated.

Perhaps the most basic way to change motivation is to change incentives (Gneezy et al. 2011). Consumers who do not feel strongly about a product or service and therefore lie in the middle of the distribution may not be motivated enough to provide a review. Providing an external incentive may motivate these consumers to provide reviews, and thereby reduce the bias commonly found in online reviews (King et al. 2014). Existing psychological research suggests that people are motivated by both the desire to benefit themselves and also by the desire to do good unto others (Buss, 1989; Dunn et al. 2008; Klein et al. 2015). In our experiment, we therefore provide different levels of monetary incentives and different kinds of pro-social incentives in order to understand the kinds and magnitude of incentives needed to reduce bias in employer reviews.

2.2. Hypothesis development

2.2.1. Glassdoor Hypotheses

People who volunteer to review their employer on Glassdoor must feel motivated to do so. This motivation may create a selection effect and thus skew the distribution of reviews: for example, people who feel angry at their employer may be more likely to provide a review and vent their emotions online than observationally similar people who are indifferent. In general, voluntary reviews are not likely to be representative of the true underlying opinion of the population of all employees. In fact, prior literature on “Online Word of Mouth” suggests that consumers of products are not all equally likely to review. Very dissatisfied or very satisfied customers are more likely to provide a review (King, Racherla, and Bush 2014). However, the

9 Reducing Bias in Online Reviews

existing literature has not explored the online reviewing behavior of employees rather than consumers.

Incentives for reviewing can provide a motivation for employees to review, independently of how strongly they feel about the employer or the value of providing a review. Therefore, incentives should be able to encourage a broader array of people to review, fostering a more representative sample of employees and therefore a less biased distribution of reviews.

Hypothesis 1: The distribution of incentivized reviews is different from the distribution of voluntary reviews on Glassdoor.

If the distribution of reviews is different, this will also lead to differences in the perception of different industries, leading to our next hypothesis.

Hypothesis 2: The ranking of industries by average employer rating is different when using incentivized reviews rather than voluntary reviews on Glassdoor.

2.2.2. MTurk Hypotheses

Observational data from online platforms such as Glassdoor cannot measure the “true” underlying distribution of employer reviews because of self-selection on the reviewers’ part. To measure the magnitude of bias in employer reviews, we conducted an experiment that allowed us to compare the distributions of reviews with and without the ability to choose whether to provide a review (i.e. with or without self-selection). The experiment manipulated whether participants were forced to review their main employer or had the choice of whether or not to review their main employer. The Choice condition corresponded to the process by which voluntary reviews are collected in practice on Glassdoor, whereas the Forced condition provided an approximation of the true underlying distribution of reviews. We compared the distribution of self-selected reviews to the distribution of reviews without self-selection. Comparing these two distributions allowed us to measure the extent of selection bias in employer reviews.

10 Reducing Bias in Online Reviews

Hypothesis 3: In the MTurk sample, the distribution of forced reviews will differ from the distribution of self-selected reviews in the choice condition.

In addition, we also tested whether different types of incentives can mitigate selection bias differently. The MTurk experiment tested five types of incentives in total, determined by random assignment. Two of the incentives were monetary—a low monetary incentive provided participants with a 25% increase in their payment for the experiment if they reviewed their employer, and a high monetary incentive provided participants with a 75% increase in their payment for the experiment if they did so. The other three incentives were pro-social, and focused on different facets of helping others through the contribution of a review. First, a nonspecific prosocial incentive framed employer reviews as a way to help others make better employment decisions. Second, a negative prosocial incentive framed employer reviews as a way to protect employees from the worst employers to work for. This type of incentive is actually used by non-profits that attempt to encourage people to review their employers in order to expose employers who mistreat their (mostly low-income) employees. Third, to test for potential asymmetries in the valence of the prosocial incentive, we included a positive prosocial incentive that framed employer reviews as a way to inform employees about the best employers to work for. We predicted that because of the motivational deficit that discourages “middle-of-the-road” reviews, any incentive that increases the motivation to provide employer reviews should also reduce selection bias.

Hypothesis 4: Incentives that are successful in increasing response rates in self-selected reviews will also reduce bias in the distribution of reviews.

3. Glassdoor Data and Methods

3.1. Glassdoor Methods

Glassdoor is an online job site that houses content such as anonymously reported salaries, online job postings, and anonymous employer reviews. The employer ratings scale at Glassdoor

11 Reducing Bias in Online Reviews

follows a classic Likert ratings scale: 1 stars to 5 stars, with 5 stars representing the highest level of employee satisfaction. Like other websites that house ratings and reviews, any person is free to visit Glassdoor to post employer reviews. We treat people who log onto the website and post a review without being prompted to do so as providing voluntary reviews. In contrast, Glassdoor also has an alternative method of employer review generation. When a user first visits the site, after viewing three pieces of content (such as three salaries, one review and two salaries, or any other combination of three pieces of online content), he or she is forced to submit a piece of content themselves in order to continue viewing additional content. This economic incentive to contribute content is referred to as the company's Give-to-Get (GTG) policy.

We treat people who post a review after being prompted to contribute content in exchange for access for more information as providing incentivized reviews. As of January 2018, roughly 24 percent of employer reviews collected by Glassdoor were contributed immediately after facing the GTG policy; the remaining 76 percent were either voluntarily contributed or left by users who had faced the GTG policy at some earlier time and returned to the site to contribute. The GTG policy has been in place since the company's founding in 2007, and is deployed uniformly across all industries and occupations. More information about the company's GTG policy is available at http://help.glassdoor.com/article/Give-to-get-policy/en_US/.¹

We use a sample of 188,623 U.S. employer reviews published on Glassdoor from 2013 to 2016. We keep in the sample only the most recent review of a person's current employer. To be able to control for demographic bias, we keep only Glassdoor users for which we have available age, gender, and highest education. Additionally, all the reviews we used came from a recognizable device—mobile phone, desktop computer, or tablet. To be able to control for differences in ratings because of a company's size or status, the reviews used were also only those for which the reviewed employer belonged to a known industry, geographic state, and had

¹ For an example of previous research examining the external validity of Glassdoor reviews relative to a well-known measure of employee satisfaction from Fortune's "100 Best Companies to Work For," see Huang et al. (2015), Section 2.3.

12 Reducing Bias in Online Reviews

a known number of employees. Lastly, to focus on individuals who are thorough in their reviews and therefore more likely to provide quality reviews, all the reviews we used had every rating field in the online review survey filled in.

Table 1 shows summary statistics for the Glassdoor sample of reviews, as well as for the MTurk sample we used in the subsequent experiment.

3.2. Glassdoor Results

We test for differences between voluntary and GTG reviews in terms of both the mean of the distribution and the overall shape of the distributions. Graphically, we can see that the distribution of voluntary ratings includes more one star and five star ratings than the distribution of GTG ratings (Figure 1). The difference between the two distributions is statistically significant at the 1% level according to a chi-squared test.

When running OLS regressions in Table 2, we can see that voluntary reviews tend to be slightly more positive: after controlling for observables, we find that a voluntary review gives a significant 0.035 star extra on average (column 2). After controls, voluntary reviews are 1.4 percentage points more likely to be one star (column 4), and they are 4.3 percentage points more likely to be five stars (column 6). This pattern explains the positive bias in the average number of stars resulting from voluntary reviews. Using an ordered logit in columns 1 and 2, and a logit in columns 3-6 leads to the same qualitative results (results not shown). In further analysis available upon request, we show that these results are robust to controlling for observables by propensity score matching and cross-validation rather than by adding observable characteristics in a linear fashion.

These results confirm our Hypothesis 1, i.e. that the distribution voluntary reviews is different from the distribution of reviews that are incentivized via the GTG policy.

Voluntary reviews are more positive than incentivized reviews, but does this matter in practice? When people browse Glassdoor, they typically aim to find information about

13 Reducing Bias in Online Reviews

employers that could help them decide which employer to work for. Therefore, if the bias in reviews due to self-selection does not affect the ranking of employers, then this bias may not be important in practice.

It turns out that the difference in the distribution of observed voluntary reviews relative to give-to-get reviews is not innocuous. Instead, it can substantially affect the ranking of industries² to which employers belong. Figure 2 plots the ranking (lower rank means better reviews) of frequent industries (those with at least 500 reviews collected via the GTG policy) for GTG vs. voluntary reviews. The 45-degree line indicates that the rank of an industry is the same under GTG and voluntary reviews: an example of such an industry is colleges & universities. Industries below the 45-degree line are ranked worse under GTG than under voluntary reviews, and there are many such industries in the graph, which is consistent with the fact that voluntary reviews tend to be more positive. To the extent that incentivized GTG ratings are more accurate, the consulting industry is a much more desirable (lower rank) industry than the advertising & marketing industry. Yet, if we rely only on voluntary reviews, advertising & marketing appears more desirable than consulting. Those who only have access to voluntary reviews may gravitate toward jobs in advertising & marketing as a result, even though comparable jobs in the consulting industry are in fact more “desirable” from the perspective of employees. These results confirm our Hypothesis 2.

An important limitation of this analysis is that observational data alone do not necessarily reveal the true population distribution of employer ratings: even with GTG incentives, not all employees will rate their employers. Though there are reasons to believe that GTG reviews are less biased than voluntary reviews, we cannot know this with certainty without information about the true distribution of employer ratings. To assess this bias more rigorously, we next turn to an experiment on MTurk where we measure both the true underlying distribution of employer

² In principle, one could rank employers (rather than industries) according to give to get ratings vs. incentivized ratings. However, we did not extract this data to protect employers’ personally identifiable information. Furthermore, the ranking of industries arguably provides more general information that many workers are likely to be interested in.

14 Reducing Bias in Online Reviews

reviews and the self-selected distribution when people have the option not to provide a review. The experiment will further allows us to test which types of incentives are most effective in getting people to review and correct bias from self-selection.

4. MTurk Experiment

4.1. Participants

Participants ($N = 639$) were recruited from Amazon's Mechanical Turk (MTurk) to participate in a five-minute survey about employer reviews in exchange for \$0.20, a typical payment in this marketplace. We selected our sample size to have at least 50 participants per cell in our experiment, which gave us at least 80% chance of detecting differences between our conditions based on a power analysis. MTurk is an online marketplace matching researchers with participants interested in doing experiments in exchange for monetary compensation (Buhrmester et al. 2011; Paolacci et al. 2010). To be eligible, participants had to be U.S. residents, employed in a job outside of Amazon MTurk (referred to as their "main employer"), and could not be self-employed. Table 1 provides demographic details about this sample. The only notable difference between the Glassdoor sample and the Amazon MTurk sample based on the available demographic data³ is the greater representation of large employers on the Glassdoor website. With this exception, the Glassdoor and MTurk samples appear very similar.

4.2. Procedures

The experiment included two factors and 12 experimental conditions, resulting in a $2(\text{Choice vs. Forced Review}) \times 6(\text{Incentive: None, High Monetary, Low Monetary, Nonspecific Prosocial, Positive Prosocial, Negative Prosocial})$ between-subjects design. Participants were

³ There could be unobserved differences between the Glassdoor and MTurk populations. For example, the average Glassdoor website visitor might be less likely to agree to do a survey for low payment than the average MTurk participant. However, this does not diminish the ability to compare between self-selected MTurk reviews and MTurk reviews without self-selection, as this experiment does.

15 Reducing Bias in Online Reviews

first randomly assigned to either the Choice or the Forced condition. In the Choice conditions, participants were asked whether they were interested in providing a review of their main employer (refusing to do so did not affect their base compensation for this experiment). Thus, the Choice conditions were a proxy for what the distribution of reviews looks like when participants self-select whether to review or not. This approximates the collection of voluntary reviews in the Glassdoor sample. In the Forced conditions, participants were simply instructed to review their main employer, and refusing to do so meant terminating their participation and canceling their base compensation for the experiment; no participant in these conditions terminated their participation. Thus, the Forced condition is a proxy for what the true underlying distribution of reviews looks like without self-selection. There was no theoretically equivalent condition in the Glassdoor sample to this Forced condition, because in the real world people are never forced to log onto the website and provide reviews.

The incentives for reviewing were also randomly assigned. In the No Incentive condition, participants did not receive any additional compensation for reviewing their employer. In all other conditions, participants were given an incentive to provide a review. Two incentives conditions were monetary. In the Low Monetary Incentive condition, participants were given an additional \$0.05 if they reviewed their main employer (a 25% increase to base compensation). In the High Monetary Incentive condition, participants were given an additional \$0.15 if they reviewed their main employer (a 75% increase to base compensation). These monetary incentives are low in terms of raw amounts, which makes for a conservative test of whether they can increase people's willingness to review their employers. At the same time, because these monetary incentives are high relative to the common payment in this marketplace, they could change people's behavior. The other three incentives were pro-social, focusing on different ways in which participants' reviews can help others. In the Nonspecific Prosocial condition, participants were asked to provide their review because it would help communicate important information to people and help them make educated decisions about working for different

16 Reducing Bias in Online Reviews

employers. In the Positive Prosocial condition, participants were asked to provide their review to “expose and reveal the best employers to work for” and thereby help people seek out these good employers. Finally, in the Negative Prosocial condition, participants were asked to provide their review to “expose and reveal the worst employers to work for” and thereby help people avoid these bad employers. All of the manipulations in this experiment were “between-subjects,” whereby each participant was assigned to either the Choice or the Forced conditions and to only one incentive regime.

After learning their incentive regime, participants in the Choice conditions were asked whether they are willing to review their main employer. Choice participants who agreed were asked to provide their overall rating of their main employer on a scale identical to the one used on the Glassdoor website. Choice participants who declined were not asked to review their main employer. Participants in the Forced conditions completed reviews of their main employer on an identical scale without being given a choice of whether to do so. The scale for reviewing employers comprised of five stars, with five stars representing the highest possible rating and one star the lowest. This was our main dependent variable.

All participants (including those who declined to review their main employer) then provided details about their employer, including tenure with this employer, the industry of the employer, and the size of employer. Finally, all participants completed questions about their personal demographics, were thanked, and dismissed.

4.3. Results

4.3.1. Efficacy of Incentives. We first analyzed the Choice conditions to test the efficacy of the different incentives in motivating participants to elect to provide employer reviews. Figure 3 presents the results. An omnibus chi-square test across incentives revealed that the incentive affected the choice to provide a review, $\chi^2 = 10.50, p = 0.062$. Compared to the No Incentive condition ($M = 66.7\%$), the High Monetary incentive ($M = 83.9\%$) significantly increased

17 Reducing Bias in Online Reviews

reviews, $\chi^2 = 4.42$, $p = 0.036$, and the Nonspecific Prosocial incentive ($M = 81.5\%$) marginally increased reviews, $\chi^2 = 3.09$, $p = 0.079$. The other incentives did not meaningfully increase reviews compared to the No Incentive condition, $\chi^2s < 0.03$, $ps > 0.88$.

Because response rates in the No Incentive condition were relatively high, this limited the room for meaningful increases. Nevertheless, these results suggest that two types of incentives increased response rates, namely the High Monetary incentive and the Nonspecific Prosocial incentive (albeit marginally). The latter is more cost-effective, because it requires merely reminding participants of the prosocial benefits of their reviews rather than paying them with additional funds. Interestingly, the Low Monetary incentive did not increase response rates, consistent with existing research suggesting that the effects of monetary incentives on behavior are nonlinear (Gneezy et al. 2011). To increase response rates for employer reviews requiring less than a minute, a relatively high monetary incentive (75% of base payment) was required.

4.3.2. Bias in Employer Reviews Without Incentives.

We assume that the Forced condition without incentives is the closest approximation to the true underlying distribution of employer ratings because it is not affected by incentives or self-selection. We tested selection effects in the absence of incentives by comparing the positivity of employer reviews in the Choice condition without incentives and the Forced condition without incentives. On average, employer ratings were significantly more negative when participants had the choice of whether to provide them ($M = 2.30$, $SD = 1.89$) relative to when participants were forced to provide them ($M = 4.02$, $SD = 0.86$), $t(106) = -6.10$, $p < 0.001$, $d = 1.18$. Moreover, the distributions of the reviews differed between the Choice and Forced conditions without incentives, $\chi^2(4) = 8.54$, $p = 0.074$. As Figure 4 shows, voluntary reviews exhibited a downward bias in employer ratings. This result is consistent with our Hypothesis 3. When left to make their own choices, people provide more negative reviews compared to the distribution of forced reviews. In contrast to what we observed in Glassdoor data, here selection effects did not

polarize ratings toward both extremes. Whereas voluntary Glassdoor reviews had both more negative and more positive extremes, in the MTurk sample selection effects biased the distribution downwards. We discuss one possible explanation of this difference between the Glassdoor and MTurk datasets below in the Discussion.

4.3.3. Bias in Employer Reviews with Incentives.

Next, we examined whether the different incentive regimes affected the selection bias in employer ratings. We first examined the incentives we found to be effective in increasing response rates, namely the High Monetary incentive and the Nonspecific Prosocial incentive. As Figures 5 and 6 show, neither of these incentives resulted in a biased distribution of reviews compared to the Forced condition with no incentives (which we treat as an approximation of the true distribution), $\chi^2_s < 4.02$, $p > 0.403$. In addition, we conducted a regression with the choice condition as the independent variable and employer ratings as the dependent variable along with control and demographic variables. As Table 3 shows, participants who received the High Monetary or Nonspecific Prosocial incentives and could *choose* whether to review provided reviews that were *not* biased compared to participants who received these incentives and were forced to review. In sum, these results suggest that these two types of incentives not only increase response rates, but also result in review distributions that more closely mirror the true distribution (i.e., the distribution in the Forced response without incentives), consistent with our Hypothesis 4.

We next compared the distribution of reviews in the High Monetary incentive condition in the MTurk experiment and the Glassdoor GTG reviews. As Figure 7 shows, the two distributions were not different from each other, suggesting that across these two different sample of participants, a self-oriented incentive (GTG in the Glassdoor case, High Monetary Incentive in the MTurk case) resulted in similar distributions of reviews.

19 Reducing Bias in Online Reviews

We next examined bias in reviews for the other 3 incentives that did not increase response rates, namely the Low Monetary, Positive Prosocial, and Negative Prosocial incentives. As Table 4 shows, compared to the no incentive condition in which participants were forced to provide reviews, none of these incentive conditions resulted in biased distributions of reviews. Thus, although the low monetary, positive prosocial, and negative prosocial incentives failed to motivate responses, they nevertheless eliminated the selection effects found in voluntary reviews. This result suggests that incentives can reduce bias without increasing overall response rates, presumably because they change the composition of individuals willing to provide reviews.

In sum, we find that the two incentives most effective in increasing response rates also do not exhibit detectable selection effects. The distributions of reviews resulting from the High Monetary and Nonspecific Prosocial incentives are not statistically different from the distribution resulting from Forced reviews with no incentives, suggesting that these two incentive regimes not only increase response rates, but also reduce bias from self-selection.

4.3.4. Framing Effects.

We next tested for framing effects, whereby the incentives themselves can affect the distribution of forced reviews without any effects on selection. In other words, participants may have provided systematically more positive or negative reviews as a result of merely thinking about different incentives even when they did not have a choice about whether or not to review their main employer (i.e. in the Forced condition). For example, the Positive Prosocial incentive, because it brings to mind good employers, might increase reported positive ratings. To test for a framing effect, we again assume that the true distribution of employer reviews is best approximated by the Forced condition without incentives. A framing effect is then defined as the impact of an incentive in the Forced condition compared to the Forced condition without incentives. Table 5 presents the results of a regression that separates framing effects and selection effects. The framing effects are measured by the effects on employer reviews of the

20 Reducing Bias in Online Reviews

different incentives in the Forced condition (first set of coefficients in Table 5); the selection effects are measured by the interaction between the Choice conditions and these incentives (coefficients on incentives in lines below Choice* in Table 5 are interaction effects between Choice and the specific incentive). All effects are expressed relative to the Forced condition without incentives. The simple coefficients associated with the different incentives relative to the Forced No Incentive condition correspond to framing effects, and the interaction terms correspond to selection effects.

Relative to the Forced No Incentive condition, the Nonspecific Prosocial incentive and the Positive Prosocial incentives resulted in more negative ratings. These results suggest that these two incentives were associated with negative framing effects—merely thinking about how one’s reviews will help others (Nonspecific Prosocial incentive) or about revealing the best employers to work for (Positive Prosocial incentive) led participants to provide more negative ratings relative to the Forced No Incentive condition. This result appears consistent with existing research that suggests that when thinking of others, people err on the side of caution because the possibility that they would lead others to make a wrong decision looms large in people’s minds (Dana and Cain 2015). By providing more negative reviews, participants in these prosocial incentive conditions may have been trying to avoid giving overly rosy views of their employers to others. The other incentives were not associated with framing effects.

Interestingly, the nonspecific prosocial incentive also resulted in a positive selection effect, because the interaction between the Choice condition and the Positive Prosocial condition was significantly positive. The magnitudes of the framing effect and the selection effect for the nonspecific prosocial incentive were similar, leading them to cancel each other out. This resulted in an unbiased distribution of reviews for the Nonspecific Prosocial condition relative to the Forced No Incentive condition. Thus, the Nonspecific Prosocial incentive reduced bias in reviews because of two contrasting effects: A negative framing effect, whereby thinking about helping others led to more negative reviews; and a positive selection effect, whereby thinking

about helping others led more participants with positive evaluations of their employers to provide reviews.

5. Discussion and Conclusion

5.1. General Discussion

Employer reviews can be a useful resource for workers in choosing where they want to work. However, voluntary online reviews may not always be reliable. Using an experiment, we have shown that the distribution of voluntary employer reviews differs significantly from the distribution of forced reviews. While selection bias is an issue, we have demonstrated that it is possible to reduce this bias by providing incentives to review. These incentives are also effective in a real-world setting, as we demonstrate using data from Glassdoor's Give-to-Get policy.

Because many aspects of a workplace are only revealed over time when working there, it can be difficult for prospective employees to assess the desirability of different employers. Online platforms like Glassdoor can help fill this gap by providing employer reviews from current (and past) employees. This should in theory help workers make more informed choices. However, such information is useful only to the extent that it paints a truthful picture of the underlying distribution of opinion about what it is like to work for different employers. Glassdoor's use of incentives through its Give-to-Get policy does appear to decrease selection bias in online reviews, and can thus better reveal information about the desirability of different employers. Future research should explore additional incentives-based strategies that can provide a less biased distribution of employer reviews on websites like Glassdoor.

A key methodological innovation of our paper is in providing unbiased reviews in our experiment (reviews in the Forced condition). Indeed, the literature on bias in online word of mouth is typically unable to compare reviews with a meaningful "true" assessment for products and services. The work by Hu et al. (2009), using a strategy similar to ours, compares the ratings for a CD on Amazon to the ratings of a group of participants who had to review the CD. They find that the J-shaped reviews on Amazon can be explained by a combination of "purchasing

22 Reducing Bias in Online Reviews

bias” and “reporting bias”: i.e. only people who like the CD tend to buy it, and then, conditional on buying, extreme opinions are more likely to be reported. Lu and Rui (2017) use a different strategy to get at “ground truth”: they compare cardiac surgeons’ reviews on RateMD with their medical outcomes. They show that reviews are correlated with medical outcomes, which establishes that reviews are informative about this important life outcome. However, their work does not explicitly treat bias in reviews because they do not compare RateMD reviews with a set of unbiased reviews. Other studies attempt to resolve this problem by comparing consumers’ reviews to reviews of experts, but this strategy leaves gaps because the two populations tend to evaluate products and services based on different criteria (Simonson 2016).

Interestingly, we find that the direction of selection bias in Glassdoor data differs to some extent from that of our MTurk data. Glassdoor voluntary employer ratings were more polarized in both the positive and negative directions compared to the GTG employer ratings. In contrast, voluntary non-incentivized MTurk employer ratings were biased *only in the negative direction* compared to forced MTurk employer ratings. This inconsistency could be explained in part by employers’ strategic behavior, as employers may encourage employees to provide positive reviews on Glassdoor.⁴ If employers can exert influence over employees and motivate them to provide positive reviews, the distribution of voluntary reviews may exhibit more positive extremes than incentivized reviews (with the negative extremes found in voluntary Glassdoor reviews attributable to the high motivation to contribute poor reviews for bad employers). In the MTurk sample, however, employers do not have the ability to motivate positive reviews, potentially eliminating positive extremes that might otherwise exist. Whatever the reason behind these differences, both the Glassdoor and MTurk datasets are consistent in two important

⁴ Note: Glassdoor’s terms of use prohibit employers from providing monetary compensation in exchange for employees leaving online reviews, and reviews in violation of that policy are removed when identified. However, it is not a violation of the site’s terms of use to encourage employees to leave reviews without offering a direct incentive. See: <https://www.glassdoor.com/employers/start/common-questions.htm>.

respects: Both reveal evidence of selection bias in voluntary, non-incentivized reviews, and both reveal that incentives to provide reviews can reduce bias.

5.2. Incentives As a Way of Reducing Bias in Voluntary Reviews

This research contributes to our understanding of incentives, both monetary and pro-social. Existing literature suggests that monetary incentives work best when they encourage precisely the desired behavior and when they are high enough to justify the effort required to attain them (Gneezy and Rustichini 2000). In line with existing research, we find that the magnitude of monetary incentives matters, whereby only the high monetary incentive increased the motivation to review sufficiently to increase response rates and change the review distribution.

Less is known about the factors that determine the efficacy of pro-social incentives. At a basic level, it is clear that people are motivated by the desire to do good by others because pro-social behavior increases psychological well-being, especially happiness and a sense of meaning in life (Dunn, Aknin, and Norton 2014; Klein 2017). Here we attempted to provide new insight into pro-social incentives by unpacking the motivation to help others into either the desire to help people identify the best employers or the desire to help protect people from the worst employers. We find that neither of these positive and negative pro-social motivations increased response rates in our employer reviews, perhaps because it is difficult for people with moderate opinions of their employers to connect with incentives that ask them to provide reviews of the “best” or “worst” employers. Moreover, people may be unlikely to believe that their employers are extreme enough to be the “best” or “worst” employers. For these reasons, persuasion appeals that emphasize extreme employers in attempt to motivate people to provide online reviews may have limited efficacy. The subtler pro-social motivations we tested were less effective than the more generalized and non-specific motivation of providing employer reviews in order to help others.

The present research joins a number of studies finding that social incentives can be just as effective in motivating behavior as monetary incentives can (e.g. Bandiera, Barankay, and Rasul

2010; Heyman and Ariely 2004; Huang, Ribeiro, Madhyastha, and Faloutsos 2014). In our case, the Nonspecific Prosocial incentive increased response rates by the same margin as the High Monetary incentive did. The commonly cited economic argument in favor of employing social incentives compared to monetary incentives is that they should be used whenever possible because they are less costly than monetary incentives. However, our Glassdoor data show that one can use incentives oriented towards self-gain without an explicit out-of-pocket cost. Recall that Glassdoor's GTG policy allows users to see employer information only if they themselves provide employer reviews or other information such as salary. This incentive is self-oriented, in that users provide employer reviews mainly to unlock valuable information for themselves. But this incentive does not cost Glassdoor money, and it in fact benefits the company by increasing the number of reviews contained in the website while also reducing the selection bias in reviews, thus improving the overall quality of the service. This will more generally also be the case for any company that aggregates user data—companies can increase user participation without expending money to incentivize users by conditioning access to valuable data on user participation. Thus, in some cases, self-oriented incentives can be as costless as non-monetary incentives.

5.3. Practical Implications

We have shown that incentives can reduce bias in employer reviews. This suggests that websites and government surveys alike can use incentives to increase the response rate and reduce bias. However, it is important to recognize that not all incentives work: in order to significantly increase response rates, relatively high monetary incentives must be provided, making this strategy impractical in many cases. Moreover, the precise magnitude of “high” monetary incentives will differ by context. A certain level of payment can be considered high in one industry while being considered low in another. Thus, companies and governments will have to experiment on a small scale to calibrate monetary incentives before rolling them out on a

larger scale. Using pro-social cues as incentives seems more promising in this respect because they do not require calibration.

The response rate to government surveys is likely declining in part because people are over-surveyed (Meyer et al. 2015). If all surveyors provide higher incentives, this will not necessarily improve response rates much because respondents will still be pressed for time. Our finding that incentives work is nevertheless crucial in a cost-benefit context: if the benefits of a high response rate and low bias are high enough, there are costly but effective ways of getting these results.

5.4. Online reviews As a Tool For Communicating Reputation: Promise and Perils

The present research also contributes to our understanding of broader issues related to the advent of online reviews as a means of quickly propagating reputation in the marketplace. There are obvious advantages to online reviews. They are easily accessible and often free to use, and therefore have the potential to increase the efficiency of markets and allow consumers to make more informed and more optimal decisions.

However, online reviews also have less obvious disadvantages. First, consumers' reviews of products and services lack objectivity, and often diverge from the opinions of experts (De Langhe, Fernbach, and Lichtenstein 2015). Second, the consumption of online reviews may not be systematic. Consumers may engage in selective or incomplete information search when evaluating reviews (Ariely, 2000; Urbany, Dickson, and Wilkie 1989). For example, existing research suggests that the characteristics of the choice set affect how people search for information. Larger choice sets (corresponding to websites with large amounts of consumer reviews) lead people to stop information search earlier in part to conserve cognitive resources (Diehl 2005; Payne, Bettman, and Johnson 1988; Iyengar, Wells, and Schwartz 2006). The order of the reviews people read can also affect when they stop searching for more information, as existing research finds that search strategies adopted in initial search environments tend to persist into different search environments (Broder and Schiffer 2006; Levav, Reinholtz, and Lin 2012).

The interpretation of online reviews may also be susceptible to information-processing biases observed in other domains (Kahneman and Tversky 1982). For example, consumers may not intuit the selection biases inherent in employer reviews, failing to appreciate the non-representative polarity of the typical J-shaped review distribution. As another example, consumers may myopically focus on the “star” rating of a product or a service while failing to take into account the reference point upon which the rating is based. For example, a financial start-up company that has an average review score of 4.5 stars will likely differ from an established investment bank that has the same average review, because the reference point of employees working in these two companies differ in many ways. However, prospective employees may not fully account for these inherent differences when evaluating the two companies, and may instead myopically focus on their similar average ratings. Overall, while online reviews are an important tool for communicating reputation in the marketplace, their limitations can be consequential and warrant further study.

5.5. Conclusion

We assess the reliability of online employer reviews and the role that incentives can play in reducing bias in the distribution of employee opinions. Using data from a leading online employer rating website Glassdoor, we have shown that voluntary reviews are more likely to be one star or five stars relative to incentivized reviews. On average, this difference in the distribution leads to voluntary reviews being slightly more positive than incentivized reviews. Using an experiment on Amazon’s Mechanical Turk, we show that voluntary employer reviews are biased relative to forced reviews. Forced reviews in our experimental setting provide an unbiased distribution of reviews that is not available in observational data from Glassdoor, and allow us to rigorously demonstrate bias in voluntary reviews. Our experiment allows us to show that certain monetary and pro-social incentives can increase the response rate and also reduce the bias in voluntary reviews.

Our results reinforce the conclusion from the existing word-of-mouth literature that users should know that the distribution of voluntary online reviews can be biased and should be taken with a grain of salt. At the same time, we also demonstrate that bias in reviews can be reduced by using adequate incentives. Our results suggest that websites and government entities alike should experiment with the use of incentives – monetary and pro-social – to increase survey response rates and thereby reduce bias. Such experimentation has the promise to both improve the quality of information at consumers' disposal and allow companies and governments to optimize the informative signal contained in their surveys.

References

- Ariely D (2000) Controlling information flow: Effects on consumers' decision making and preference. *J. Consumer Res.* 27(2):233-248.
- Bandiera O, Barankay I, Rasul I (2010) Social incentives in the workplace. *Rev. of Economic Studies* 77(2):417-458.
- Benson, Alan, Aaron Sojourner, and Akhmed Umyarov. 2015. "Can Reputation Discipline the Gig Economy? Experimental Evidence from an Online Labor Market." SSRN Scholarly Paper ID 2696299. Rochester, NY: Social Science Research Network.
<https://papers.ssrn.com/abstract=2696299>.
- Berger J (2011) Arousal increases social transmission of information. *Psychological Sci.*, 22(7):891-893.
- Berger J, Milkman K (2012) What makes online content viral? *J. Marketing Res.* 49(2):192-205.
- Berger J., Schwartz EM (2011) What drives immediate and ongoing word of mouth? *J. Marketing Res.* 48(5):869-880.
- Broder A, Schiffer S (2006) Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *J. Experimental Psychology: Learning, Memory, and Cognition* 32(4):904-918.
- Burmester M, Kwang T, Gosling SD (2011) Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives Psychological Sci.* 6(1):3-5.
- Buss DM (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sci.* 12:1-49.
- Card, D, Mas A, Moretti E, Saez E (2012) Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Rev.* 102(6):2981-3003.

29 Reducing Bias in Online Reviews

- Chatterjee, P (2001) Online reviews: Do consumers use them? in NA - Advances in Consumer Res. Vol. 28, eds. Mary C. Gilly and Joan Meyers-Levy, Valdosta, GA : Association for Consumer Research, 129-133.
- Chen Z (2017). Social acceptance and word of mouth: How the motive to belong leads to divergent WOM with strangers and friends. *J. Consumer Res.* 44(3):613-632.
- Chintagunta PK, Gopinath S, Venkataraman S (2010) The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Sci.* 29(5):944–957.
- Dana J, Cain DM (2015) Advice versus choice. *Current Opinion Psych.* 6:173-176.
- De Langhe B, Fernbach PM, Lichtenstein DR (2015) Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *J. Consumer Res.* 42:817-833.
- Diehl K (2005) When two rights make a wrong: Searching too much in ordered environments. *J. Marketing* 42(3):313-322.
- Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4 edition. Wiley.
- Dunn EW, Aknin, LB, Norton, MI (2008). Spending money on others promotes happiness. *Science*, 319:1687–1688.
- Dunn EW, Aknin, LB, Norton, MI (2014). Prosocial spending and happiness: Using money to benefit others pays off. *Current Directions Psychological Sci.* 23(1):41-47.
- Filippas, Apostolos, Horton, John, and Joseph Golden. 2017. “Reputation in the Long-Run” Work. Pap., NYU.
- Floyd, K, Freling R, Alhoqail S, Young Cho H, Freling T (2014) How online product reviews affect retail sales: A meta-analysis. *J. Retailing* 90:217–232.
- Forbes 2014: <https://www.forbes.com/sites/jaysondemers/2014/09/09/how-negative-online-company-reviews-can-impact-your-business-and-recruiting/#1409f0c91d9b>

- Fradkin, Andrey, 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. ACM.
- Gneezy U, Meier S, Rey-Biel P (2011) When and why incentives (don't) work to modify behavior. *J. Econom. Perspectives* 25(4):191-209.
- Gneezy U, Rustichini A (2000). A fine is a price. *J. Legal Studies* 29(1):1-17.
- Heyman J, Ariely D (2004) Effort for payment: A tale of two markets. *Psychological Sci.* 15(11):787-793.
- Hu N, Zhang J, Pavlou PA (2009) "Overcoming the J-shaped distribution of product reviews. *Commun. ACM* 52, no. 10 (October 2009): 144–147.
- Huang M, Li P, Meschke F, Guthrie J (2015) Family firms, employee satisfaction, and corporate performance. *Journal of Corporate Finance*, 34: 108-127.
- Huang TK, Ribeiro B, Madhyastha HV, Faloutsos M (2014) The socio-monetary incentives of online social network malware campaigns. *Proceedings of the second ACM conference on Online social network* 259-270.
- Iyengar SS, Wells RE, Schwartz B (2006) Doing better but feeling worse: Looking for the 'best' job undermines satisfaction. *Psychological Sci.* 17(2):143–150.
- Kahneman D, Tversky A (1982) On the study of statistical intuitions. *Cognition*, 11:123–141.
- King RA, Racherla P, Bush VD (2014) What we know and don't know about online word-of-mouth: A review and synthesis of the literature. *J. Interactive Marketing* 28(3):167-183.
- Klein N, Grossman I, Uskul AK, Kraus AA, Epley N (2015) It generally pays to be nice, but not really nice: Asymmetric reputations from prosociality across 7 countries. *Judgment and Decision Making* 10:355-364.
- Klein N (2017) Prosocial behavior increases perceptions of meaning in life. *J Positive Psychology* 12(4):354-361.

31 Reducing Bias in Online Reviews

- Levav J, Reinholtz N, Lin C (2012) The effect of ordering decisions by choice-set size on consumer search. *J. Consumer Res.* 39:585-599.
- Lu SF, Rui, H (2017) Can we trust online physician ratings? Evidence from cardiac surgeons in Florida. *Management Sci.* (published online June 13, 2017).
- Luca M (2016). Reviews, reputation, and revenue: The case of Yelp.com. *Harvard Business School NOM Unit Working Paper No. 12-016.*
- Mayzlin D, Dover Y, Chevalier J (2013) Promotional reviews: An empirical investigation of online review manipulation. *American Economic Rev.* 104(8):2421-2455.
- Meyer, BD, Mok WKC, Sullivan JX (2015) Household surveys in crisis. *J. Economic Perspectives* 29 (4):199–226.
- Moe WW, Trusov M (2011) The value of social dynamics in online product ratings forums. *J. Marketing Res.* 49:444–456.
- Pallais, Amanda. 2014. “Inefficient Hiring in Entry-Level Labor Markets.” *American Economic Review* 104 (11):3565–99. <https://doi.org/10.1257/aer.104.11.3565>.
- Payne JW, Bettman JR, Johnson EJ (1988) Adaptive strategy selection in decision making. *J. Experimental Psychology* 14(3):534–52.
- Philipson, Tomas. 2001. “Data Markets, Missing Data, and Incentive Pay.” *Econometrica* 69 (4): 1099–1111. <https://doi.org/10.1111/1468-0262.00232>.
- Senecal S, Nantel J (2004) The influence of online product recommendations on consumers’ online choices. *J. Retailing* 80:159-169.
- Simonson I (2016) Imperfect progress: An objective quality assessment of the role of user reviews in consumer decision making: A commentary on de Langhe, Fernbach, and Lichtenstein. *J. Consumer Res.* 42(6):840-845.
- Urbany JE, Dickson PR, Wilkie WL (1989) Buyer uncertainty and information search. *J. Consumer Res.* 16(2):208-215.

Variable	Mturk			Glassdoor		
	Observations	Mean	SE	Observations	Mean	SE
Age	639	35.462	10.887	188,623	34.314	10.550
Female	639	0.518	0.500	188,623	0.419	0.493
More than 1000 employees	639	0.421	0.494	188,623	0.703	0.457
Education (years)	639	15.283	2.006	188,623	15.505	1.343
Tenure (years)	639	3.563	4.344	188,623	3.543	4.790

Table 1: Summary statistics: Mturk vs. Glassdoor datasets

34 Reducing Bias in Online Reviews

	Rating	Rating	Is 1 star	Is 1 star	Is 5 stars	Is 5 stars
	(1)	(2)	(3)	(4)	(5)	(6)
Voluntary	0.0240*** (0.00800)	0.0345*** (0.00798)	0.0190*** (0.00169)	0.0143*** (0.00169)	0.0467*** (0.00301)	0.0432*** (0.00302)
Age		-0.00881*** (0.000299)		0.00225*** (7.05e-05)		-0.000635*** (0.000104)
Female		-0.120*** (0.00575)		0.0147*** (0.00128)		-0.0273*** (0.00210)
More than 1000 employees		-0.285*** (0.00644)		- 0.00934*** (0.00141)		-0.174*** (0.00242)
Education (years)		0.0514*** (0.00220)		- 0.00843*** (0.000524)		0.00897*** (0.000756)
Tenure (years)		0.00250*** (0.000636)		- 0.00215*** (0.000142)		-0.00221*** (0.000232)
Constant	3.611*** (0.00738)	3.349*** (0.0371)	0.0639*** (0.00155)	0.130*** (0.00875)	0.262*** (0.00279)	0.290*** (0.0128)
Observations	188,623	188,623	188,623	188,623	188,623	188,623
R-squared	0.000	0.022	0.001	0.011	0.001	0.035

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Glassdoor selection bias: more polarized ratings

	No incentive	No incentive	Nonspecific prosocial	Nonspecific prosocial	Monetary high	Monetary high
	(1)	(2)	(3)	(4)	(5)	(6)
Choice	-0.574** (0.225)	-0.660*** (0.228)	0.050 (0.190)	-0.017 (0.192)	-0.295 (0.200)	-0.290 (0.216)
Age		-0.004 (0.010)		0.006 (0.008)		-0.009 (0.010)
Female		-0.034 (0.225)		0.284 (0.176)		0.223 (0.207)
More than 1000 employees		-0.473** (0.223)		-0.464** (0.196)		-0.319 (0.196)
Education (years)		-0.026 (0.048)		-0.017 (0.041)		-0.053 (0.061)
Tenure (years)		-0.007 (0.022)		-0.053* (0.029)		0.005 (0.024)
Constant	4.019*** (0.117)	4.846*** (0.709)	4.019*** (0.117)	4.315*** (0.683)	4.019*** (0.117)	5.171*** (1.023)
Observations	90	90	98	98	101	101
R-squared	0.077	0.144	0.001	0.157	0.022	0.074

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Impact of selection and incentives on average ratings in MTurk sample (relative to no incentive, forced review)

	Positive prosocial	Positive prosocial	Negative prosocial	Negative prosocial	Monetary low	Monetary low
	(1)	(2)	(3)	(4)	(5)	(6)
Choice	0.043 (0.252)	-0.048 (0.267)	-0.047 (0.215)	-0.042 (0.195)	-0.289 (0.225)	-0.236 (0.232)
Age		-0.006 (0.012)		0.002 (0.008)		0.004 (0.009)
Female		0.347 (0.255)		0.288 (0.201)		0.034 (0.228)
More than 1000 employees		0.140 (0.262)		-0.616*** (0.201)		-0.366* (0.210)
Education (years)		0.081 (0.063)		0.031 (0.046)		0.011 (0.049)
Tenure (years)		0.037 (0.035)		-0.013 (0.020)		-0.005 (0.022)
Constant	3.686*** (0.139)	2.391** (1.163)	4.019*** (0.117)	3.611*** (0.830)	4.019*** (0.117)	3.851*** (0.753)
Observations	88	88	89	89	91	91
R-squared	0.000	0.056	0.001	0.136	0.020	0.055
Robust standard errors in parentheses						
*** p<0.01, ** p<0.05, * p<0.1						

Table 4: Impact of selection and incentives on average ratings for incentives that did not increase response rates in the MTurk sample (relative to no incentive, forced review)

37 Reducing Bias in Online Reviews

	Rating	Rating
	(1)	(2)
Monetary high	-0.001 (0.180)	-0.002 (0.180)
Monetary low	0.022 (0.187)	0.050 (0.191)
Negative prosocial	-0.325 (0.198)	-0.319 (0.199)
Nonspecific prosocial	-0.332* (0.190)	-0.345* (0.194)
Positive prosocial	-0.332* (0.181)	-0.331* (0.182)
Choice*		
Control (no incentive)	-0.574** (0.225)	-0.574** (0.227)
Monetary high	-0.294 (0.212)	-0.297 (0.215)
Monetary low	-0.311 (0.241)	-0.323 (0.243)
Negative prosocial	0.278 (0.241)	0.273 (0.236)
Nonspecific prosocial	0.382* (0.212)	0.402* (0.213)
Positive prosocial	0.043 (0.252)	0.029 (0.253)
Constant	4.019*** (0.117)	3.645*** (0.378)
Control		X
Observations	546	546
R-squared	0.030	0.039

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5: Framing effects: do incentives affect forced average ratings in MTurk sample?

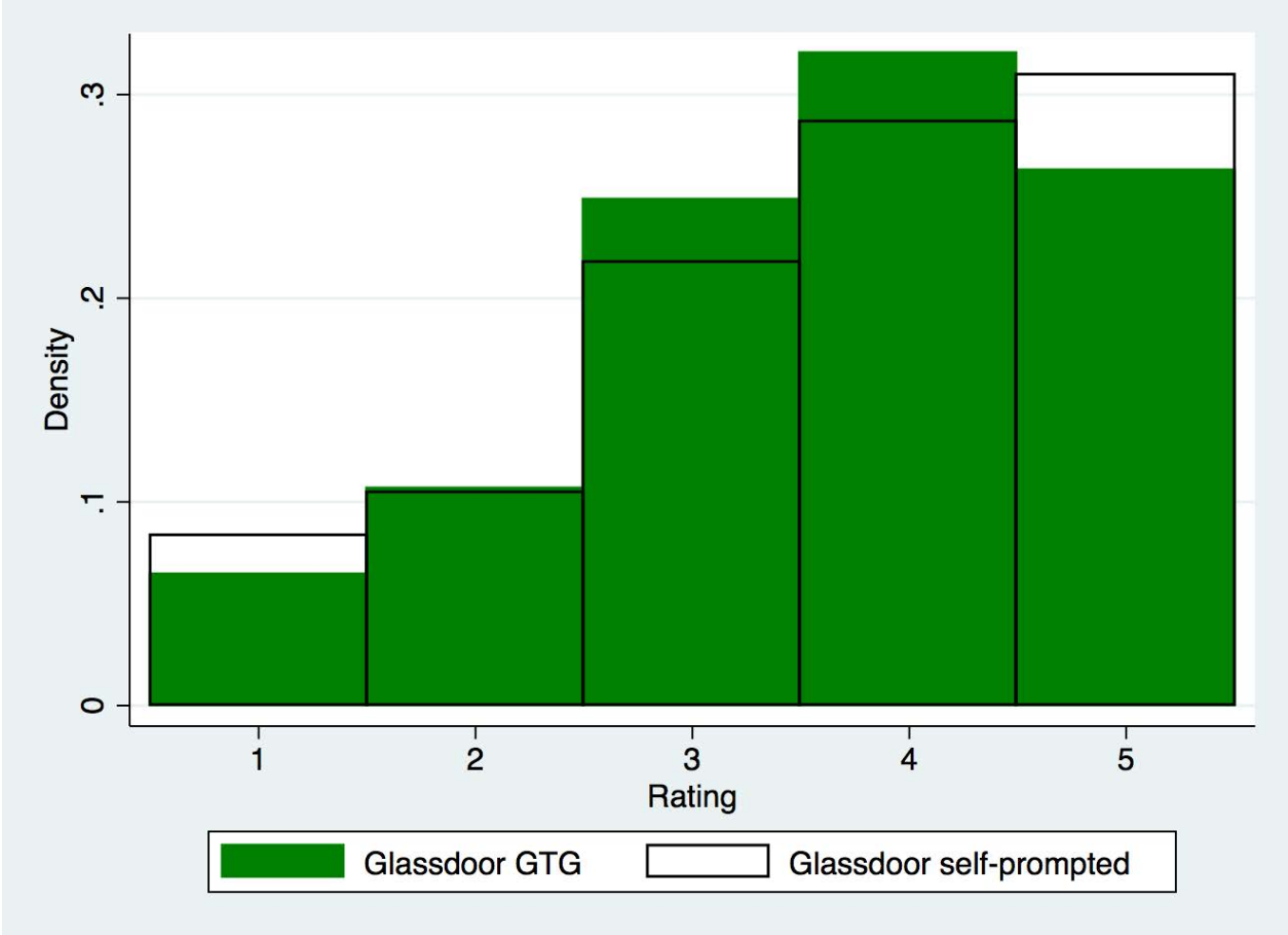


Figure 1: Glassdoor GTG vs. voluntary reviews

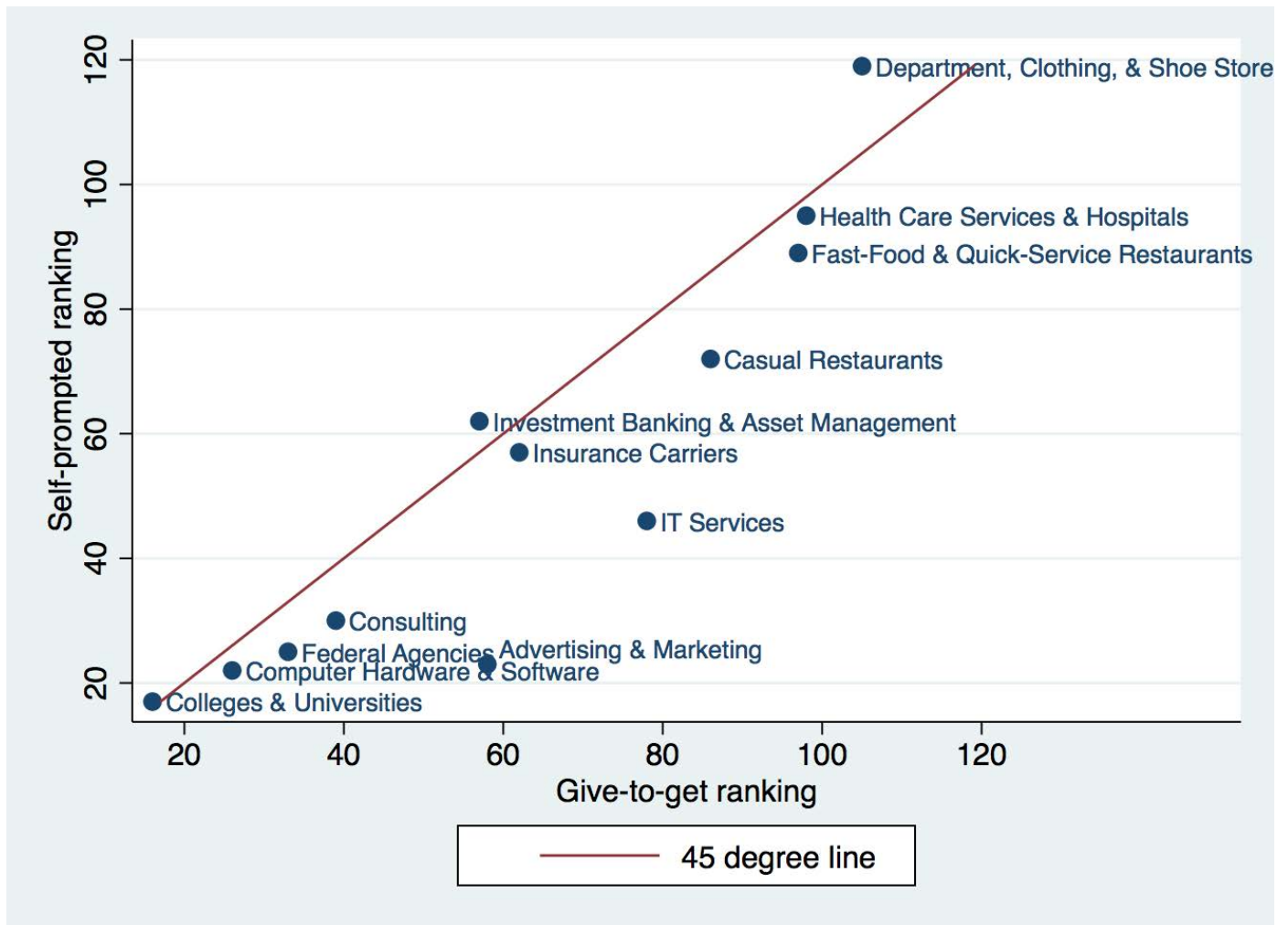


Figure 2: Glassdoor: changes in rankings of frequent industries due to bias

41 Reducing Bias in Online Reviews

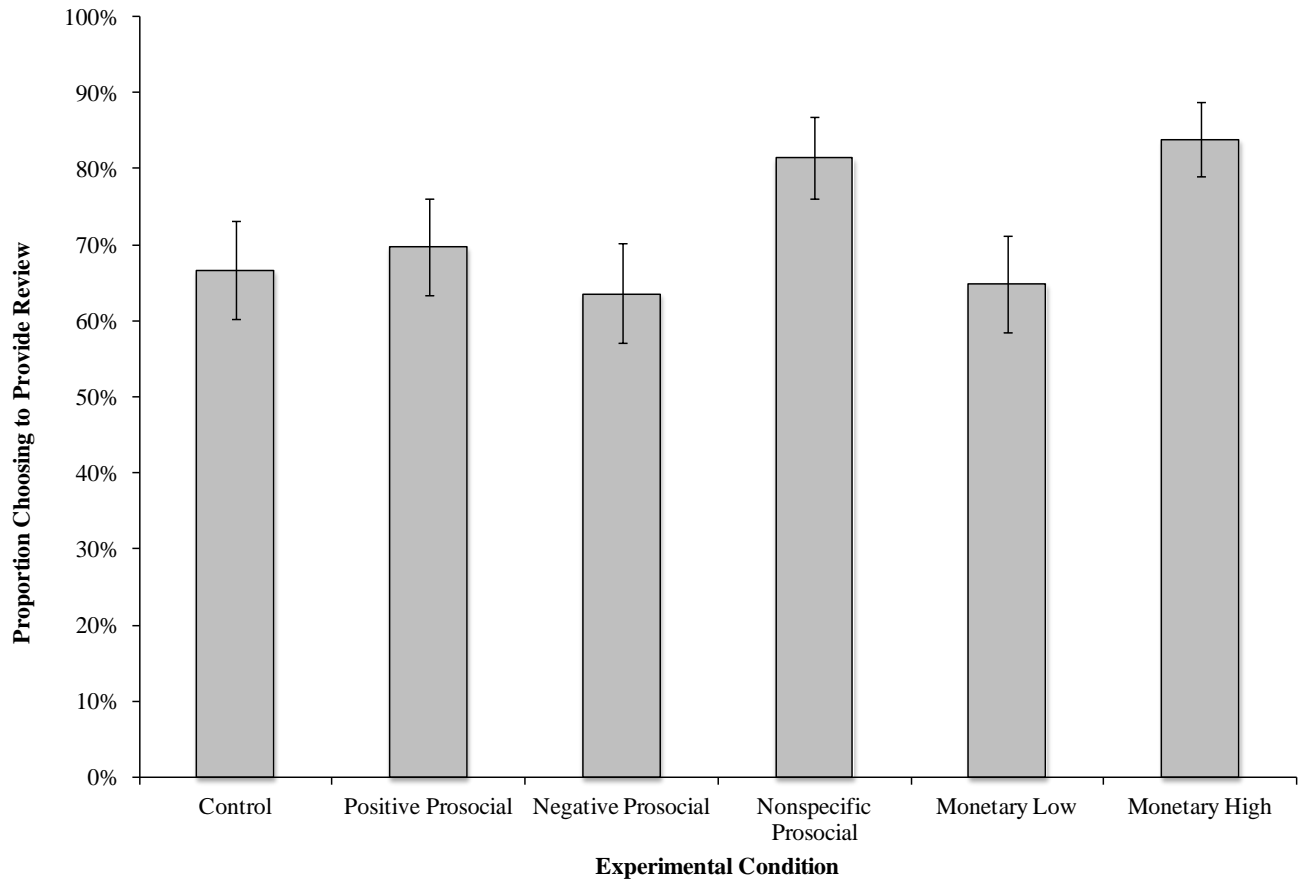


Figure 3: MTurk experiment: efficacy of different incentives in increasing response rates

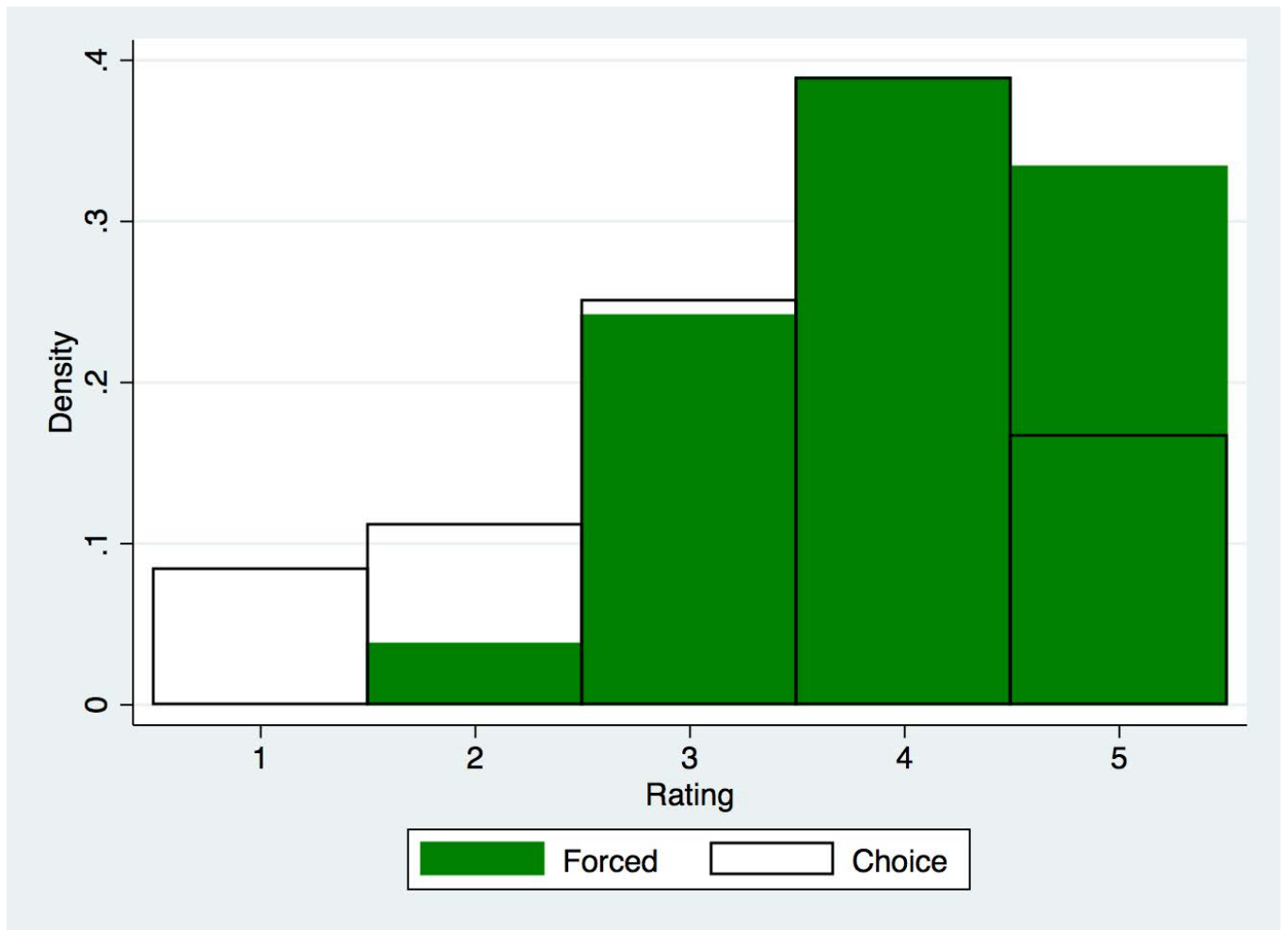


Figure 4: Bias in reviews in the absence of incentives in the MTurk experiment

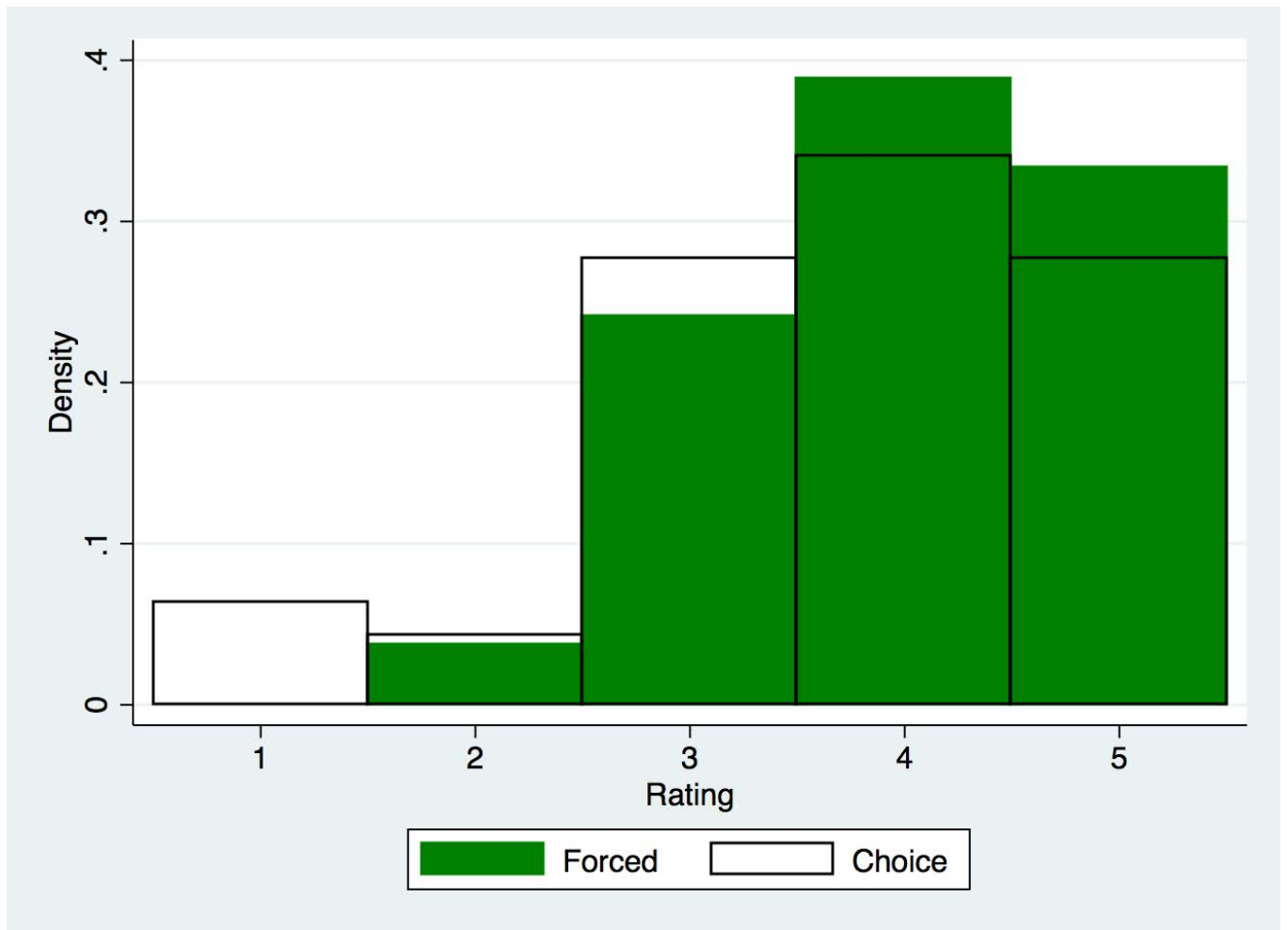


Figure 5: No bias in reviews with high monetary incentive (75% payment increase) in the MTurk experiment

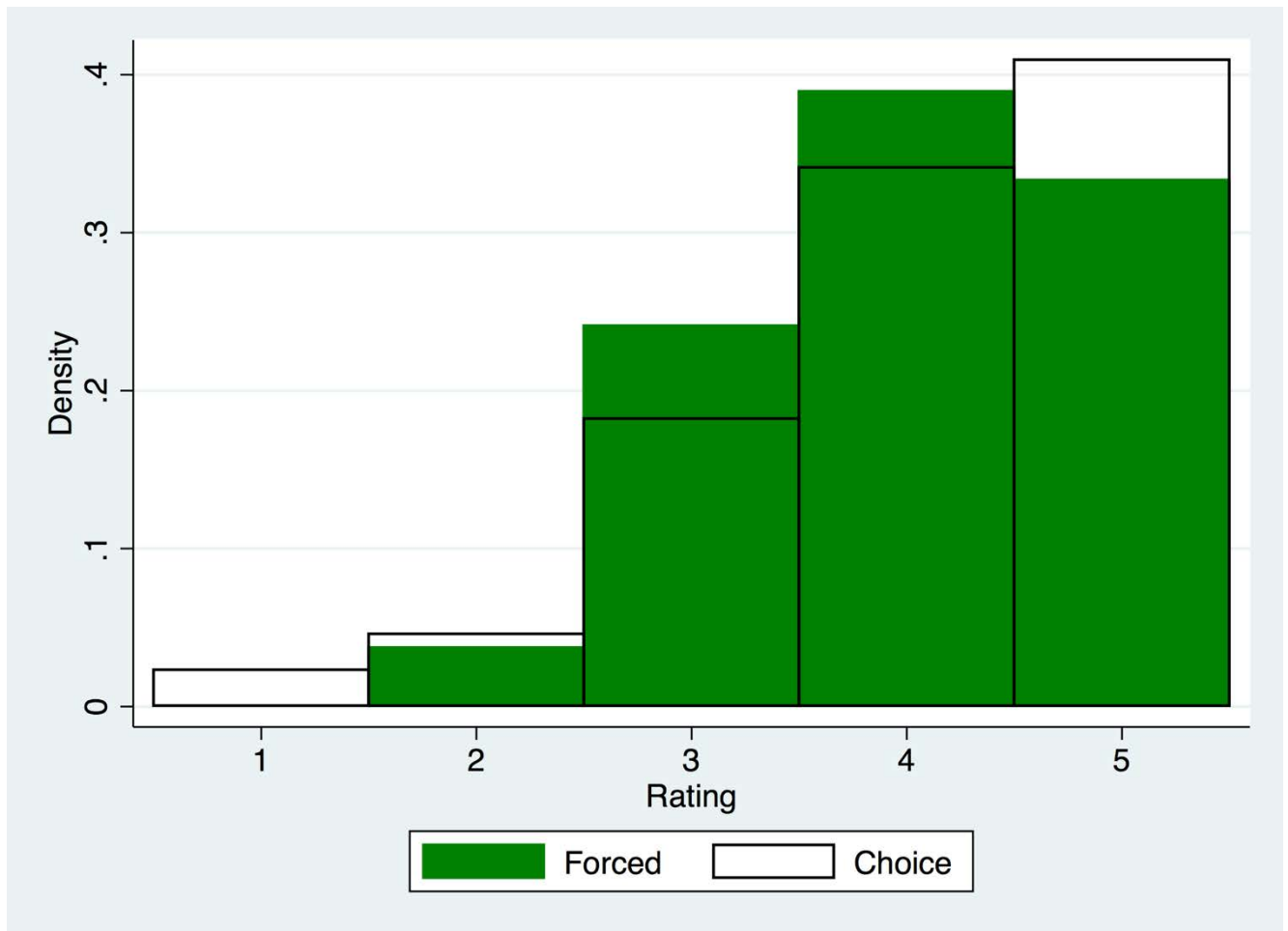


Figure 6: No bias in reviews with nonspecific prosocial incentives in the MTurk experiment

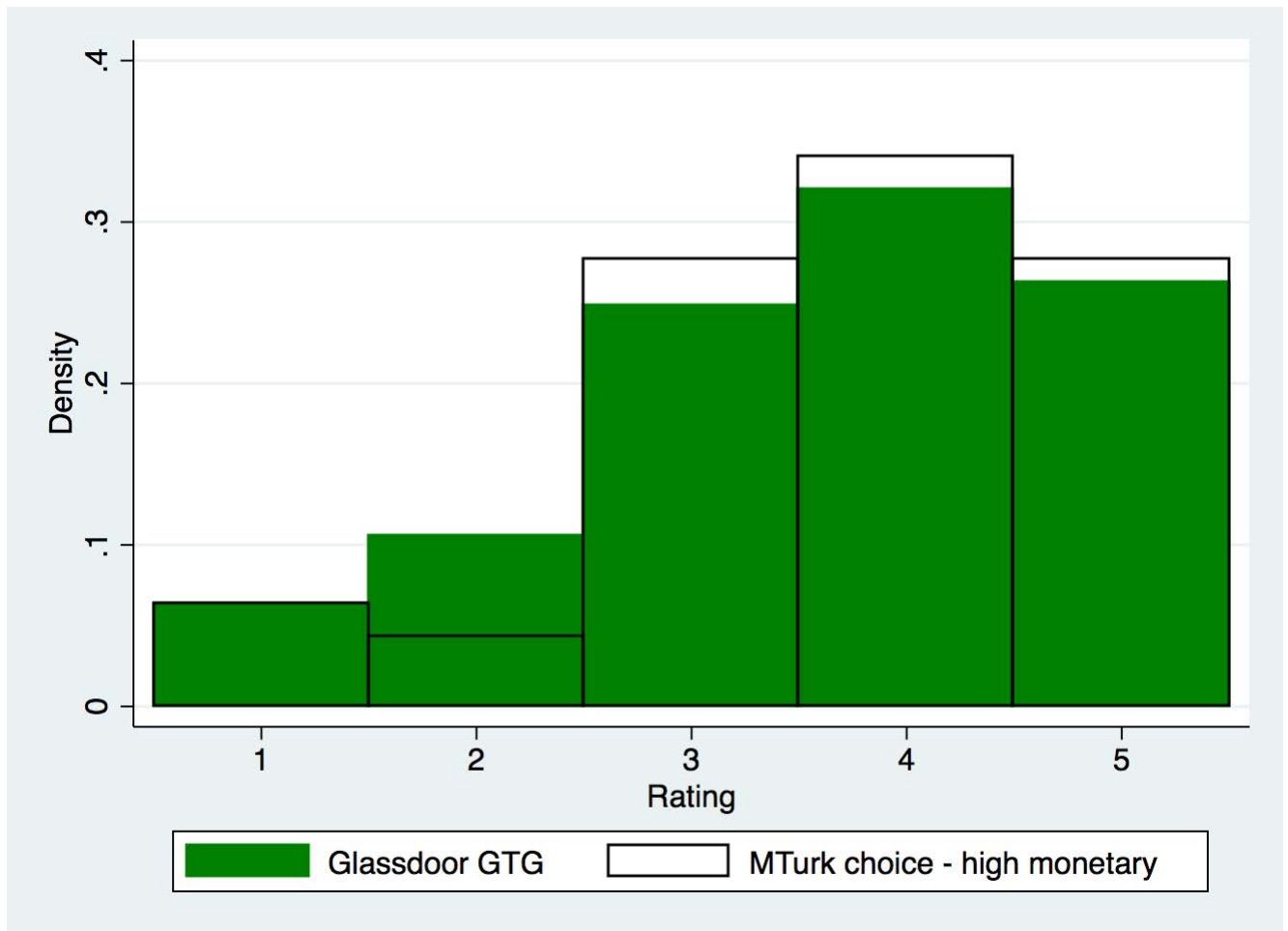


Figure 7: Glassdoor GTG vs. Mturk choice, high monetary: similar distributions