Incentives can reduce bias in online employer reviews

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

This paper is in the "Revise & Resubmit" stage in the *Journal of Applied Psychology*

Abstract

Online reviews are a powerful means of propagating the reputations of employers. However, existing research suggests that online reviews often suffer from selection bias—people with extreme opinions are more motivated to select into sharing them than people with moderate opinions, resulting in biased distributions of ratings. 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 find that voluntary ratings have a different distribution from incentivized ratings. 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 cannot 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 selection 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.

Keywords: online reviews, job choice, incentives, selection bias, reputation

Incentives can reduce bias in online reviews

In the age of the internet, an employer's reputation is almost never a blank slate. Job seekers can easily find employer reviews because employees have gone to the trouble of posting their opinions online. As in other areas of the economy, these online reviews are an important decision aid (Card et al., 2012; Chatterjee, 2001; Chintagunta et al., 2010; Floyd et al., 2014; Luca, 2016; Mayzlin et al., 2013; Moe & Trusov, 2011; Senecal & Nantel, 2004). Accordingly, organizations devote resources to managing how they are viewed in these channels (Cable & Yu, 2006).

However, emergent research suggests that these widely-used online reviews may suffer from selection bias (Hu et al., 2009), defined as reviews that are not representative of the actual experience of employees at a given company. Selection bias is thought to occur because the motivation to publicly share opinions may depend on how extreme these opinions are. Feeling strongly—either positively or negatively—about an employment experience may increase the motivation to post reviews. Thus, non-representative extreme reviewers may be more likely to share reviews than more representative moderate ones. In this article, we propose that providing incentives for reviewing can reduce this selection bias and discuss why this method is preferable even if applicants can recognize review bias when assessing potential employers.

Specifically, we use data from Glassdoor, an employer review website used by a substantial proportion of applicants according to industry surveys (Childress, 2018; DeMers, 2014). We find that incentivized employer reviews exhibit less polarized distributions than non-incentivized reviews, and that this difference changes the relative rankings of industries between which applicants might plausibly choose. We complement these observational data with a controlled experiment in which we manipulate participants' ability to opt out of reviewing their employer. We treat forced reviews as the closest measure of the "true" distribution of employer ratings and measure the degree of bias relative to these forced reviews. We find that non-incentivized reviews are negatively biased. Our experiment also finds that two types of incentives 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 prosocial incentive (a request to consider how one's review will be helpful to others).

Existing research conceptualizes selection bias in online reviews as a combination of "employment bias"[1] and "reporting bias" (Hu et al., 2009). Employment bias originates from an employee's decision to join a company, which renders his or her evaluation of that company more positive than an objective observer's. Employment bias is unavoidable in online reviews because only current and past employees can draw on their experiences when reviewing their employers. Meanwhile, reporting bias is undesirable because it suggests that the propensity to review an employer is affected by one's experience, whereby extreme opinions are more likely to be reported than moderate ones.

Consistent with the selection bias hypothesis, existing research finds that the distributions of online reviews of retail products, motion pictures, books, and medical physicians are "J-shaped," wherein extreme positive reviews are more common than extreme negative reviews, which are in turn more common than moderate reviews (Chevalier & Mayzlin, 2006; Hu et al., 2009; Liu, 2006; Lu & Rui, 2017). These skewed distributions contrast with "bell curve" normal distributions found in performance appraisal contexts (Beck, Beatty, & Sackett, 2014) and obtained in randomized experiments in which participants do not have a choice but are instead forced to provide reviews of products (Hu et al., 2009).

Why are distributions of reviews skewed in these ways? One reason is that people with moderate opinions may be less motivated to provide reviews than people with extreme opinions. Existing research on information sharing suggests that emotionally charged information is more likely to be shared than non-arousing information (Berger, 2011; Berger & Milkman, 2012; Heath, 1996; Heath et al., 2001; Peters et al., 2009; Rimé, 2009). Accordingly, extreme employment experiences—negative and positive—may be more likely to create arousal and emotion and therefore be shared as online reviews than moderate employment experiences.

[1] The original term is "purchasing bias" and refers to a consumer's decision to purchase a given product of service. We adapted the term so that it is more relevant for organizational research.

If moderate employment experiences result in a motivational deficit to post online reviews, then incentives may compensate for this deficit and ultimately reduce selection bias. Psychological research suggests that incentives can tap into not only self-interest but also prosocial interest (Barclay, 2004; Dunn et al., 2008; Gintis et al., 2003; Fehr and Schmidt, 1999; Klein, 2017; Klein et al., 2015). Prosocial incentives[2] can be effective in this context because people can be made to construe online reviews as a public good that benefits other people. In our experiment, we therefore provide different levels of monetary incentives and different kinds of prosocial incentives in order to understand the kinds and magnitudes of incentives needed to reduce selection bias in employer reviews. Overall, we test four hypotheses:

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

Hypothesis 2: Differences between the distributions of incentivized and non-incentivized reviews on Glassdoor lead to differences in the perceptions of different industries. The ranking of industries as measured through average employer ratings differs when using incentivized reviews compared to voluntary reviews on Glassdoor.

Hypothesis 3: In line with the Glassdoor results, the distribution of forced reviews in our controlled experiment differs from the distribution of self-selected reviews.

As mentioned, our experiment tested the efficacy of different types of incentives. Two of the incentives were monetary and differed in their magnitude. Based on past research (Gneezy, Meier, & Rey-Biel, 2011), we predicted that people will be sensitive to the magnitude of monetary incentives and will be more likely to post reviews in exchange for more money. The other three incentives were prosocial, focusing on different facets of helping others by contributing a review. First, the nonspecific prosocial incentive framed employer reviews as a way to help others make better employment decisions. Second, the negative prosocial incentive

[2] Prosocial incentives are sometimes called "social incentives" in some literatures. Here we focus on incentives that tap into the desire to help other people, and so the former term is more precise.

framed employer reviews as a way to protect employees from the worst employers to work for.[3] Third, to test for potential asymmetries in the valence of prosocial incentives 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 any incentive that increases the motivation to provide employer reviews should also reduce selection bias, because it would disproportionately encourage "middle-of-the-road" reviewers to share reviews. Thus:

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

Our work makes a number of contributions to existing research on reputation, recruitment, and applicant behavior. First, we help answer the call by organizational researchers for a better understanding of the effects of technology on recruitment (Ployhart, Schmitt, & Tippins, 2017), specifically in promoting our understanding of the general characteristics that affect the usefulness of online reviews (McFarland & Ployhart, 2015). Second, whereas much of existing research in this area uses hypothetical websites as stimuli (Allen, Mahto, & Otondo, 2007; Dineen, Ash, & Noe, 2002; Dineen & Noe, 2009; Dineen, Ling, Ash, & DelVecchio, 2007), here we add the realistic environment of Glassdoor reviews. Third, whereas prior studies discuss selection bias in online reviews of products and services (Chevalier & Mayzlin, 2006; Hu et al., 2009), here we focus on a market relevant to work and organizations.

The present work also has important implications for recruitment and applicant behavior. Existing research suggests that organizational image affects applicant attraction (Bretz & Judge, 1998; Cable & Yu, 2006; Chapman et al., 2005; Gatewood et al., 1993; Ryan et al., 2000). Specifically, organizational desirability as expressed in online ratings increases applicant attraction (Benson, Sojourner, & Umyarov, 2015). Selection bias in online reviews can mislead applicants because it can increase or decrease the average rating of an employer in unexpected ways. If extreme negative experiences are overly represented in online reviews, then applicants

[3] 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.

may be led to believe that the organization is less desirable than would be the case in a world without selection bias. If extreme positive experiences are overly represented, then the opposite will occur. Optimal assessments of organizations from the applicant's perspective partly depend on having accurate reputational information about the organization (Caldwell & O'Reilly, 1985). Perceptions of organizational image can solidify at the beginning of the job search process (Kappes, Balcetis, & De Cremer, 2018; Swider, Zimmerman, & Barrick, 2015), which is often when applicants consult online reviews. For these reasons, mitigating selection bias can be important for improving recruitment and job selection decisions.

Existing research on online reviews tested the effects of monetary incentives on the *average* reviews of products and services (Burtch et al., 2018; Fradkin et al., 2015; Stephen et al., 2012; Wang et al., 2012). Here we measure the *distribution* of reviews to assess the degree to which incentives affect selection bias (and consequently the average review). The one exception is work done by Khern-am-nuai et al. (2016), who find statistically similar distributions of reviews before and after an American retailer introduced incentive in the form of loyalty points equivalent to 50 cents per review. However, the current work remains necessary because (1) employment markets differ from retail products in many ways and (2) the effects of prosocial incentives in this space are unknown.

Study 1: Glassdoor

Sample and Method. Glassdoor is a leading employer review website. People must be current or past employees of an employer to review it. The employer ratings scale at Glassdoor 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 or non-incentivized reviews. In contrast, Glassdoor also uses an alternative method for generating employer reviews. 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), a first-time visitor must submit a review 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 to 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.[4]

We use a sample of 188,623 U.S. employer reviews published on Glassdoor[5] from 2013 to 2016[6]. We keep in the sample only the most recent review of a person's current employer. To be able to control for bias due to different characteristics among GTG vs. voluntary review, we keep only Glassdoor users for which we have age, gender, and highest education. and the reviewed employer belongs to a known industry, geographic state, and has a known number of employees. Table 1 shows summary statistics for the Glassdoor sample of reviews, as well as for the MTurk sample we used in the subsequent experiment.

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

[4] More information about the company's GTG policy is available at http://help.glassdoor.com/article/Give-to-get-policy/en_US/. 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.

[5] These data can be requested from Glassdoor by other researchers for the purposes of replication by contacting pr@glassdoor.com.

[6] Of the GTG reviews in our sample, 8% were from 2013, 31% from 2014, 35% from 2015 and 26% from 2016.

of GTG ratings (Figure 1). The difference between the two distributions is statistically significant, Pearson χ^2 (4) = 427.22, $p < 0.001$.

OLS regressions (Table 2) show that on average voluntary reviews tend to be slightly more positive: after controlling for observables, we find that the average voluntary review is higher by 0.035 stars on average (column 2). More importantly, voluntary reviews are also *more extreme* than incentivized reviews. After controls, voluntary reviews are 1.4 percentage points more likely to be one star (column 4), and 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. Replacing OLS regressions with an ordered logit model finds substantively similar results. In further analysis available upon request, we show that these results are robust to controlling for observables by propensity score matching and cross-validation (these methods are used in place of adding observable characteristics in a linear fashion). These results confirm our Hypothesis 1. The distribution of voluntary reviews is different from the distribution of reviews incentivized via the GTG policy.

Voluntary reviews are more polarized than incentivized reviews, but does this matter in practice? When people browse Glassdoor, they typically aim to find information that could help them decide which employer to work for. Therefore, if this apparent selection bias affects the ranking of employers, it is likely to also affect employment choices.

It turns out that the difference between the distribution of voluntary reviews and GTG reviews is not innocuous. Instead, it can substantially affect the ranking of industries[7] 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 points where 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-

[7] One could rank individual employers (rather than industries). However, we did not extract these data to protect employers' identifiable information. Moreover, we did not have enough granular data to pursue this analysis. Finally, many job seekers do consider entering a new industry (either beginning their career or considering a mid-career transitions) and so compare the reviews of different industries.

degree line are ranked worse under GTG than under voluntary reviews, and the opposite is true for industries above the 45-degree line. To the extent that incentivized GTG ratings are more accurate, the consulting industry is more desirable (lower rank) than the advertising & marketing industry. Yet, if we rely only on voluntary reviews, advertising & marketing appears more desirable than consulting. Similarly, the insurance industry appears more desirable under voluntary ratings than the investment banking industry, but these rankings are reversed under GTG ratings. These industries draw on candidates with similar backgrounds, and so a job seeker might plausibly choose between them. These results broadly confirm our Hypothesis 2.

An important limitation of this analysis is that observational data alone do not 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. Thus, we turn to an experiment on MTurk where we measure both the "true" underlying distribution of employer reviews and the self-selected distribution when people have the option not to provide a review. The experiment also tests which types of incentives are more effective in getting people to review and mitigate bias from self-selection.

Study 2: Controlled Experiment

Participants. Participants ($N = 639$) were recruited from Amazon's Mechanical Turk (MTurk) for a five-minute study advertised as a 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 main notable difference between the Glassdoor sample and the Amazon MTurk sample based on the available demographic data[8] is the greater representation of large employers on Glassdoor.

Procedure. The experiment included two factors and 12 experimental conditions, resulting in a 2(Choice vs. Forced Review)× 6(Incentive: None, High Monetary, Low Monetary, Nonspecific Prosocial, Positive Prosocial, Negative Prosocial) between-subjects design. Participants were 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, and told that 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, and is theoretically similar to Glassdoor voluntary reviews. In the Forced conditions, participants were 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 participation). Thus, the Forced condition is a proxy for what the true underlying distribution of reviews looks like without self-selection. Note that there is no theoretically equivalent condition in the Glassdoor sample to this Forced condition because in the real world people are never forced to provide online 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

[8] 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.

reviewed their main employer (a 75% increase to base compensation). These monetary incentives are low in terms of raw amounts but are commensurate with monetary incentives used in previous research investigating the willingness of MTurk workers to provide online reviews (Burtch et al. 2018). Moreover, using low raw monetary incentives represents a conservative test of whether they can increase people's willingness to review their employers even for relatively low raw amounts.[9] The other three incentives were prosocial, 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 employment decisions. 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.

[9] It is also important to note that from a practical perspective, companies cannot afford to pay high amounts for short reviews that require completing only a Likert scale. Thus, if low monetary incentives fail to motivate more reviews, then payment for reviews is unlikely to be a viable business strategy.

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

Results

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 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, χ^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 increase 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 effort expenditure is sensitive to the magnitude of monetary incentives (Gneezy et al., 2011).

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 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, Pearson $\chi^2(4) = 8.54$, p = 0.074. As Figure 4 shows, self-selected non-incentivized reviews exhibited a downward bias in employer ratings. This result is consistent with our Hypothesis 3. When left to make their own choices, participants provided 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 for this finding in the General Discussion.

Bias in Employer Reviews with Incentives. Next, we examined whether the different incentives 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), Pearson χ^2s < 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. 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: indeed the coefficient on Choice in Table 3, columns 3-6, is not statistically significantly different from zero. These results suggest that these two types of incentives not only increased response rates, but also resulted in review distributions that more closely mirrored the true distribution (i.e., the distribution in the Forced response without incentives), consistent with our Hypothesis 4.

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 (the coefficient on Choice is not statistically significantly different from zero). Thus, although the low monetary, positive prosocial, and negative prosocial incentives failed to motivate more responses, they nevertheless eliminated the selection effects found in voluntary non-incentivized 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 bias. 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 increased response rates, but also reduced selection bias.

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 Negative Prosocial incentive might bring to mind bad employers and thus might increase reported negative 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 coefficients of the different incentives within the Forced condition with incentives relative to the Forced No Incentive condition (first set of coefficients in Table 5); the selection effects are measured by the interaction between the Choice condition and these incentives (coefficients on incentives in lines below "Choice ×" in Table 5 are interaction effects between Choice and the specific incentive).

Relative to the Forced No Incentive condition, the Nonspecific Prosocial incentive and the Positive Prosocial incentive 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 the make a wrong decision looms large in people's minds (Dana & 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 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 in the nonspecific prosocial incentive were similar, so these effects cancelled 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 to provide reviews.

General Discussion

Online reviews may not always be reliable because workers with extreme experiences are more likely to share reviews than workers with moderate ones. A key methodological innovation of our paper is in providing unbiased reviews in our experiment (Forced condition, no incentives). Approximating the "true" distribution of online reviews without self-selection is central to measuring selection bias. Hu et al. (2009) use a strategy similar to ours in comparing the ratings of a CD on Amazon to the ratings of participants forced to review the CD. Other studies attempt to resolve this problem by comparing laypeople's reviews to expert reviews, but this strategy leaves gaps because experts and laypersons use different criteria in their evaluations (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 strategic behavior, as employers may encourage employees to provide positive reviews on Glassdoor.[10] If employers can exert influence over employees, this will increase positive extremes in the distribution of voluntary reviews. In contrast, employers cannot exert influence in the MTurk sample, potentially eliminating positive extremes that would otherwise exist. Nevertheless, both the Glassdoor and MTurk datasets are consistent in two important respects: Both reveal evidence of selection bias in non-incentivized reviews, and both reveal that incentives can reduce this bias.

One worry for our Glassdoor findings is that the difference between incentivized and non-incentivized distributions are due not to incentives but rather to a large sample size that can render even insubstantial effects statistically significant. To test this possibility, we use nearest neighbor matching to find 643 Glassdoor observations most similar in characteristics to the MTurk data ($N = 639$) and compare the review distributions of incentivized and non-incentivized Glassdoor reviews within this subsample (Figure 7). The difference between the two

[10] 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.

distributions is (marginally) significant, $\chi^2(4) = 8.022$, $p = 0.091$. Note that this result is obtained despite the non-controlled and noisy environment of real-world online reviews, suggesting that selection bias seen in the full Glassdoor data is not solely due to a large sample size.

Selection bias reduces the accuracy of online reviews because it can modify the mean ratings of employers in unexpected ways. If the proportion of extreme negative reviews outweighs that of extreme positive reviews, then selection bias will detract from a company's average rating. If the proportion of extreme positive reviews outweighs that of negative extremes, the opposite will occur. Selection bias can thus mislead applicants by shifting their perceptions of companies to be either more positive or more negative than would be the case without selection bias.

Applicants' perceptions of an organization's desirability matter not only for their inclination to apply for jobs (Bretz & Judge, 1998; Chapman et al., 2005), but also for calibrating applicants' expectations about what it is like to work there. Misguided expectations can derail job performance because applicants have to adjust to a reality different from what they anticipated (Kappes et al., 2018; Swider et al., 2015; Uggerslev et al., 2012).

Selection bias in online reviews also matters in terms of applicants' preferences for representative distributions. We conducted a follow-up study to investigate this point. This study ($N = 101$) asked participants on MTurk to assess how likely people with moderate experiences, negative experiences, and positive experiences at a workplace are to post an online review. Participants were also asked how representative each of these three types of reviews would be of what it was like to work at a given company. Participants correctly recognized that employees who had either negative experiences ($M = 5.50$, $SD = 1.64$) or positive experiences ($M = 5.06$, $SD = 1.38$) at a workplace were more likely to post an online review than employees who had moderate experiences ($M = 3.85$, $SD = 1.64$), *paired t*s > 6.69, *p*s < .001, *d*s > .66. However, participants also believed that moderate reviews are more representative of what it is like really like to work at a company ($M = 5.13$, $SD = 1.37$) compared to negative reviews ($M = 4.32$, $SD =$

1.66) and positive reviews ($M = 4.78$, $SD = 1.34$), *paired* ts > 1.92, ps < .058, ds > .19. Thus, people appear to have correct intuitions about the causes of selection bias in online reviews but also realize that extreme reviews are non-representative.

However, correcting for selection bias in online reviews is more difficult for individual applicants than merely recognizing that it exists. The first difficulty is that selection bias can either increase or decrease the mean rating of employers, and so the direction of correction is unclear. Moreover, online reviews of different employers in different industries may suffer from different degrees of selection bias (see our Figure 2), complicating on-the-spot correction. Finally, mentally correcting for selection bias likely requires time and cognitive resources, which applicants may not always have when browsing online reviews. Thus, relying on the applicants themselves to correct for selection bias in online reviews may not be sustainable.

A better approach to ensuring that applicants have the right information about employers may be changing the design of online review platforms so that applicants can easily see not only the mean rating of an employer but also the distribution of reviews.[11] This approach is in line with behavioral science, which advocates changing the decision environment to accommodate for people's cognitive limitations and biases (Thaler & Sunstein, 2008). However, changing the decision environment would not reduce the selection bias inherent in online reviews, but only mitigate its adverse effects by allowing applicants to mentally correct for it. For this reason, the possibility that incentives can reduce selection bias in the first place is important for creating sustainable improvements in how organizational reputation is communicated without having to rely on individual applicants making mental corrections. More accurate information will benefit both job seekers and employers because it will likely improve the matching process involved in recruitment.

[11] The distribution of employer reviews is accessible on Glassdoor, but only after clicking to get additional details about the rating.

References

Allen, D. G., Mahto, R. V., & Otondo, R. F. (2007). Web-based recruitment: Effects of information, organizational brand, and attitudes toward a Web site on applicant attraction. *Journal of Applied Psychology, 92,* 1696–1708.

Barclay, P. (2004). Trustworthiness and competitive altruism can also solve the “tragedy of the commons.” *Evolution and Human Behavior, 25*, 209-220.

Beck, J. W., Beatty, A. S., & Sackett, P. R. (2014). On the distribution of job performance: The role of measurement characteristics in observed departures from normality. *Personnel Psychology, 67,* 531-566.

Benson, A., Sojourner, A., & Umyarov, A. (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.

Berger, J. (2011). Arousal increases social transmission of information. *Psychological Science, 22*, 891-893.

Berger, J., & Milkman, K. (2012). What makes online content viral? *Journal of Marketing Research, 49*, 192-205.

Bretz, R. D., & Judge, T. A. (1998). Realistic job previews: A test of the adverse self-selection hypothesis. *Journal of Applied Psychology, 83,* 330-337.

Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (2018). Stimulating online reviews by combining financial incentives and social norms. *Management Science.*

Burmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science. 6*, 3-5.

Cable, D. M., & Yu, K. Y. T. (2006). Managing job seekers’ organizational image beliefs: The role of media richness and media credibility. *Journal of Applied Psychology, 91,* 828 – 840.

Caldwell, D. F., & O’Reilly, C. A. (1985). The impact of information on job choice and turnover. *Academy of Management Journal, 28,* 934-943.

Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review, 102*, 2981–3003.

Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A., & Jones, D. A. (2005). Applicant attraction to organizations and job choice: A meta-analytic review of the correlates of recruiting outcomes. *Journal of Applied Psychology, 90,* 928–944.

Chatterjee, P. (2001). Online reviews: Do consumers use them? in M. C. Gilly & J. Meyers-Levy (eds.), *NA - Advances in Consumer Research* (Vol. 28), Valdosta, GA: Association for Consumer Research (pp. 129-133).

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*, 345-354.

Childress, L. (2018). 6 HR & recruiting stats you need to know for 2018 and Beyond. Retrieved at: https://www.glassdoor.com/employers/blog/6-hr-recruiting-stats-you-need-to-know-for-2018-and-beyond/

Chintagunta, P. K., 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 Science*, *29*, 944–957.

Dana, J., & Cain, D. M. (2015). Advice versus choice. *Current Opinion in Psychology, 6*, 173-176.

Dineen, B. R., Ash, S. R., & Noe, R. A. (2002). A Web of applicant attraction: Person-organization fit in the context of Web-based recruitment. *Journal of Applied Psychology, 87,* 723–734.

Dineen, B. R., Ling, J., Ash, S. R., & DelVecchio, D. (2007). Aesthetic properties and message customization: Navigating the dark side of web recruitment. *Journal of Applied Psychology, 92*, 356–372.

Dineen, B. R., & Noe, R. A. (2009). Effects of customization on application decisions and applicant pool characteristics in a web-based recruitment context. *Journal of Applied Psychology, 94*, 224–234.

Dunn, E. W., Aknin, L. B., & Norton, M. I. (2008). Spending money on others promotes happiness. *Science, 319*, 1687–1688.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*, 817-868.

Floyd, K., Freling, R., Alhoqail, S., Young Cho, H., & Freling, T. (2014). How online product reviews affect retail sales: A meta-analysis. *Journal of Retailing, 90*, 217–232.

DeMers, J. (2014). How negative online review can impact your business and recruiting. *Forbes,* retrieved at: https://www.forbes.com/sites/jaysondemers/2014/09/09/how-negative-online-company-reviews-can-impact-your-business-and-recruiting/#1409f0c91d9b

Fradkin, A., Grewal, E., Holtz, D., & Pearson, M. (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. *Association for Computer Mahcinery.*

Gatewood, R. D., Gowan, M. A., & Lautenschlager, G. J. (1993). Corporate image, recruitment image, and initial job choice decisions. *Academy of Management Journal, 36,* 414-427.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24,* 153-172.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives, 25*, 191-209.

Heath, C. (1996). Do people prefer to pass along good news or bad news? Valence and relevance of news as predictors of transmission propensity. *Organizational Behavior and Human Decision Process, 68,* 79-94.

Heath, C., Bell, C., & Sternberg, E. (2001). Emotional selection in memes: the case of urban legends. *Journal of Personality and Social Psychology, 81*, 1028-1041.

Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM, 52,* 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, T. K., Ribeiro, B., Madhyastha, H. V., & 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.

Kappes, H. B., Balceits, E., De Cremer, D. (2018). Motivated reasoning during recruitment. *Journal of Applied Psychology, 103,* 270-280.

Khern-am-nuai, W., Kannan, K., & Ghasemkhani, H. (2016). Extrinsic versus intrinsic rewards to participate in a crowd context: An analysis of a review platform. *Working paper, Purdue University, West Lafayette, IN.*

Klein, N., Grossman, I., Uskul, A. K., Kraus, A. A., & 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. *The Journal of Positive Psychology, 12,* 354-361.

Lu, S. F., & Rui, H. (2017). Can we trust online physician ratings? Evidence from cardiac surgeons in Florida. *Management Science* (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 Review, 104*, 2421-2455.

McFarland, L. A., & Ployhart, R. E. (2015). Social media: A contextual framework to guide research and practice. *Journal of Applied Psychology, 100,* 1653–1677.

Moe, W.W., Trusov, M. (2011) The value of social dynamics in online product ratings forums. *Journal of Marketing Research, 49*, 444–456.

Paoloacci, G., Chandler, J., & Ipeirotis, G. (2010). Running experiments on Amazon Mechnical Turk. *Judgment and Decision Making, 5,* 411-419.

Peters, K., Kashima, Y., & Clark, A. (2009). Talking about others: Emotionality and the dissemination of social information. *Eropoan Journal of Social Psychology, 39*, 207-222.

Ployhart, R. E., Schmitt, N., & Tippins, N. T. (2017). Solving the supreme problem: 100 years of selection and recruitment at the Journal of Applied Psychology. *Journal of Applied Psychology, 102,* 291-304.

Rimé, B. (2009). Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review, 1*, 60-85.

Ryan, A. M., Sacco, J. M., McFarland, L. A., & Kriska, S. D. (2000). Application self-selection: Correlates of withdrawal from a multiple hurdle process. *Journal of Applied Psychology, 85,* 163-179.

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of 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. *Journal of Consumer Research, 42*, 840-845.

Stephen, A., Bart, Y., Du Plessis ,C., & Goncalves, D. (2012). Does paying for online product reviews pay off? The effects of monetary incentives on content creators and consumers. In Gürhan-Canli Z, Otnes C, Zhu RJ (eds.). *NA–Advances in Consumer Research* (Vol. 40), Duluth, MN: Association for Consumer Research (pp. 228–231).

Swider, B. W., Zimmerman, R. D., & Barrick, M. R. (2015). Searching for the right fit: Development of applicant person-organization fit perceptions during the recruitment process. *Journal of Applied Psychology, 100,* 880-893.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Yale University Press.

Uggerslev, K. L., Fassina, N. E., & Kraichy, D. (2012). Recruiting through the stages: A meta-analytic test of predictors of applicant attraction at different stages of the recruiting process. *Personnel Psychology, 65,* 597-660.

Wang, J., Ghose, A., & Ipeirotis, P. (2012) Bonus, disclosure, and choice: What motivates the creation of high-quality paid reviews? In *International Conference on Information Systems (ICIS 2012)*. Orlando, FL: Association for Information Systems.

	Mturk			Glassdoor		
Variable	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

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

Note: Ordinary least square regressions. The variable “voluntary” is an indicator equal to 1 if the review was provided voluntarily and to 0 if the review was provided through Give-to-Get. In column 1 and 2, “Rating” is the 1 to 5 employer rating given by each individual. In column 3 and 4, “Is 1 star” is an indicator variable taking the value of 1 if the rating is 1 star, and 0 otherwise. In columns 5 and 6, “Is 5 stars” is an indicator variable taking the value of 1 if the rating is 5 stars, and 0 otherwise.

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**	-0.660***	0.050	-0.017	-0.295	-0.290
	(0.225)	(0.228)	(0.190)	(0.192)	(0.200)	(0.216)
Age		-0.004		0.006		-0.009
		(0.010)		(0.008)		(0.010)
Female		-0.034		0.284		0.223
		(0.225)		(0.176)		(0.207)
More than 1000 employees		-0.473**		-0.464**		-0.319
		(0.223)		(0.196)		(0.196)
Education (years)		-0.026		-0.017		-0.053
		(0.048)		(0.041)		(0.061)
Tenure (years)		-0.007		-0.053*		0.005
		(0.022)		(0.029)		(0.024)
Constant	4.019***	4.846***	4.019***	4.315***	4.019***	5.171***
	(0.117)	(0.709)	(0.117)	(0.683)	(0.117)	(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

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

Note: Ordinary least square regressions. In all columns, the dependent variable is the 1 to 5 employer rating given by each individual. The variable "Choice" is an indicator equal to 1 if the review was provided in the choice condition and to 0 if the review was provided in the forced condition. In column 1 and 2, only data from the "No incentive" condition is used. In column 3 and 4, only data from the "Nonspecific prosocial" condition is used. In columns 5 and 6, only data from the "Monetary high" condition is used.

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	-0.048	-0.047	-0.042	-0.289	-0.236
	(0.252)	(0.267)	(0.215)	(0.195)	(0.225)	(0.232)
Age		-0.006		0.002		0.004
		(0.012)		(0.008)		(0.009)
Female		0.347		0.288		0.034
		(0.255)		(0.201)		(0.228)
More than 1000 employees		0.140		-0.616***		-0.366*
		(0.262)		(0.201)		(0.210)
Education (years)		0.081		0.031		0.011
		(0.063)		(0.046)		(0.049)
Tenure (years)		0.037		-0.013		-0.005
		(0.035)		(0.020)		(0.022)
Constant	3.686***	2.391**	4.019***	3.611***	4.019***	3.851***
	(0.139)	(1.163)	(0.117)	(0.830)	(0.117)	(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

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

Note: Ordinary least square regressions. In all columns, the dependent variable is the 1 to 5 employer rating given by each individual. The variable "Choice" is an indicator equal to 1 if the review was provided in the choice condition and to 0 if the review was provided in the forced condition. In column 1 and 2, only data from the "Positive Prosocial" condition is used. In column 3 and 4, only data from the "Negative prosocial" condition is used. In columns 5 and 6, only data from the "Monetary low" condition is used.

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)

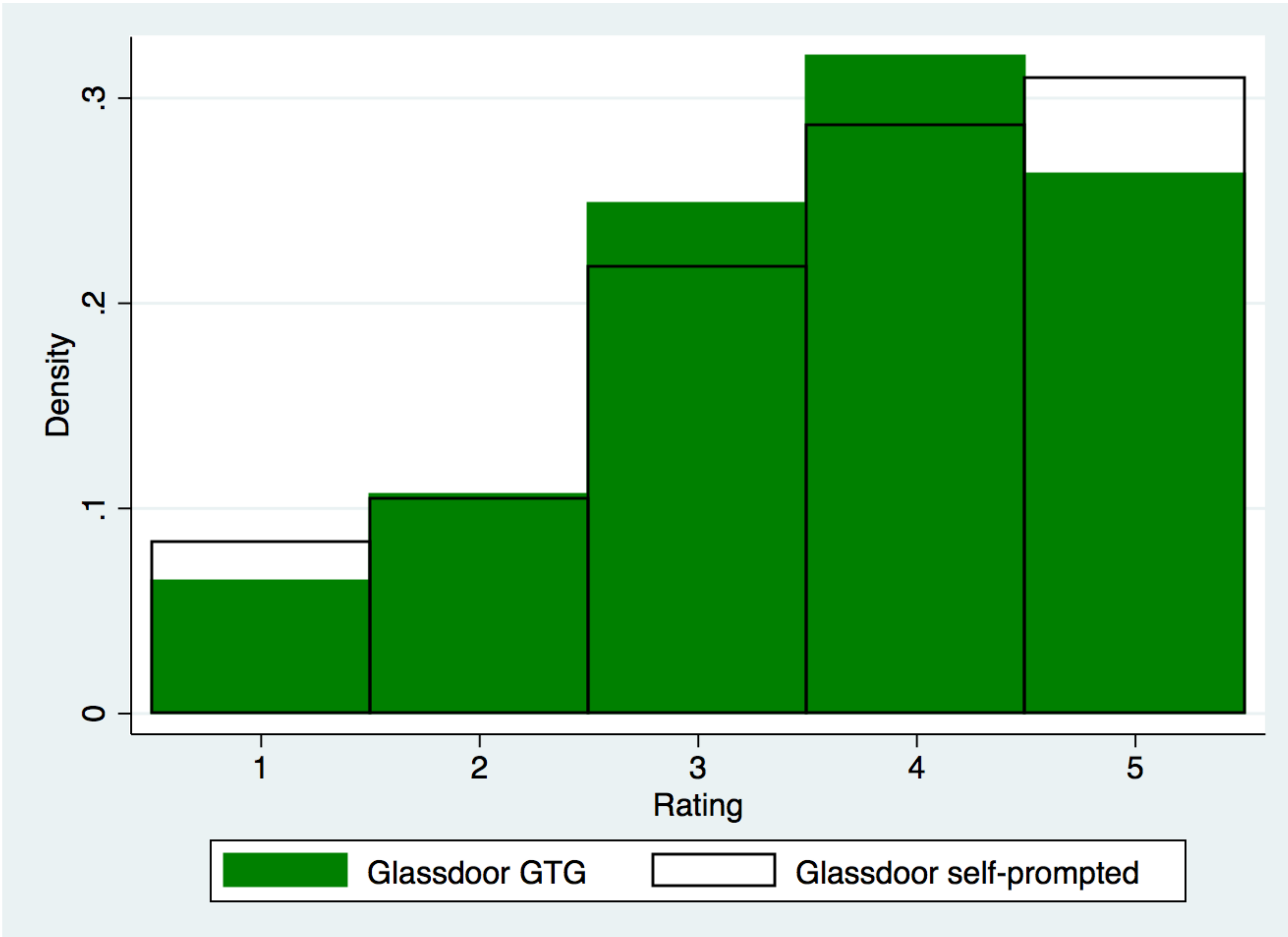
	Rating	Rating
	(1)	(2)
Monetary high	-0.001	-0.002
	(0.180)	(0.180)
Monetary low	0.022	0.050
	(0.187)	(0.191)
Negative prosocial	-0.325	-0.319
	(0.198)	(0.199)
Nonspecific prosocial	-0.332*	-0.345*
	(0.190)	(0.194)
Positive prosocial	-0.332*	-0.331*
	(0.181)	(0.182)
Choice ×		
Control (no incentive)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

Robust standard errors in parentheses

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

Note: Ordinary least square regressions. The sample includes data from all conditions. In all columns, the dependent variable is the 1 to 5 employer rating given by each individual. The variable "Choice" is an indicator equal to 1 if the review was provided in the choice condition and to 0 if the review was provided in the forced condition; the second set of variables in the table (beginning after "Choice $\times$") are interactions between each incentive condition and the Choice indicator. Column 2 adds the control variables shown in Table 4.

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



Test for significant difference in distributions: Pearson χ^2 (4) = 427.2152 $p < 0.001$

Figure 1: Glassdoor GTG vs. voluntary reviews

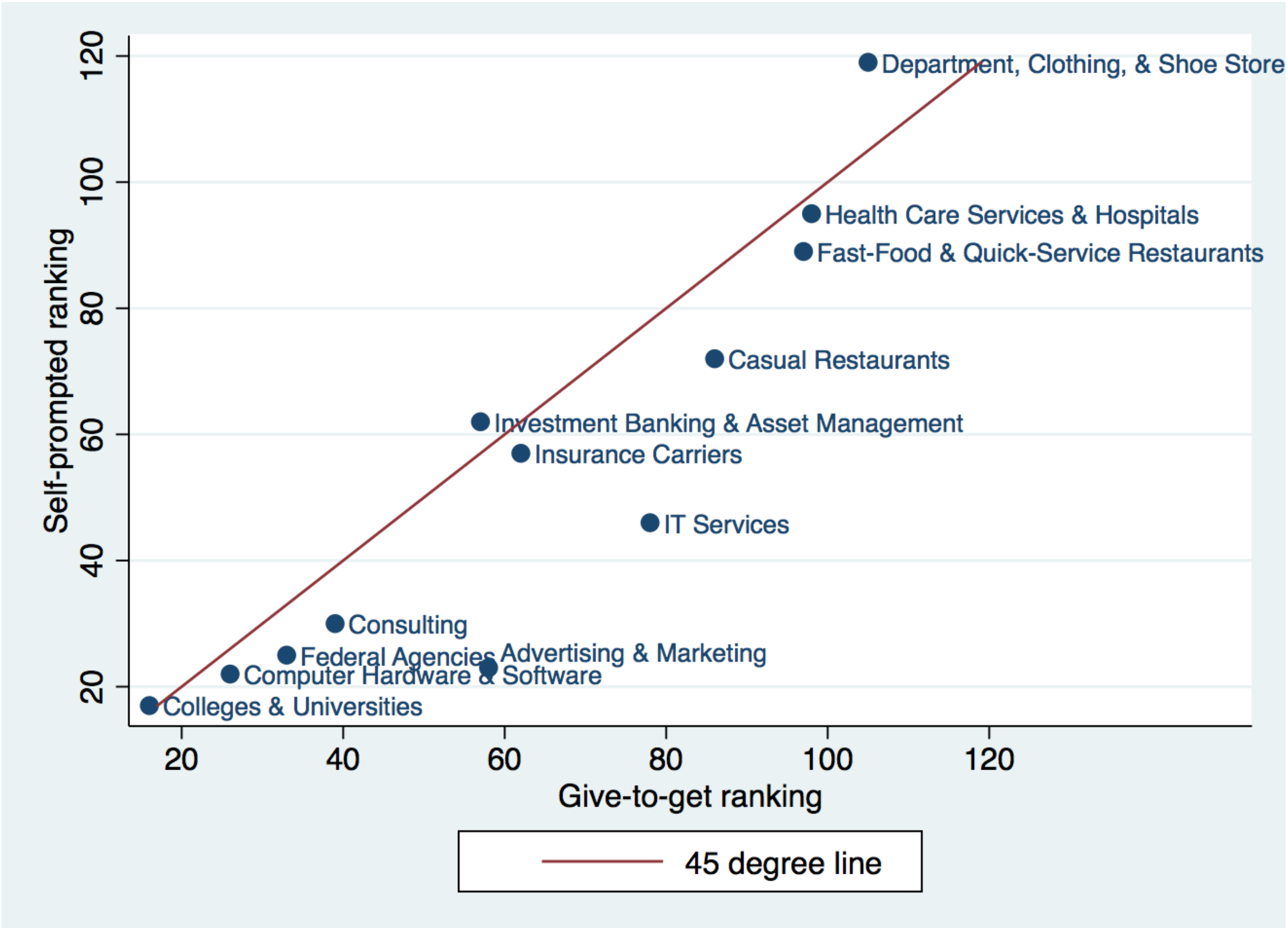


Figure 2: Glassdoor: changes in rankings of frequent industries due to bias

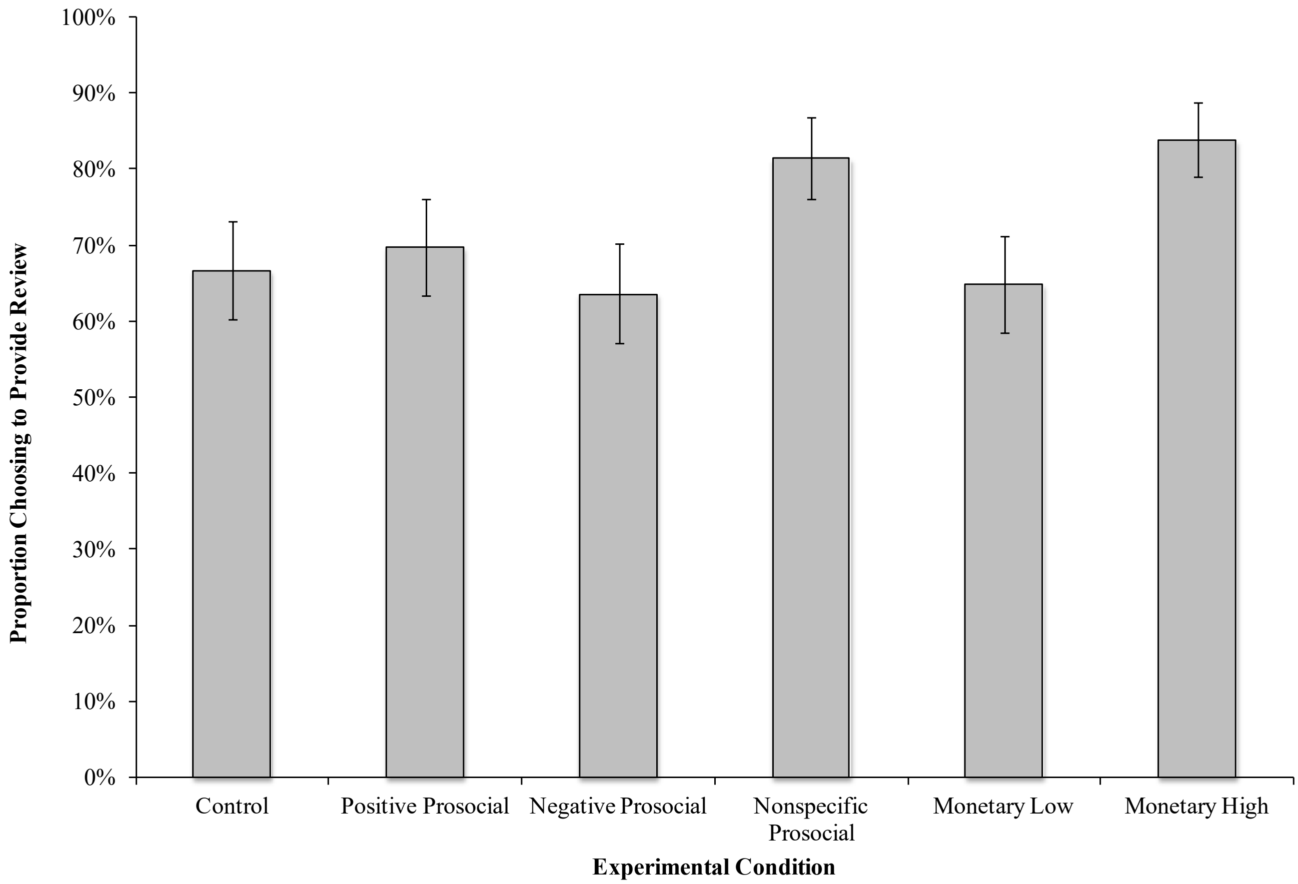
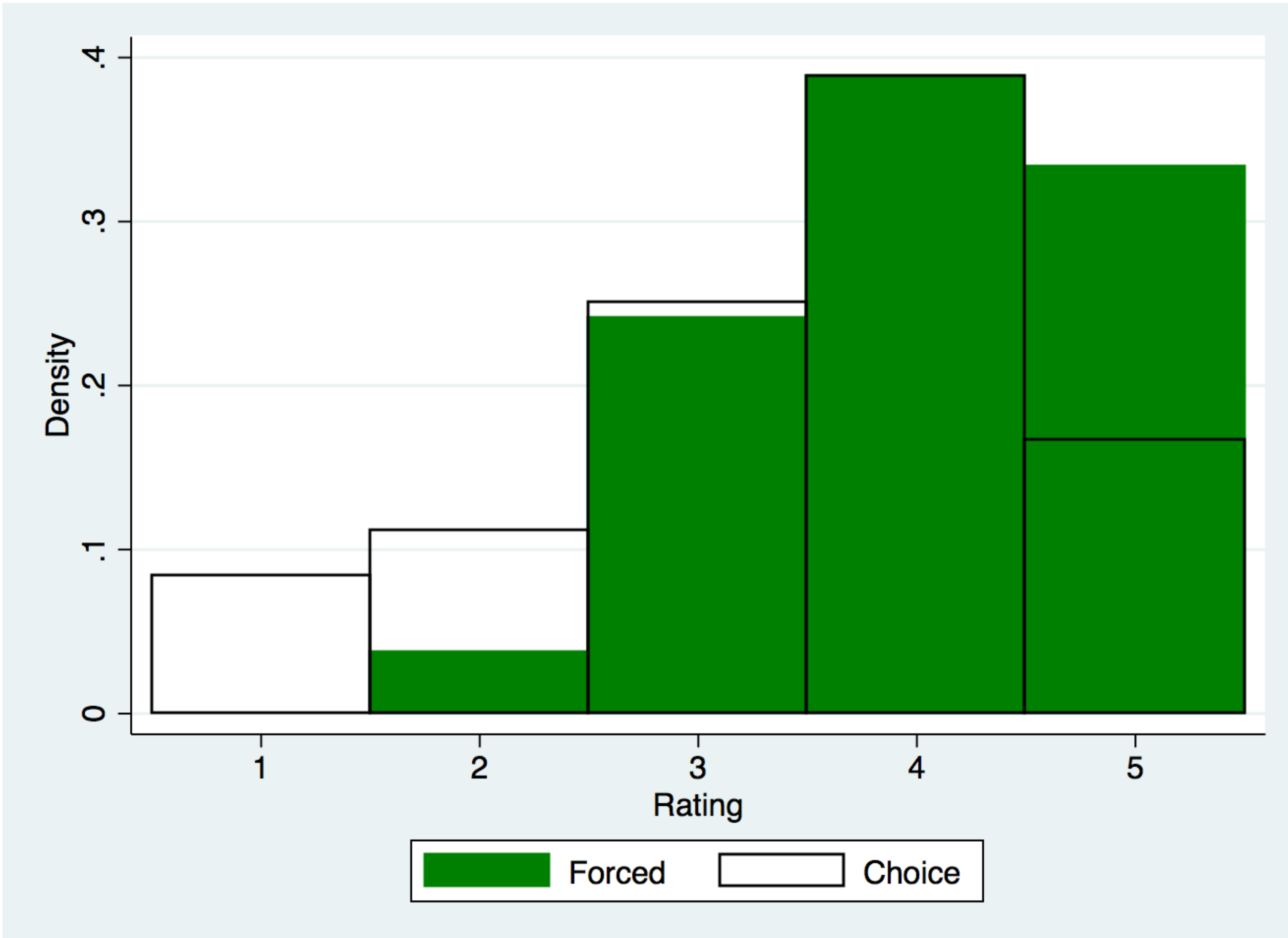
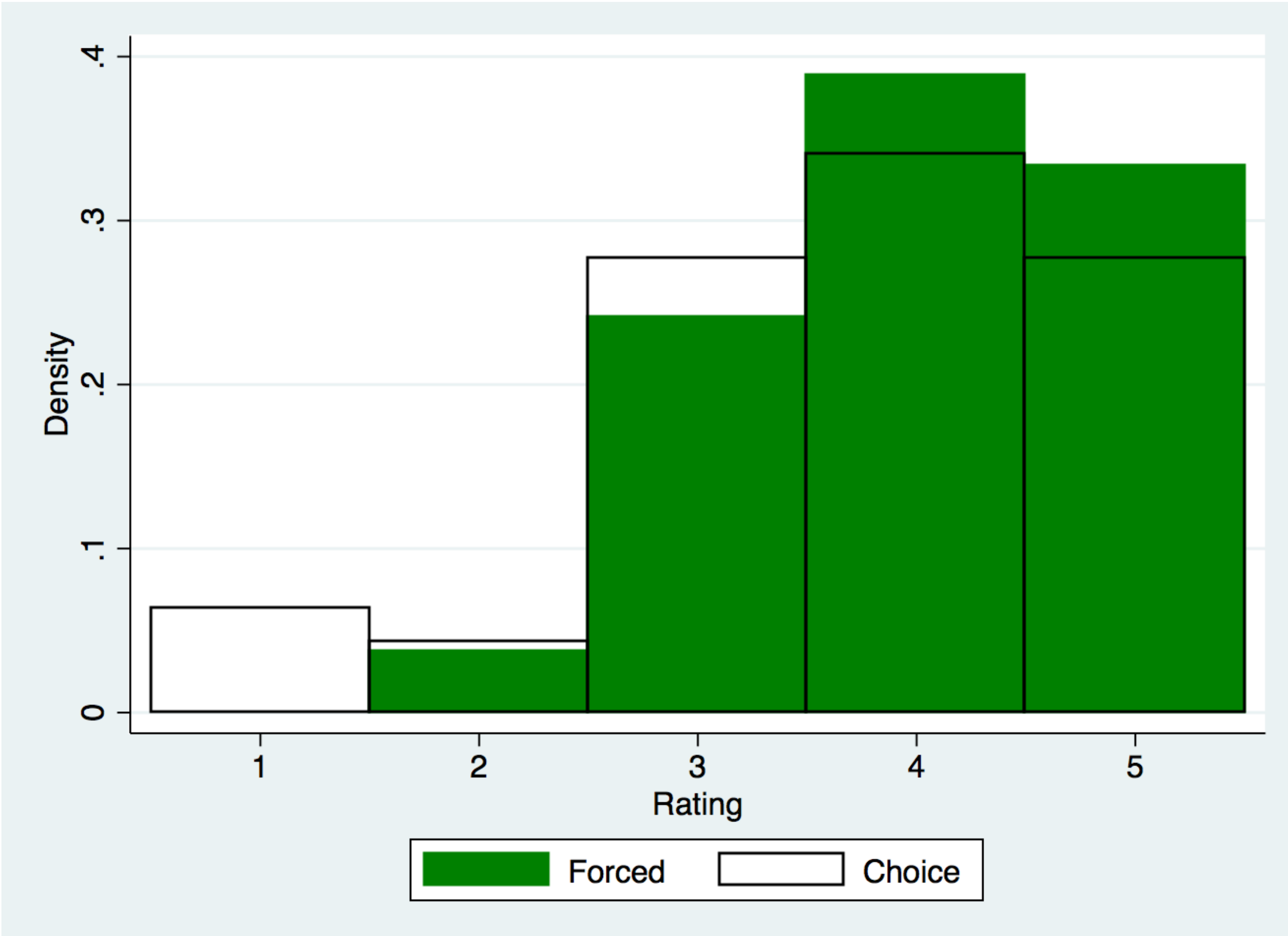


Figure 3: MTurk experiment: efficacy of different incentives in increasing response rates



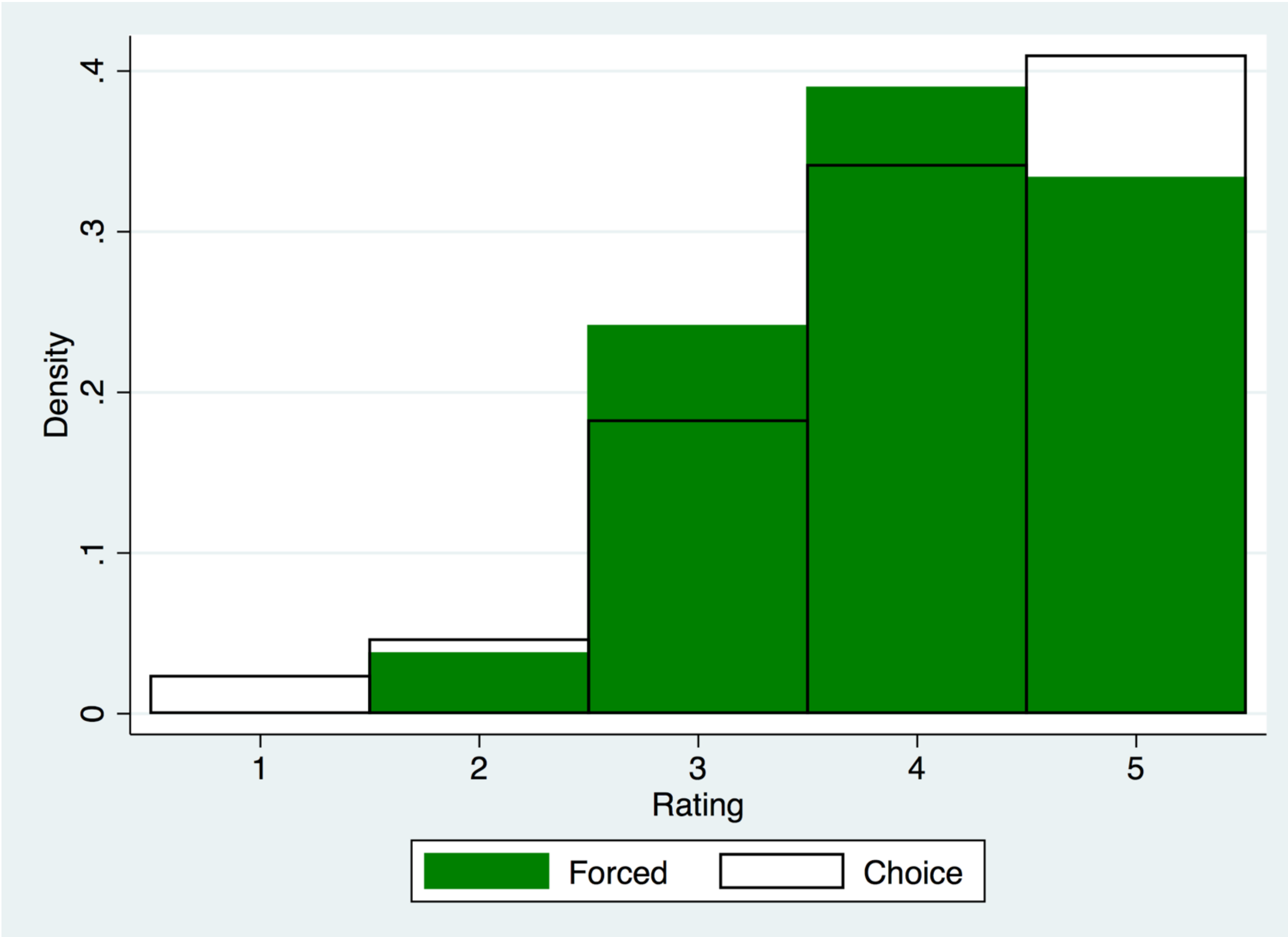
Test for significant difference in distributions: Pearson χ^2 (4) = 8.5354 p = 0.074

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



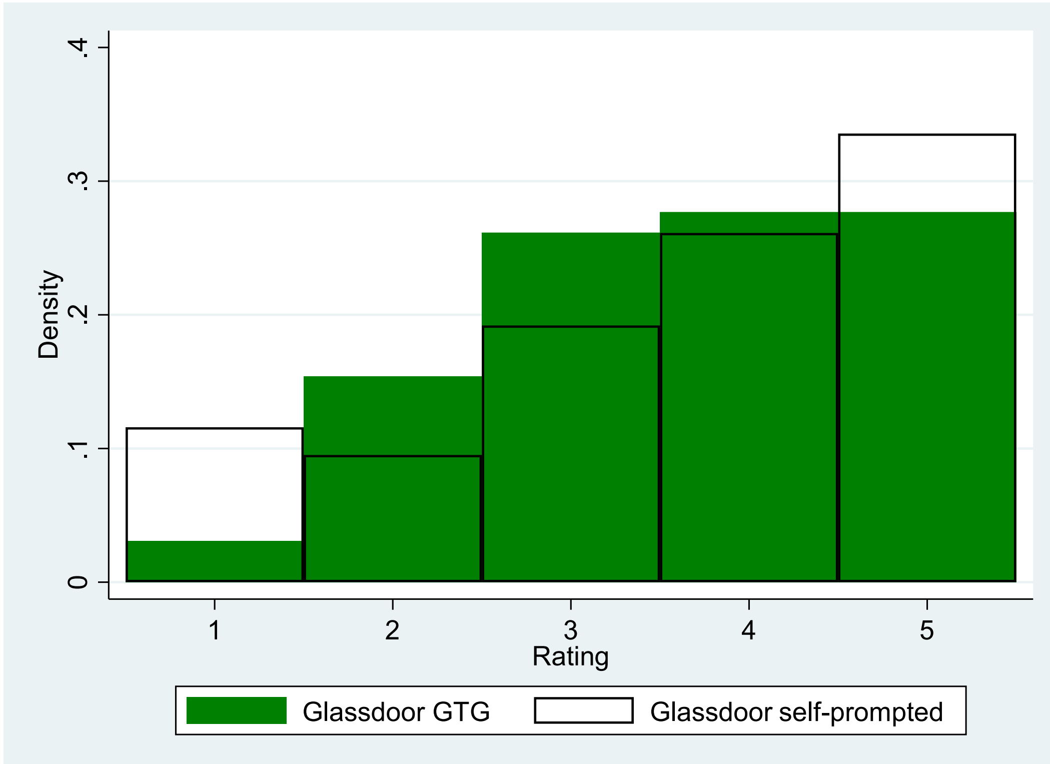
Test for significant difference in distributions: Pearson χ^2 (4) = 4.0163 p = 0.404

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



Test for significant difference in distributions: Pearson χ^2 (4) = 2.1929 p = 0.700

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



Test for significant difference in distributions: Pearson $\chi^2(4) = 8.022$ $p = 0.091$

Figure 7: Glassdoor GTG vs. voluntary reviews based on nearest-neighbor matching to MTurk sample (sub-sample of 643 observations most similar to observations in MTurk)