The Tradeoffs of Transparency:
Measuring Discrimination When Subjects Are Told They Are in an Experiment

Amanda Agan
Rutgers University

Bo Cowgill
Columbia University

Laura Gee
Tufts University*

January 10, 2024

Abstract

Correspondence audit studies have sent almost one-hundred-thousand resumes without informing subjects they are in a study — increasing realism, but without being fully transparent. We study the potential trade-offs of this lack of transparency by running a hiring field experiment with recruiters in a natural setting. One group of recruiters is told they are screening for an employer, and another is told they are part of an academic study. Job applicants' gender is randomly assigned. When subjects are told they are in an experiment, callback rates and willingness-to-pay for male candidates decline relative to female candidates (with no detectable change for female candidates). This suggests that telling subjects they are in an experiment would underestimate gender inequality.

JEL Classification: J7, C9, D9
Keywords: Field Experiments, Methodology, Gender, Discrimination

*The authors thank David Autor, Zoe Cullen, Amy Finkelstein, Yana Gallen, Judd Kessler, Corinne Low, Prachi Jain, Anya Samek, Olga Stoddard and participants at the AEA Annual Meeting, the Advances with Field Experiments Conference, the Economic Science Association World Meetings, and seminars at Nuffield College (Oxford), Stanford and MIT for valuable feedback. The experiments in this paper were approved by an Institutional Review Board and registered at the AEA RCT Registry under AEARCTR-0003088.

"...a plausible concern is that when subjects know they are participating in an experiment... they may react to that knowledge in a way that leads to bias in the results." – List (2011)

1 Introduction

Disparities on the basis of race, gender, and other characteristics persist in labor markets and other domains (e.g. Altonji and Blank 1999; Blau and Kahn 2017; Chetty et al. 2020). Understanding the causes of these disparities is imperative for developing strategies to remedy and alleviate these inequalities. A common methodology for estimating the scope of discrimination is correspondence audit studies, which have been used to study discrimination in multiple domains (Bertrand and Duflo, 2017; Neumark, 2018). They are most commonly used to study labor market discrimination, with over 300 such experiments run between 2005 and 2020 which sent out almost 965,000 applications (Baert, 2018; Lippens et al., 2023).

The realism of these natural field experiments offers the ability to detect discriminatory behavior, not just attitudes, in a real-world setting. However, audit studies also have drawbacks including a lack of informed consent, and uncompensated time costs of subjects (Jamison et al., 2008; Ortmann and Hertwig, 2002; Charness et al., 2022; Salganik, 2019).[1] These concerns have led some researchers — as well as some unwilling subjects — to advocate against deceptive audit studies (McDaniel and Starmer, 1998; Ortmann and Hertwig, 2002).[2]

However, norms and policies in favor of more transparency could produce tradeoffs on other margins. Studies with more transparency may produce biased estimates of disparities or discrimination, leading researchers to misinterpret the extent and reasons for inequality. Natural field experiments where subjects are not told they are in an experiment may be particularly useful for studying sensitive or potentially illegal behavior such as discrimination on the basis of race, gender, disability, or other protected class. In this paper we ask: how does subject behavior — particularly with regards to gender discrimination in hiring — differ when subjects are told they are in an academic study, compared to a (perceived) real-world setting?

[1] There are also critiques of audit studies' ability to uncover discrimination, notably in Heckman and Siegelman (1993) and Heckman (1998), however given their proliferation we abstract away from this concern here.

[2] Government officials in the United Kingdom recently criticized academics for sending them fictitious emails, apparently randomized for research. https://www.theguardian.com/politics/2021/feb/24/mps-criticise-academics-for-sending-them-fictitious-emails-for-research. Complaints have been filed by professors and restaurants who later realized they were in audit studies without their informed consent, and some professors have been the subjects of lawsuits (King et al., 2013). These concerns have translated into some Institutional Review Boards becoming more cautious of the traditional audit study methodology with Connor (2000) offering a case study of an IRB who declined to approve an employment audit study.

To study the potential tradeoffs of transparency, we compare the average gender gaps in hiring outcomes between two groups of recruiters who both participated in an adaptation of the traditional audit study. Both groups were hired through a third party platform and were asked to evaluate job candidates for payment (using incentives similar to natural settings for recruiters). One group was *not* told they were part of an academic study (as is typical for an audit study) and were led to believe they were evaluating real applicants for a real employer. The other group was explicitly told they were participating in an academic study.

We call our adaptation of the standard audit study methodology a two-sided audit study. This methodology allows the researcher to randomize attributes of two sides of the labor market: job candidates and the firms that hire them. In a two-sided audit study researchers hire professional freelance recruiters in a natural setting to aid a (unknown-to-them fictitious) firm in the hiring of applicants. This allows researchers to study variations in the firm's hiring method (e.g. asking for salary history) as well as variations in candidates' attributes (e.g. the candidate's gender), and the interaction between these variations. This paper is not focused on firm-level variations, but the two-sided audit study is an ideal setting for testing whether the subjects behave differently if they are participating under natural conditions (unaware of the experiment) versus if they are told they are taking part in academic research with fictitious candidates. This is because we can hold all other aspects of the audit study the same, while only varying whether or not the subjects are told they are in an experiment.

Our standard two-sided audit study results come from Agan, Cowgill and Gee (2022) (henceforth ACG (2022)) where we hired freelance recruiters from a third-party platform to evaluate software engineer applications for a firm. We paid these recruiters their market wage and a bonus. Across applicants, we randomly varied: gender, salary at current employer, and whether current salary is disclosed.[3] For each of eight applications sent to a subject, recruiters were asked for feedback about the candidates in a typical way. This feedback included whether the firm should interview the candidate, the take-it-or-leave-it salary amount the firm should offer, and the willingness-to-pay (WTP) the firm should have for each candidate.[4]

We also ran an arm of this experiment pulling from the same pool of recruiters with 448 applications reviewed by 56 recruiters with similar market rate compensation — but these recruiters were explicitly told the task was hypothetical and being run by an academic re-

[3] Across recruiters, the job application form completed by candidates differed. Some forms asked for salary history, and others did not.

[4] For more on the realism of our structured response form and the recruitment outsourcing industry more generally, see Agan, Cowgill and Gee (2022).

search team. In this arm of the experiment the recruiters signed a standard IRB informed consent form and were told the following before beginning their evaluations: "Our research endeavors to help companies and recruiters optimize their work, so please respond as if this was a real recruiting/screening task... In addition, we will pay you a bonus based on data from how real companies would evaluate these candidates."

We find that when recruiters are aware they are part of an experiment, this changes their behavior and our estimate of the level of gender disparities in this market. In particular, when subjects are told they are in an experiment, this generally harms outcomes for male applicants without detectable changes for women. In particular, recruiters reduce the average male callback rate by 12 percentage points relative to the change for women (a 18% decline over the male callback rate when recruiters are not told) and they reduce their average willingness to pay by $4,050 for men relative to women (a 4% decline over WTP for men when recruiters are not told). Thus being told you are in an experiment improves female outcomes relative to male. Women are called back more than men, and gender disparities in female-male ratio of willingness to pay for applicants go from 0.94 to 0.98. A researcher using data only from subjects told they were in an experiment would misinterpret the level of gender inequality in this market.

There are multiple reasons why recruiters' decisions could differ in these two contexts. When recruiters are told they are in a study they may: exert less effort knowing that the stakes are not real, react to the knowledge their responses are being recorded, respond to potential experimenter demand effects, change actions based on social image concerns (avoid looking biased), or incorporate social justice goals (avoid perpetuating bias; Levitt and List, 2007). The idea that subjects may respond differently in natural environments than in a known experiment has been called the "Hawthorne effect" (Levitt and List 2011).[5] In addition, subject characteristics (observable and unobservable) may differ between the two settings: individuals who are prone to discrimination may deliberately avoid selecting into experiments, or there may be other differences in the sample of individuals willing to participate in an academic study (Levitt and List 2007, Crabtree and Dhima 2022).

The traditional audit study context likely gives a more realistic estimate of the discrimination that actual job applicants experience than alternative methodologies. Audit studies achieve their realism by misleading subjects (Jamison et al., 2008; Ortmann and Hertwig, 2002; Charness et al., 2022). Obfuscation is controversial in experimental economics partially because it does not allow for voluntary participation, uses subject effort on fictitious candidates, usually does not allow subjects to be compensated for these efforts, and it may

[5]Though Levitt and List 2011 study of the original Hawthorne plant experiments uncovers little evidence for the so-called "Hawthorne effect", the moniker remains common.

cause further harm by changing subjects' beliefs moving forward. Given these concerns, we designed the less transparent arm of our experiment to avoid some of these downsides. In our two-sided audit approach, all subjects are compensated at their market rate. We inhibit recruiters from learning or generalizing about candidate's (fictitious) outcomes by withholding candidate-specific feedback about the outcomes from the applications reviewed by the recruiters. We elaborate on these and other ways our design minimized negative externalities in Appendix B.

However, our two-sided audit study, like traditional audit studies, still involves two types of obfuscation: first, subjects are not informed they are participating in an experiment, and second, unbeknownst to them, the candidates they are evaluating are not real (and in our case the employer is not real either). In a different context, Gallen and Wasserman (2022) reduced the amount of misinformation in an e-mail correspondence audit study by recruiting real student subjects who were interested in career advice from professionals to test how these professionals responded to emails from male versus female students. While the professionals did not consent to be in a study, the requests they were receiving were real and could help actual students. It might be feasible to translate this to a labor market setting by finding and recruiting job-seekers with similar resumes but different genders, though it may be difficult to scale to the number needed for sufficient power.

While the audit methodology has costs (obfuscation and use of subjects' time), it may also have benefits in uncovering discrimination compared to other potential methods (Boggs et al., 1993; Edley Jr, 1993; Riach and Rich, 2004; Zschirnt, 2019; Crabtree and Dhima, 2022). One reason to value realism is the longstanding fear that people will act differently when they know they are being observed (Levitt and List, 2007; McCambridge and Day, 2008; List, 2011; McCambridge et al., 2014; De Quidt et al., 2018). Perhaps subjects hope to appear socially desirable to an audience (Jann et al., 2019), or because they want to please the experimenter with the "right" actions (Orne, 1962; Zizzo, 2010). These concerns may be particularly strong when studying sensitive topics like gender or racial discrimination. Beyond hoping to appear a certain way, subjects may exert lower effort if the task is known to be lower stakes or totally hypothetical (Cameron, 1999).

In other settings researchers have found differences in behavior between laboratory experiments and more natural settings (Laury and Taylor, 2008; Galizzi and Navarro-Martinez, 2019; List, 2006). In the labor market setting, two papers have looked at hiring outcomes in both a traditional audit study and lab studies, but comparisons across the field and lab settings were not the point of either paper (the laboratory setting was used to test mechanisms). Neither elicited the same outcome across study conditions (Derous et al., 2012; Kristal et al.,

2023). Furthermore, in line with our own findings, studies that compare traditional audit studies with survey responses find a reduction in disparities in the survey data when subjects know they are part of a research study versus the traditional audit methodology (Pager and Quillian, 2005; Wulff and Villadsen, 2020). To our knowledge there has been no study in the labor market replicating an audit study procedure by manipulating whether subjects are told they are in a study.

Some researchers have developed methodologies to try to uncover evidence of discrimination with more transparency. Landgrave (2020) measures willingness of political actors to take part in an academic research survey where compensation is randomized between a donation to a scholarship fund that benefits specifically Hispanic students or nondescript students; though this inspired design is hard to translate to a hiring setting. Additionally, innovative partnerships between researchers and practitioners may offer alternatives to lack of transparency (Kessler et al., 2019; Low, 2021; Chan, 2022; Macchi, forthcoming). In a labor market setting, Kessler et al. (2019) asks real firm representatives to rate fictitious job candidates and uses the ratings of fictional candidates to develop recommendations for real job candidates. Subjects have an incentive to reveal their preferences truthfully in order to obtain better recommendations. Kessler et al. (2019) still informs employers that their anonymized data will be analyzed for research, which could shape their responses to sensitive characteristics. Their innovative methodology also requires practitioner cooperation (which can be challenging to identify and implement; List, 2011), and firms willing to participate may be systematically different. Baker and Grimm (2021) find that explicitly being told one is part of research on Mturk affects the level of racial discrimination in that setting. As far as we know, we are the first to test the effect of being told you are in a study in a natural field experimental labor market hiring setting, which more closely reproduces the structure of traditional audit studies. And our findings have important implications for and add context to the interpretation of studies that focus on the behavior of recruiters and other actors with increased transparency.

2 Methodology

To compare the responses of recruiters who are told they are in a study to recruiters who are not, we hired real professional recruiters via a third party website to each evaluate eight job candidates for a firm.[6]

[6]To find these real professional recruiters in both arms of our experiment we made a list of independent freelance workers on the platform with profiles that mentioned keywords like "recruiter", "staffing", and "human resources." A random-subsample of the potential experimental subjects were reviewed by our research

Our design takes advantage of the practice of delegating recruitment decisions to specialists. This has become increasingly common in the past two decades.[7] Nearly all firms that use outsourced recruitment ask recruiters to provide input on salary and 85% of this work is done remotely (even before COVID-19).[8] We identified recruiters who are representative of those hired by companies through the recruiting industry, and hired them through a freelance platform for outsourcing administrative and other white-collar work. We offered to pay them both their hourly wage and a bonus, which is a common combination of incentives for this type of work (Adler, 2020).[9]

These subjects were randomized into either the experimental arm where they were explicitly told they were in a study, or into the original Agan, Cowgill and Gee (2022) study where they are not told they are in an experiment (and to our knowledge all the subjects in ACG (2022) believed they were interacting with a real firm and evaluating real job candidate). Subjects were invited directly, then accepted our invite, and were enrolled in the experiment until we reached our predetermined goals for the number of subjects. In both the original study and the "told in experiment" version, 21% accepted our invite and 87% completed the experiment conditional on being enrolled in our study. In this paper we focus on 76 recruiters from ACG (2022) who viewed the same sets of eight job candidates as the 56 recruiters who were explicitly told they were in an experiment.[10]

team to determine if they had prior real-world experience in hiring or recruiting and charged less than $100 per hour for their services. Because the pool of potential subjects for both arms was created collectively, those eventually invited in both conditions are similar on observable variables such as: hourly rate, total hours billed, having previous feedback and average feedback measured before our invites were sent (results available from authors upon request). Hourly rate is the hourly rate in dollars subjects provide to the platform, total hours billed is the number of hours they have worked so far on the platform, having feedback is a binary variable for whether the subject has feedback on the platform, and average feedback conditional on feedback denotes their average feedback score (1 to 5) conditional on having any feedback on the platform.

[7] https://www.alliedmarketresearch.com/recruitment-process-outsourcing-market-A31397. See also, e.g., Landay and DeArmond (2018); Black et al. (2020); Cappelli (2019); Cowgill and Perkowski (forthcoming).

[8] See *Staffing Industry Analysts*, RPO Market Developments, December 2017, https://staging.kornferry.com/media/sidebar_downloads/Measuring-Up-A-new-research-report-about-RPO-metrics.pdf, and https://www.shrm.org/ResourcesAndTools/business-solutions/Documents/Talent-Acquisition-Report-All-Industries-All-FTEs.pdf.

[9] Because there were no actual candidates nor firms, bonuses were paid based on simulated outcomes using data from comparable settings. The "told experiment" subjects were told this, and the "not told" subjects were told bonuses would be paid based on outcomes at the client firm.

[10] In our original study we created 32 specific job candidate biographies that were evaluated by a total of 256 recruiters. In our analysis we include job candidate biography fixed effects so that inferences are made about differences in outcomes for all male candidates versus all female candidates while controlling for the individual attributes of each specific fictitious candidate. To ensure we had more evaluations of each specific fictitious candidate while limiting the number of recruiters we hired we limited ourselves to only 16 candidate biographies in the experimental arm where subjects are explicitly told they are in an experiment. The inclusion of job candidate biography fixed effects precludes the use of the full 256 recruiters from ACG (2022). For a comparison of the full 256 subjects in ACG (2022) to the 76 reported on in our main text see Table A1. Summary

While our original two-sided audit study design avoided some of the downsides of lack of transparency — subjects were paid, unlike most resume/correspondence audit studies, and subjects are not given individual-level feedback about each job candidate they review to avoid belief-updating[11] — subjects were still participating without full transparency. By contrast, in the "told in experiment" condition, subjects are told they are in an experiment, and the subjects were recruited with a message that differed from ACG (2022). In this condition, we explicitly mentioned the names of the authors and their academic affiliations (as suggested by IRB protocol). Recruiters were explicitly told that the company is hypothetical as detailed in Table B1. In addition, subjects had to sign and return an informed consent form — as typically suggested by IRB protocol, but different from the ACG (2022) condition.[12] Subjects in both conditions were sent the same instructions including information about the position they were hiring for. They also received eight one-page job candidate applications and a structured evaluation form to report assessments and decisions about the eight candidates. The last distinction between conditions appeared on the evaluation form. In the "told in experiment" condition, recruiters saw university branding in the background of the form (versus no branding in ACG (2022); details of the recruitment wording and style can be found in the Appendix).

All subjects were paid their advertised hourly wage, and a bonus (see ACG (2022) for details). Our study is a between-subjects design, so no single recruiter experiences both versions of our experiment. An additional benefit of our two-sided audit design is that we collect more than the binary callback outcome appearing in traditional audit studies. Gaps in callback rates are an important component of discrimination, but wage gaps are also a major concern. Although many natural field experimental audit studies have found evidence of lower callback rates for racial minorities, the evidence is more mixed for discrimination in the form of lower callback rates for women (Lippens et al., 2023). Yet there is a persistent gender wage gap in the real world which the two-sided audit methodology is more able to explore.

We drew subject recruiters for the "told in experiment" condition from the same pool as the recruiters in our two-sided audit study. As such, generalizability concerns that plague lab experiments (e.g., often performed by college students) are less of an issue. We also kept the stakes the same. Besides the possibility that in the audit study the recruiters may have thought they could be hired again in the future — different opportunity costs should not have played a role in creating different outcomes. Despite these, there still could have

statistics and results remain similar when using the full sample.

[11] We elaborate on these issues in Appendix B.

[12] For ACG (2022), we obtained IRB permission to proceed with the level of obfuscation in the experiment.

been selection into the types of recruiters willing to say yes and participate in an experiment rather than in a job. We return to this question in our results section.

In both studies recruiters were given a structured evaluation form asking for several pieces of information about each candidate. First, they recommend whether each candidate should get a callback and also a take-it-or-leave-it salary offer that should be made to each candidate if they were hired, regardless of callback recommendation. On the next page of the structured assessment subjects were asked for the maximum salary request from the job candidate the firm should accept, which we interpret as the firm's willingness-to-pay for the job candidate. These three items are our primary outcomes, but we also ask for information on the amount of offers that the candidate would accept with 5, 50 and 95 percent chances. We use this as a measure of the recruiters' beliefs about candidates' outside options. Recruiters were also asked whether the candidate will get more than two offers from other firms. We explain these outcomes using economics jargon here, but the questions were worded more colloquially in the structured evaluation form sent to subjects (see Appendix E).

The most important randomization in this paper is whether a recruiter has been told they are in an experiment (recruited by a named researcher with an academic affiliation, informed the firm is fictitious, and required to sign a consent form). Although recruiters in this branch were told they are in an experiment, they were not told the exact nature of the experiment or the hypothesis tested (although, they may have made inferences about it; Orne, 1962; Sears, 1986; Zizzo, 2010). Recruiters in both conditions also received job candidates containing randomized attributes. The attributes of most interest to our analysis are whether the job candidate has a male or female name, whether they disclose their current salary, and the amount they disclose. On all other attributes the job candidates are by design quite similar.

Our fictitious applicants emulate real world software engineers. They worked recently at one of the 13 biggest employers of software engineers from Monster.com and Indeed. We include the real world wage gaps between the higher and lower paying firms as found on payscale.com. Likewise, we emulate real world gender wage gaps when salaries are disclosed — female candidates who work at the same current firm were given a salary that was 85% compared to the male salary at the same current firm.[13] This difference appears on average, and the underlying data in Figure A1 shows we have male and female job candidates reporting a wide array of past salaries.

Table 1 shows summary statistics for recruiter characteristics and outcomes in our experimental conditions, which are discussed in more detail in the next section.

[13]Note no individual recruiter saw a male and female candidate who worked at the same current firm.

3 Results

We focus on three main outcomes: whether the recruiter suggested the applicant should be called back —which mimics the outcome of traditional audit studies— the take-it-or-leave-it salary offer the recruiter suggests, and the maximum salary the employer should pay the applicant.[14] Figure 1 shows how these average outcomes differ for male and female candidates in the two main treatment conditions: not told in experiment (subset of results from ACG (2022)) and told in an experiment.

Overall women are treated relatively better than men when recruiters are told they are in an experiment. This happens because recruiters on average lower callback rates, salary offers, and willingness-to-pay for men when they are told they are in an experiment. Table 2 shows these point estimates from regressions that include controls for recruiter attributes, job candidate fixed effects, and fixed effects for the exact make-up of the eight-candidate packet sent.[15] Average callback probabilities for male candidates decline by 12 percentage points when recruiters know they are in an experiment (an 18% decline over the callback rate in the more standard audit branch of the experiment, $p < 0.05$); and willingness-to-pay is \$4,050 less for men relative to women (a 4% decline over the \$111,826 WTP in the more standard audit branch, $p < 0.05$). Point estimates imply they also give lower salary offers to men as well (2% decline over \$105,351) though this is not significant at conventional levels (p=0.20). Overall, average callback rates, salary offers, and willingness to pay are very similar for women whether or not the recruiter is told they are in an experiment. The difference in equality appears to be driven by the choices about male candidates.

Our experiment also randomized whether the applicant disclosed their current salary or not. In Figure 2 and Table 3 we break down these results by whether the applicant disclosed their salary. We see an interesting suggestive pattern emerge: the callback rate decline for men occurs mainly when salaries are disclosed (representing a 28% decline in callbacks over the mean for men in the audit-study branch of the experiment when salaries are disclosed, $p < 0.10$). When candidates do not disclose a salary there is little difference between callback rates by whether the recruiter is explicitly informed they are in an experiment.

[14] Results for other collected outcomes can be found in the Appendix.

[15] We include controls for recruiter attributes to improve precision even though we are balanced on most elements as shown in Table 1. Additionally, we control for job candidate attributes with job candidate fixed effects, although we are also balanced upon most features of job candidates with for example 50% of job candidates being female, 13% working at Amazon, 6% working at Apple and 13% working at IBM in both the told and not told arms of the experiment. Last, we include fixed effects for the exact eight-candidate packet, which controls for some imbalance we have across treatment arms with 21% getting candidate packets with zero disclosures in not told arm while 32% get these zero disclosure packets in the told in an experiment condition.

When salaries are not disclosed, however, there is a suggestive decline in male salary offers when subjects have been told they are in an experiment versus the natural field experiment setting (point estimate represents a 4% decline, however $p = 0.16$ for salary offers). It may be that recruiters want to be perceived as relatively kinder to women when they are told they are in an experiment, but the device they use to achieve this better treatment of women varies by whether salary is disclosed. Suggestively recruiters use callbacks when salary is disclosed, but amounts (of offers or willingness-to-pay) when salary history is hidden. Most of these differences are not statistically significant at conventional levels.

Appendix Table A2 shows impacts on some of the inputs we asked of recruiters like salary offers they would take with a 5, 50 or 95 percent chance and whether they thought the candidate had more than 2 outside offers. There is not much movement on these outcomes - though they are more likely to say the women have 2+ outside offers when they know they are in an experiment than otherwise (a 23.5% rise, $p < 0.05$).

There are several potential mechanisms that could explain why we get different results when subjects are aware they are in an experiment. One possibility is that the experimental disclosure — which occurred during recruitment — altered selection of subjects into the experiment, and thus the composition of subjects. When we examine observable characteristics in Table 1, we find that most attributes of recruiters in both branches of the experiment are balanced.[16] The acceptance rate is not statistically significantly different for the two arms of the study and is similar to the acceptance rate observed by others using this platform (Coffman et al., 2023). For those recruiters who accepted our task, the completion rate is 87% across both arms. These recruiters are shown the same sets of fictitious job candidates; the distribution of candidate characteristics (e.g. gender, current employer) is therefore also balanced by design. Differential selection does not appear to be a driving factor explaining our results. However if there were any differences in the makeup or acceptance/completion rates of the final subject pool, then it would highlight another possible tradeoff of explicitly telling subjects they are in a study.

Another possibility is that subjects put less effort into the "told experiment" condition because it was not real or they felt it was lower stakes. In our structured assessment the recruiters could leave comments in a notes box when suggesting callbacks, salaries, or WTP;

[16]The one place of imbalance on observables is that in the final subject pool (Panel A) about one-third of those in the told in an experiment arm had feedback on the online freelance platform, while only about one-sixth in the other arm had feedback. The invited subjects in both arms were equally likely to have feedback, about 11% (data available from authors upon request), however it seems subjects with feedback for both treatment arms were more likely to accept and complete our study with that being slightly more the case for the "told in experiment" arm. It is possible that human resource professionals who do not have any feedback yet on the platform are reluctant to get their first feedback for an academic study versus an human resources related task.

at the end of the survey we also asked them whether they reported using any outside research to help them with salary recommendations. Using these as measures of effort, we find that recruiters in the two conditions put in similar effort on average. In addition, we can rule out that "told experiment" subjects reduce outside-research as compared to those "not told" by more than 17% and can rule out they reduce comment-making by more than 27%. In our case, significantly reduced effort is thus not likely to explain our results though we only have these two crude measures of effort.[17]

Other potential mechanisms, as discussed in the Introduction, include social desirability bias, experimenter demand, and effects of observation more generally. (Cameron, 1999; Levitt and List, 2007; McCambridge and Day, 2008; Zizzo, 2010; List, 2011; McCambridge et al., 2014; De Quidt et al., 2018; Jann et al., 2019). Our setting does not allow us to differentiate between these potential mechanisms. But our results do imply caution when interpreting the results of a labor market discrimination study where subjects are aware they are in an experiment. We hope to encourage future research to better understand these mechanisms.

4 Discussion

Our results suggest that informing research subjects that they are in an experiment alters their behavior in comparison to more natural settings, and thus alters the potential conclusions of the research. In our setting, the change goes in a particular direction: male applicants are treated more negatively when subjects are told about the experiment (and more similar to female applicants). For this reason, the level of gender inequality in the "told in experiment" condition is lower than our other condition. Had we relied on this solely, we would underestimate gender discrimination in this market.

The results highlight the potential tradeoff between research design and external validity. While informing subjects about their participation in an experiment may be an ethical consideration, it may also lead to underestimation of the extent of discrimination or other illegal or taboo behaviors in real-world settings. This is particularly relevant for studies on sensitive topics such as racial or gender biases, where subjects may alter their behavior to appear socially desirable or to avoid perpetuating bias.

Our study also contributes to the literature on experimental methodology. The findings

[17] In Table A3 we show that recruiters in both conditions leave a similar amount of comments (7.4 when not told and 7.9 when told, p=0.52, if anything those in the told experiment condition leave *more* comments) and a similar percentage report doing research to help them answer these questions (81.6% when not told and 80.4% when told, p=0.86).

underscore the importance of considering the potential impact of experimental disclosure on subjects' behavior and the subsequent interpretation of research results. The study further demonstrates the value of natural field experiments as tools for studying discriminatory behavior, while also highlighting the limitations and ethical considerations associated with lack of transparency in research.

Several questions remain open for exploration. First, it would be valuable to investigate the mechanisms behind the observed changes in behavior when subjects are informed about the experiment. While in our particular context differential selection of recruiters or differential effort into the told in a study condition do not appear to be driving mechanisms, we could not tease apart other potential mechanisms. In addition to social desirability bias and experimenter demand effects, it is possible that our experimental treatment changes the subject's objectives. A recruiter may approach a real world task with one set of objectives, and may approach participation in an academic experiment with different objectives emphasizing social image concerns or social justice. If recruiters normally provide effort for clients because of the possibility of repeat business with the client, then the academic experiment would change these payoffs. Our treatment may have also changed information without changing objectives. Subjects may feel it was unlikely that an academic would ever present candidates for review in which (say) the men were scored as better – even if they honestly felt that the real world is full of such imbalance. Even if the recruiters were naively trying to pick the "best" candidates, the academic disclosure could alter expectations about which candidates had been pre-screened as "the best."

Second, it would be interesting to explore what attributes people consider "acceptable" to differentiate on, and how this may affect the effectiveness of hiding attributes as a strategy to reduce discrimination. Our main experiment suggests that recruiters do find it acceptable to differentiate on salary history, even though this has a gendered component – although they seem to try to make up for this in other ways in order to compensate women. Studies using methodologies where the subjects know they are in an experiment sometimes find evidence of discrimination on topics like women's age (Low, 2021) and obesity (Macchi, forthcoming), but find mixed evidence on race and gender on other topics (e.g. Kessler et al. (2019) do not find overall discrimination on race/gender in their labor market setting study while they do find it for some subsets of employers; likewise Chan (2022) and Baker and Grimm (2021) find discrimination on race in healthcare choices and the assessment of MTurk workers). Lastly, further experiments could be conducted to assess the impact of experimental disclosure in other contexts.

Our study provides insights into the effects of experimental disclosure on subjects' behav-

ior in the context of gender discrimination in resume screening. The findings have implications for research design and the broader understanding of discrimination in labor markets. While informed consent is a crucial ethical requirement, researchers should be aware of its potential effect on their findings.

References

Adler, Laura, "What's a Job Candidate Worth? Explaining Variation in Pay-Setting Practices," *Working Paper*, 2020.

Agan, Amanda, Bo Cowgill, and Laura Gee, "Salary History and Employer Demand: Evidence from a Two-Sided Audit," *NBER Working Paper 29460*, 2022.

Altonji, Joseph G and Rebecca M Blank, "Race and gender in the labor market," *Handbook of labor economics*, 1999, *3*, 3143–3259.

Baert, Stijn, "Hiring discrimination: an overview of (almost) all correspondence experiments since 2005," in "Audit Studies: Behind the Scenes with Theory, Method, and Nuance," Springer, 2018, pp. 63–77.

Baker, Edmund and Veronika Grimm, "Monitoring Decreases Discrimination," *Available at SSRN 3858441*, 2021.

Bertrand, Marianne and Esther Duflo, "Field experiments on discrimination," in "Handbook of economic field experiments," Vol. 1, Elsevier, 2017, pp. 309–393.

Black, Ines, Sharique Hasan, and Rembrand Koning, "Hunting for talent: Firm-driven labor market search in the United States," *Working paper*, 2020.

Blau, Francine D and Lawrence M Kahn, "The gender wage gap: extent, trends, and explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.

Boggs, Roderick, Joseph Sellers, and Marc Bendick Jr, "Use of testing in civil rights enforcement," in "Clear and convincing evidence: Measurement of discrimination in America," Urban Institute Press Washington, DC, 1993, pp. 345–376.

Cameron, Lisa A, "Raising the stakes in the ultimatum game: Experimental evidence from Indonesia," *Economic Inquiry*, 1999, *37* (1), 47–59.

Cappelli, Peter, "Your approach to hiring is all wrong," *Harvard Business Review*, 2019, *97* (3), 48–58.

Chan, Alex, "Discrimination and Quality Signals: A Field Experiment with Healthcare Shoppers," *Working Paper*, 2022.

Charness, Gary, Anya Samek, and Jeroen van de Ven, "What is considered deception in experimental economics?," *Experimental Economics*, 2022, *25* (2), 385–412.

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter, "Race and economic opportunity in the United States: An intergenerational perspective," *The Quarterly Journal of Economics*, 2020, *135* (2), 711–783.

Coffman, Katherine, Manuela Collis, and Leena Kulkarni, "Whether to Apply," *Working Paper*, 2023.

Connor, Jane Marantz, "Studying racial bias: Too hot to handle?," *Journal of College Science Teaching*, 2000, *30* (1), 26.

Cowgill, Bo and Patryk Perkowski, "Delegation in Hiring: Evidence from a Two-Sided Audit," *Journal of Political Economy: Microeconomics*, forthcoming.

Crabtree, Charles and Kostanca Dhima, "Auditing ethics: A cost–benefit framework for audit studies," *Political Studies Review*, 2022, *20* (2), 209–216.

Derous, Eva, Ann Marie Ryan, and Hannah-Hanh D Nguyen, "Multiple categorization in resume screening: Examining effects on hiring discrimination against Arab applicants in field and lab settings," *Journal of Organizational Behavior*, 2012, *33* (4), 544–570.

Galizzi, Matteo M and Daniel Navarro-Martinez, "On the external validity of social preference games: a systematic lab-field study," *Management Science*, 2019, *65* (3), 976–1002.

Gallen, Yana and Melanie Wasserman, "Informed Choices: Gender Gaps in Career Advice," 2022.

Heckman, James J, "Detecting discrimination," *Journal of economic perspectives*, 1998, *12* (2), 101–116.

_ **and Peter Siegelman**, "The Urban Institute audit studies: Their methods and findings," *Clear and convincing evidence: Measurement of discrimination in America*, 1993, pp. 187–258.

Jamison, Julian, Dean Karlan, and Laura Schechter, "To deceive or not to deceive: The effect of deception on behavior in future laboratory experiments," *Journal of Economic Behavior Organization*, 2008, *68*, 477–488.

Jann, Ben, Ivar Krumpal, and Felix Wolter, "Social Desirability Bias in Surveys–Collecting and Analyzing Sensitive Data. Special Issue.," 2019.

Jr, Christopher Edley, "Implications of empirical studies on race discrimination," 1993, pp. 377–391.

Kessler, Judd B, Corinne Low, and Colin D Sullivan, "Incentivized resume rating: Eliciting employer preferences without deception," *American Economic Review*, 2019, *109* (11), 3713–3744.

King, Eden B, Michelle R Hebl, Whitney Botsford Morgan, and Afra Saeed Ahmad, "Field experiments on sensitive organizational topics," *Organizational Research Methods*, 2013, *16* (4), 501–521.

Kristal, Ariella S, Leonie Nicks, Jamie L Gloor, and Oliver P Hauser, "Reducing discrimination against job seekers with and without employment gaps," *Nature human behaviour*, 2023, *7* (2), 211–218.

Landay, Karen and Sarah DeArmond, "Recruitment Process Outsourcing and Recruiter and Hiring Firm Characteristics," *Journal of Personnel Psychology*, 2018.

Landgrave, Michelangelo, "Can we reduce deception in elite field experiments? Evidence from a field experiment with state legislative offices," *State Politics & Policy Quarterly*, 2020, *20* (4), 489–507.

Laury, Susan K and Laura O Taylor, "Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good?," *Journal of Economic Behavior & Organization*, 2008, *65* (1), 9–29.

Levitt, Steven D and John A List, "What do laboratory experiments measuring social preferences reveal about the real world?," *Journal of Economic perspectives*, 2007, *21* (2), 153–174.

_ **and** _ , "Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments," *American Economic Journal: Applied Economics*, 2011, *3* (1), 224–238.

Lippens, Louis, Siel Vermeiren, and Stijn Baert, "The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments," *European Economic Review*, 2023, *151*, 104315.

List, John A, "The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions," *Journal of political Economy*, 2006, *114* (1), 1–37.

_ , "Why economists should conduct field experiments and 14 tips for pulling one off," *Journal of Economic perspectives*, 2011, *25* (3), 3–16.

Low, Corinne, "Pricing the Biological Clock: The Marriage Market Costs of Aging to Women," *Journal of Labor Economics*, 2021.

Macchi, Elisa, "Worth your weight: experimental evidence on the benefits of obesity in low-income countries," forthcoming.

McCambridge, Jim and Maria Day, "Randomized controlled trial of the effects of completing the Alcohol Use Disorders Identification Test questionnaire on self-reported hazardous drinking," *Addiction*, 2008, *103* (2), 241–248.

_ , John Witton, and Diana R Elbourne, "Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects," *Journal of clinical epidemiology*, 2014, *67* (3), 267–277.

McDaniel, Tanga and Chris Starmer, "Experimental economics and deception: A comment," *Journal of Economic Psychology*, 1998, *19* (3), 403–409.

Neumark, David, "Experimental research on labor market discrimination," *Journal of Economic Literature*, 2018, *56* (3), 799–866.

Orne, Martin T, "On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications.," *American psychologist*, 1962, *17* (11), 776.

Ortmann, Andreas and Ralph Hertwig, "The costs of deception: Evidence from psychology," *Experimental Economics*, 2002, *5*, 111–131.

Pager, Devah and Lincoln Quillian, "Walking the talk? What employers say versus what they do," *American Sociological Review*, 2005, *70* (3), 355–380.

Quidt, Jonathan De, Johannes Haushofer, and Christopher Roth, "Measuring and bounding experimenter demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

Riach, Peter A. and Judith Rich, "Deceptive Field Experiments of Discrimination: Are they Ethical?," *KYKLOS*, 2004, *57* (3), 457–470.

Salganik, Matthew J, *Bit by bit: Social research in the digital age*, Princeton University Press, 2019.

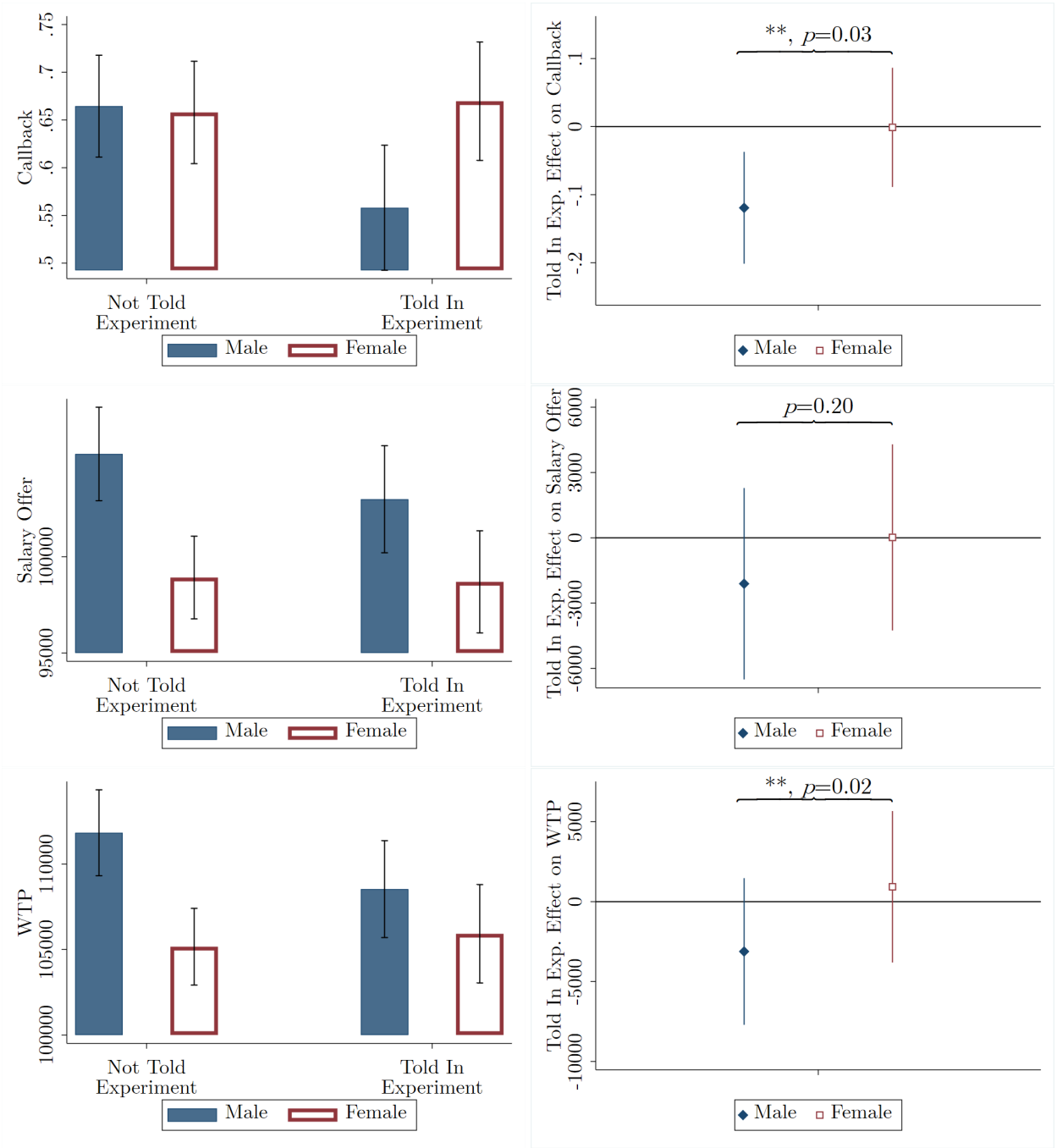
Sears, David O, "College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature.," *Journal of personality and social psychology*, 1986, *51* (3), 515.

Wulff, Jesper N and Anders R Villadsen, "Are survey experiments as valid as field experiments in management research? An empirical comparison using the case of ethnic employment discrimination," *European Management Review*, 2020, *17* (1), 347–356.

Zizzo, Daniel John, "Experimenter demand effects in economic experiments," *Experimental Economics*, 2010, *13*, 75–98.

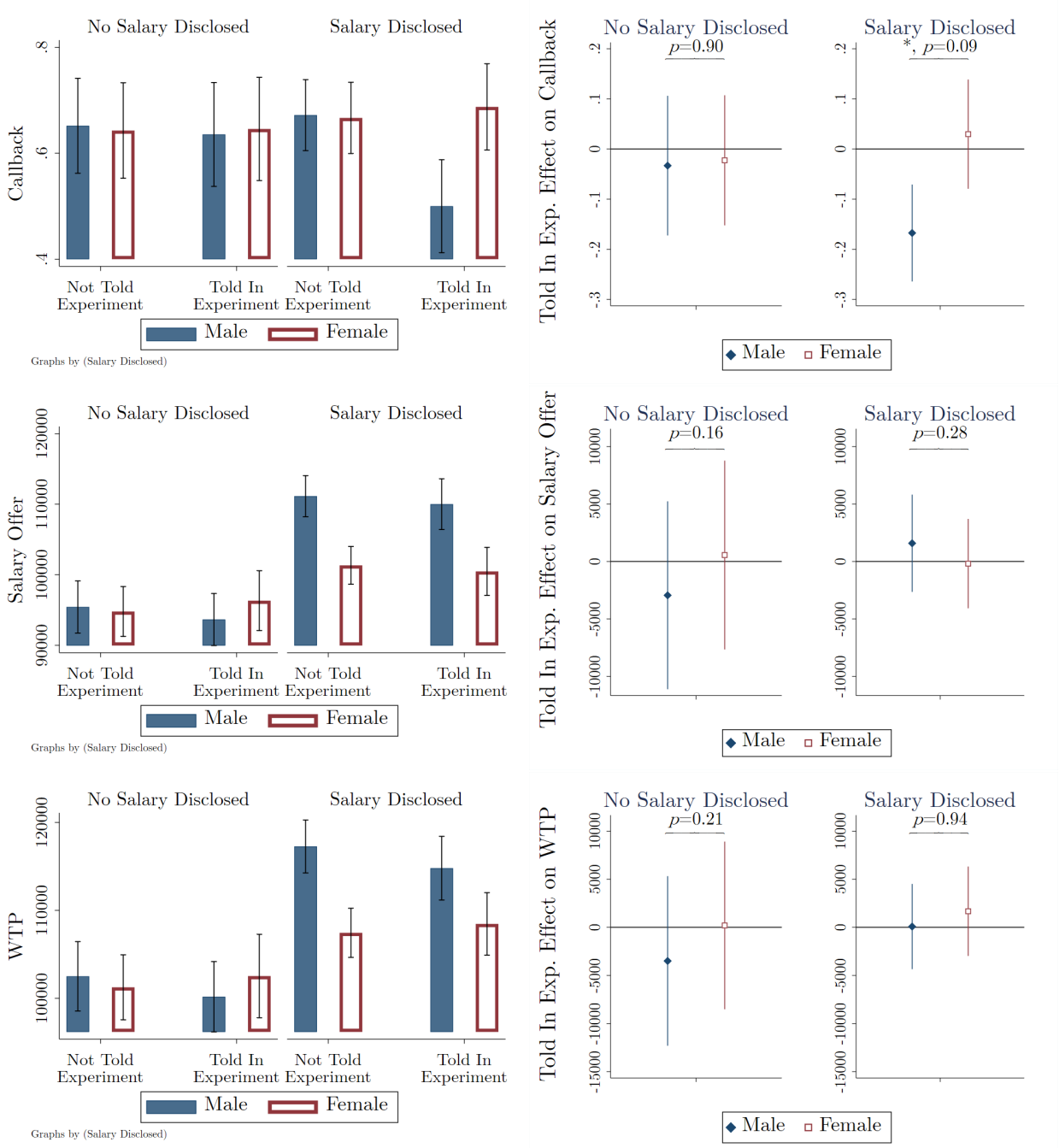
Zschirnt, Eva, "Research ethics in correspondence testing: An update," *Research Ethics*, 2019, *15* (2), 1–21.

Figure 1: Outcomes and Effects By If Told You Are In An Experiment



Notes: WTP denotes Willingness-to-Pay. On the left-hand-side of this figure we show the mean of our outcomes for male and female candidates by whether the recruiter evaluating the job candidate was told they were in an experiment or not. On the right-hand-side we show the effects of being Told You're In An Experiment on outcomes for female and male job candidates from our models reported in Table 2 with p-values and statistical significance of the tests of the effects for male versus female candidates denoted by * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$ (see results on "(Told Exp.)*(Male Candidate)".)

Figure 2: Outcomes and Effects By If Told You Are In An Experiment & Disclosure



Notes: WTP denotes Willingness-to-Pay. On the left-hand-side of this figure we show the mean of our outcomes for male and female candidates by whether the recruiter evaluating the job candidate was told they were in an experiment or not and by whether salaries were disclosed. On the right-hand-side we show the effects of being Told You're In An Experiment on outcomes for female and male job candidates and by whether salaries were disclosed from our models reported in Table 3 with p-values and statistical significance of the tests of the effects for male versus female candidates denoted by * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$ (see results on "(Told Exp.)*(Male Candidate)" and "(Told Exp.)*(Salary Disclosed)*(Male Candidate)".)

Table 1: Recruiter Attributes and Summary Statistics By If Told You Are In An Experiment

	(1) Not Told Experiment	(2) Told Experiment	(3) t-test p-value
	Panel A: Final Subject Pool		
3+ Yrs Exp	0.70	0.70	0.99
Black Recruiter	0.23	0.16	0.38
White Recruiter	0.57	0.56	0.91
Female Recruiter	0.71	0.69	0.74
Hourly Rate	46.00	45.08	0.83
Total Hours Billed	103.28	73.79	0.74
Recruiter Has Feedback	0.17	0.32	0.04
Avg. Feedback (1 to 5) | Feedback	4.86	4.83	0.73
Subject Observations	76	56	132
	Panel B: Summary Statistics		
Male Candidates			
Callback	0.66	0.56	0.01
Salary Offer	105,351	102,997	0.36
Willingness to Pay (WTP)	111,826	108,531	0.20
Female Candidates			
Callback	0.66	0.67	0.78
Salary Offer	98,924	98,698	0.92
Willingness to Pay (WTP)	105,164	105,920	0.77
Candidate Observations	608	448	

Notes: In Panel A "3+ Year Exp" denotes if the subject self-reported at least 3 years of experience in hiring. Subjects also self-report their race. All other variables in Panel A were available from the platform and were measured before subjects did our task. The exception is the variable "Female Recruiter" which was coded based on the subject's name. Self-reported gender and name-coded gender had a correlation above 0.90 and either variable is balanced across experimental conditions. "Hourly Rate" is the hourly rate in dollars subjects provide to the platform, "Total Hours Billed" is the number of hours they have worked so far on the platform, "Has Feedback" is a 0/1 variable for whether the subject has feedback on the platform, and "Avg. Feedback (1 to 5) | Feedback" denotes their average feedback score conditional on having any feedback on the platform. See Table A1 for comparison to all the 256 subjects reported on in ACG (2022) of which those in column (1) are a subset of 76 which viewed the same job candidates as those in column (2).
Panel B shows the average outcomes provided by our subjects when they evaluate male or female job candidates by whether the subject was told they were in an experiment or not. In Panel B p-values are presented from regressions clustering standard errors at the recruiter-level. The 132 final subjects evaluate eight candidates each for a total of 1056 job candidate level observations with equal numbers of job candidates being male and female.

Table 2: Gender Gap In Outcomes By Told In An Experiment

	Callback	Salary Offer	WTP
(Told Exp.)	-0.01	18	931
	(0.05)	(2,161)	(2,395)
(Told Exp.)*(Male Candidate)	-0.12**	-2,128	-4,050**
	(0.05)	(1,641)	(1,750)
Male Told Exp. Effect:			
Total	-0.12	-2,110	-3,118
p-value	0.01	0.34	0.18
Mean Told Experiment:			
Female Told Exp.	0.67	98,698	105,920
Male Told Exp.	0.56	102,997	108,531
Female:Male Told Exp. Ratio	1.20	0.96	0.98
Mean Not Told Experiment:			
Female Not Told Exp.	0.66	98,924	105,164
Male Not Told Exp.	0.66	105,351	111,826
Female:Male Not Told Exp. Ratio	0.99	0.94	0.94
R^2	0.05	0.26	0.23
Observations	1056	1056	1056

Notes: WTP denotes Willingness-to-Pay. Each of our 132 recruiters evaluated eight candidates for a total of 1056 candidate level observations. Regressions include controls for recruiter attributes, job candidate fixed effects, and fixed effects for the exact make-up of the eight-candidate packet sent. Standard errors are clustered at the recruiter level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Table 3: Gender Gap In Outcomes By Told In An Experiment & Disclosure

	Callback	Salary Offer	WTP
(Told Exp.)	-0.02	560	197
	(0.07)	(4,154)	(4,403)
(Told Exp.)*(Male Candidate)	-0.01	-3,504	-3,690
	(0.08)	(2,501)	(2,950)
(Told Exp.)*(Salary Disclosed)	0.04	-1,446	756
	(0.09)	(4,429)	(4,765)
(Told Exp.)*(Salary Disclosed)*(Male Candidate)	-0.19*	3,372	288
	(0.11)	(3,132)	(3,679)
(Salary Disclosed)	0.05	3,352	3,438
	(0.08)	(2,353)	(2,501)
(Salary Disclosed)*(Male Candidate)	0.01	9,161***	8,711***
	(0.07)	(1,835)	(2,124)
Male Told Exp. Effect without Disclosure :			
Total	-0.03	-2,944	-3,493
p-value	0.64	0.48	0.43
Mean Told Exp. No Disclosure:			
Female Told Exp. No Disclose	0.65	96,328	102,531
Male Told Exp. No Disclose	0.64	93,656	100,177
Female:Male Ratio Told Exp. No Disclose	1.02	1.03	1.02
Mean Don't Told Exp. No Disclosure:			
Female Not Told Exp. No Disclose	0.64	94,792	101,250
Male Not Told Exp. No Disclose	0.65	95,435	102,500
Female:Male Ratio Not Told Exp. No Disclose	0.99	0.99	0.99
Male Told Exp. Effect with Disclosure:			
Total	-0.18	-1,018	-2,450
p-value	0.01	0.64	0.28
Mean Told Exp. with Disclosure:			
Female Told Exp. with Disclose	0.69	100,476	108,461
Male Told Exp. with Disclose	0.50	110,003	114,797
Female:Male Ratio Told Exp. with Disclose	1.38	0.91	0.94
Mean Not Told Experiment with Disclosure:			
Female Not Told Exp. with Disclosure	0.67	101,333	107,448
Male Not Told Exp. with Disclosure	0.67	111,135	117,266
Female:Male Ratio Not Told Exp. with Disclosure	0.99	0.91	0.92
R^2	0.05	0.29	0.25
Observations	1056	1056	1056

Notes: WTP denotes Willingness-to-Pay. Each of our 132 recruiters evaluated eight candidates for a total of 1056 candidate level observations. Regressions include controls for recruiter attributes, job candidate fixed effects, and fixed effects for the exact make-up of the eight-candidate packet sent. Standard errors are clustered at the recruiter level. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Appendix: For Online Publication Only

A Appendix Tables and Figures

Figure A1: Distribution of Disclosed Salary Amounts by Told In An Experiment



Notes: This figure shows the distribution of male and female salaries (among the job applications containing a disclosure) by whether the subject was told they are in an experiment or not.

Table A1: Comparison to Full Agan et al. (2022) Data Set ($N = 256$)

	(1) Not Told Full	(2) Not Told Subset	(3) Told Experiment	(4) p-val. (1) vs. (3)	(5) p-val. (2) vs. (3)
Male Candidates					
Callback	0.63	0.66	0.56	0.03	0.01
Salary Offer	104,588	105,351	102,997	0.42	0.36
Willingness to Pay (WTP)	110,814	111,826	108,531	0.25	0.20
Female Candidates					
Callback	0.63	0.66	0.67	0.29	0.78
Salary Offer	97,327	98,924	98,698	0.48	0.92
Willingness to Pay (WTP)	103,388	105,164	105,920	0.25	0.77
Candidate Observations	1056	608	448		
Subject Observations	256	76	56		

Notes: This table compares means from the full $N = 256$ subjects reported on in Agan et al. (2022) to the subset $N = 76$ reported on in this paper which viewed the exact same job candidate packets as the $N = 56$ subjects who were told they were in an experiment. The p-values are presented from regressions clustering standard errors at the recruiter-level. There are equal numbers of male and female job candidates.

Table A2: Additional Outcomes Gender Gap In Outcomes By If Told You Are In An Experiment

	5pct Accept	50pct Accept	95pct Accept	2+ Offers
(Told Exp.)	556.28	-206.89	3,605.47	0.12**
	(1,662.64)	(1,815.33)	(4,189.88)	(0.05)
(Told Exp.)*(Male Candidate)	-942.32	-809.43	-4,292.53	-0.06
	(1,551.68)	(1,560.48)	(4,017.19)	(0.06)
Male Told Exp. Effect:				
Total	-386.04	-1,016.32	-687.06	0.06
p-value	0.81	0.59	0.79	0.27
Mean Told Experiment:				
Female Told Exp.	85,428.13	96,045.76	112435.71	0.60
Male Told Exp.	91,835.80	103451.79	116302.23	0.54
Female:Male Told Exp. Ratio	0.93	0.93	0.97	1.11
Mean Not Told Experiment:				
Female Not Told Exp.	85,659.87	97,150.99	108668.59	0.51
Male Not Told Exp.	93,009.87	105366.45	116827.63	0.51
Female:Male Not Told Exp. Ratio	0.92	0.92	0.93	1.00
R^2	0.40	0.38	0.14	0.04
Observations	1056	1056	1056	1056

Notes: We asked our subjects to please tell us a salary offer each candidate would accept with 5pt, 50pct, and 95pct probability. We also asked is each candidate would have "zero or one" or "two or more" offers. Each of our 132 recruiters evaluated eight candidates for a total of 1056 candidate level observations. Regressions include controls for recruiter attributes, job candidate fixed effects, and fixed effects for the exact make-up of the eight-candidate packet sent. Standard errors are clustered at the recruiter level. * 0.10 ** 0.05 *** 0.01

Table A3: Real Effort By If Told You Are In An Experiment

	Not Told Experiment	Told Experiment	Difference	t-test p-value	95% Confidence Interval
Reported Research	0.816	0.804	0.012	0.86	(-0.14, 0.12)
Total Comments	7.408	7.911	-0.503	0.52	(-2.05, 1.05)
Observations	76	56	132	132	132

Notes: Reported Additional Research takes the value 1 if a subject reported any of the following actions at the end of the structured evaluation form "Looked up salaries on a website like Payscale.com, glassdoor.com or others," or "Spoke with others who are familiar with salaries for software engineers." Total Comments on Callback/Salary/WTP is a count of the times a subject left a comment in the open ended "Notes" box when reporting Callback/Salary offer simultaneously, or WTP on the next page of structured evaluation form. There are a total of 16 opportunities for comments to be left by each subject on these topics.

B Obfuscation Considerations in "Not Told" Arms

Our research design is an extension of the correspondence audit methodology (Bertrand and Mullainathan, 2004; Gaddis, 2018) in which researchers send realistic-looking fake candidates to employers' job openings. We extend this design in a way that improves upon some – albeit not all – of the ethical liabilities of audit studies. Ethical critiques of audit studies have several facets (Hamermesh, 2012), and our design primarily addresses two important aspects.[1]

1) **Wasted Effort**. The most common criticism of resume audits is that researchers waste employers' time on fictitious candidates. Researchers cannot compensate employers for wasted effort without exposing the experimental intervention.

   Our design handles this directly: Subjects are paid for their time at rates they agree to. We do not disclose to subjects they are in a research study, and we instead lead them to believe they are performing their normal jobs as professional, subcontracted recruiters working on behalf of a client.

   However, they are paid for their time at rates they set. This setup allows researchers to compensate subjects under natural, non-laboratory conditions.

2) **Adverse Learning**. A second ethical concern is that audit researchers will inadvertently teach employers to associate certain candidates with non-response. Our design also includes protections against this.

   First, all inputs and outcomes that are visible to the recruiters were simulated to resemble real-world counterparts in order to preserve realistic associations.[2]

   Second (and more importantly), we do not give our recruiter subjects feedback about individual candidates' outcomes.[3] As such, subjects are blocked from the information necessary to change their beliefs.

   These two measures are intended to preserve realistic associations and to contain information that would alter the subject's beliefs or future behavior.

Although these measures may not address all possible side effects of lack of transparency, we hope they offer significant relief on the main issues above. A large survey of economists (Charness et al., 2022) shows researchers are reluctant to use obfuscation, but believe that some obfuscation is acceptable for important topics without credible alternative research designs. The absence of credible alternatives may be particularly binding for researchers studying hiring discrimination, which is both illegal and repugnant. A researcher hoping

[1] Our discussion in this section is adapted from a working paper by Cowgill and Perkowski (forthcoming), which contains a similar field experiment.

[2] Our resumes were based on real resumes, and the bonuses were based on simulated outcomes based on data from a real company like the one in our experiment.

[3] Recruiters in our design do receive a bonus which is based partly on candidates' outcomes. This is given in a lump sum aggregate over all applications; We did not break out the bonus by each candidate. The lump-sum bonus is thus a weak signal about any single candidate's outcomes.

to randomize screening practices at a real firm may develop a paper trail later used in court against the firm. Anticipating this, firms are reluctant to facilitate research partnerships on key questions. As Oyer and Schaefer (2011) wrote, "What manager would allow an academic [...] to experiment with the firm's screening, interviewing or hiring decisions?"

Our design offered us a way to reduce the drawbacks of the traditional audit methodology (wasted effort and adverse learning) in settings where the lack of alternatives make it worth considering. In general, we believe researchers should remain cautious about using obfuscation and consider alternative approaches to address research questions.

B.1 Example Invite by Condition

Table B1: Invite Message By If Told You Are In An Experiment

Condition	Invite Message
Not Told In Experiment	Dear [redacted] I looked at your profile and you seemed like a good fit for this job. I wanted to invite you directly and move forward if you're available. If you're interested, the job is yours. Just send us your LinkedIn profile (or resume) and sign the attached NDA form to get started. We are hiring a junior software engineer at our company. We have eight applications for you to review out of our applicant pool, and would like some brief feedback about who we should interview and what salaries we should offer (and accept). Our questions will require you to use your best judgment based on your hiring expertise. In our application, we ask the same set of questions to all candidates and the job applications will be provided to you in a standardized format. We will ask you to review these applications, and provide answers to multiple choice questions and salary suggestions for each candidate - no essay responses required! You can send your answers back via a link we'll provide you. For your assistance with this task, you will be paid hourly (with a maximum of 2 hours allowable), plus a bonus. You should be either based in the US, or familiar with US institutions and culture (such as US schools and employers) so that you can better understand the candidates' qualifications. We ask that you NOT contact the candidate, and instead send us your feedback after reading the job applications. To accept this offer, read our brief instructions doc (attached as a PDF to this message). Then sign the attached NDA form and return it to us with your LinkedIn profile or resume.
Told In Experiment	Dear [redacted] We looked at your profile and you seemed like a good fit for this task. I wanted to invite you directly and move forward if you're available. If you're interested, the task is yours as long as you're qualified. Just send us your LinkedIn profile (or resume), and sign the attached Consent form and NDA form to get started. We are professors at Columbia University, Rutgers University and Tufts University who are studying recruitment procedures for companies that hire software engineers. To further our research, we have developed a survey to ask you to evaluate applications to a hypothetical software company. While the choices you make are not real, our research endeavors to help companies and recruiters optimize their work, so please respond as if this was a real recruiting/screening task in order to make the research results as helpful as possible to recruiters and companies. In addition, we will pay you a bonus based on data from how real companies would evaluate these candidates. We want you to imagine that the company is hiring a junior software engineer. The company has eight applications for you to review out of their applicant pool, and the company would like some brief feedback about who they should interview and what salaries they should offer (and accept). The questions will require you to use your best judgment based on your hiring expertise. The hypothetical company asks the same set of questions to all candidates so the job applications will be provided to you in a standardized format. We will ask you to review these applications, and provide answers to multiple choice questions and salary suggestions for each candidate - no essay responses required! You can send your answers back via a link we'll provide you. For your assistance with this task, you will be paid hourly (with a maximum of 2 hours allowable), plus a bonus. You should be either based in the US, or familiar with US institutions and culture (such as US schools and employers) so that you can better understand the candidates' qualifications. We ask that you send us your feedback after reading the job applications. To accept this offer, read our brief instructions doc (attached as a PDF to this message). Then sign the attached consent form, example NDA form (this is not legally binding) and return those to us with your LinkedIn profile or resume.

C Example Job Candidate Materials (same for both the Told Experiment and Not Told Experiment Conditions)

C.1 Sample Job Application: Salary History Asked + Candidate Discloses

Samantha [redacted]

Application Details for Software Engineering position

*All sections are **required** except where noted. For candidates who are interviewed, all information entered below will be verified.*

Candidate Information

Candidate Id: 774 **Mailing Address**: [redacted] **City/State**: [redacted]

ZIP: [redacted] **Phone**: ([redacted]) [redacted]-[redacted] **Email**: [redacted] **URL**: http://[redacted]

Are you legally authorized to work in the US? Y **Are you over the age of 18?**: Y

Are you willing to relocate for this position? Y **Will you now (or in the future) require visa sponsorship?** N

Employment History (Last Three Jobs)

Title: Software Engineer **Company Name**: IBM **Location**: San Jose, CA **Dates**: 01/2015 - Present

Position Description, Duties, Responsibilities:
* Developing and implementing new feedback system for user concerns, bugs, and defect tracking regarding use and functionality of new interfaces.
* Coding web designed interfaces using Java, XML, XSL, AJAX, and JWS.
* Implement the command-line interface for the Universal Authentication Protocol (UAP) in E-directory.

Title: Software Developer **Company Name**: Amazon **Location**: Seattle, WA **Dates**: 05/2014 - 01/2015

Position Description, Duties, Responsibilities:
* Developed code and unit tests in Python for server-side and in JavaScript for web components.
* Deployed and tested code on Linux-based EC2 instances in a distributed AWS cloud environment.
* Created and maintained automated jobs to build and test software.
* Developed and implemented working plans for the formulation of front and back-end web applications.
* Developed various algorithms to mitigate program interference.

Title: Programming Intern **Company Name**: Intraix **Location**: Ayer Rajah Crescent, SG **Dates**: 05/2013 - 08/2013

Position Description, Duties, Responsibilities:
Automated black box and white box tests for an Android application “Klug,” using Appium and Espresso framework. This helped developers expand features without much worry of breaking current functionalities.

Salary History

Annual Base Salary at Current or Most Recent Job: $96,000

Education

Institution: Georgia Institute of Technology **Location**: Atlanta, GA **Dates**: 2010 - 2014 **Graduated?** Y

Level: BS (Bachelor of Science) **Subject/Major**: Computer Science

Relevant Coursework:
Database and Information Management Systems, Java, Analysis of Algorithms, Data Sytems, Matlab for Programmers, and Compiler Design

Additional Skills and Information

Experience developing in Java, HTML/CSS, JavaScript, Node.js, Ruby, Ruby on Rails, Shell, Python, SQL, LATEX.

1

C.2 Sample Job Application: Salary History Asked + Candidate Does Not Disclose

Christopher [redacted]

Application Details for Software Engineering position

*All sections are **required** except where noted. For candidates who are interviewed, all information entered below will be verified.*

Candidate Information

Candidate Id: 721 **Mailing Address**: [redacted] **City/State**: [redacted]

ZIP: [redacted] **Phone**: ([redacted]) [redacted]-[redacted] **Email**: [redacted] **URL**: http://[redacted]

Are you legally authorized to work in the US? Y **Are you over the age of 18?**: Y

Are you willing to relocate for this position? Y **Will you now (or in the future) require visa sponsorship?** N

Employment History (Last Three Jobs)

Title: Programmer **Company Name**: Apple **Location**: Cupertino, CA **Dates**: 10/2015 - Present

Position Description, Duties, Responsibilities:
Research, design, and implement scalable applications for information identification, extraction, analysis, retrieval, and indexing. Direct software design and development while remaining focused on client needs. Collaborate closely with other team members to plan, design, and develop robust solutions. Maintain front-end admin interface as well as back data processing.

Title: Programmer **Company Name**: Verizon Communications, Inc. **Location**: New York, NY **Dates**: 07/2014 - 10/2015

Position Description, Duties, Responsibilities:
Designed, developed, and integrated software with test systems hardware for test engineering applications. Supported the design and testing of space systems software in all program phases, from initial design through coding, testing, and integration. Member of team responsible for developing a new high-end software package. Led team of 3 engineers to manage Windows client (C++) including feature development, debugging, and update release.

Title: Summer Programming Associate **Company Name**: Facebook **Location**: Menlo Park, CA **Dates**: 06/2013 - 08/2013

Position Description, Duties, Responsibilities:
Intern on the Sales Platform team within Core Ads, which deals primarily with making tools to help salespeople make sales, usually by connecting them to advertisers. Worked on improving the infrastructure and data quality of our platform that helps sales teams find their clients. Languages/technologies: Hack (PHP), Python, Dataswarm.

Salary History (optional)

Annual Base Salary at Current or Most Recent Job:

Education

Institution: California Institute of Technology **Location**: Pasadena, CA **Dates**: 2010 - 2014 **Graduated?** Y

Level: BS (Bachelor of Science) **Subject/Major**: Computer Science

Relevant Coursework:
Artificial language, hardware systems, analysis of algorithms. programming abstractions, data structures and algorithms

Additional Skills and Information

Production code launched using C/C++, Java, Javascript, Python, Perl. Back-end and research experience using Linux shell scripting, R, PiCloud/Multivac, Sawzall, MapReduce.

C.3 Sample Job Application: Salary History Not Asked + Candidate Does Not Disclose

Sarah [redacted]

Application Details for Software Engineering position

*All sections are **required** except where noted. For candidates who are interviewed, all information entered below will be verified.*

Candidate Information

Candidate Id: 1724 **Mailing Address**: [redacted] **City/State**: [redacted]

ZIP: [redacted] **Phone**: ([redacted]) [redacted]-[redacted] **Email**: [redacted] **URL**: http://[redacted]

Are you legally authorized to work in the US? Y **Are you over the age of 18?**: Y

Are you willing to relocate for this position? Y **Will you now (or in the future) require visa sponsorship?** N

Employment History (Last Three Jobs)

Title: Coder **Company Name**: Facebook **Location**: Menlo Park, CA **Dates**: 06/2014 - Present

Position Description, Duties, Responsibilities:
Enhancing existing web applications to meet current standards. Constructing complex queries using SQL in the IBM DB2 Database. Designing technical structure and modules for a new and better UX. Collaborating with senior developers to execute client work. Introducing automated acceptance and unit tests, while increasing coverage.

Title: Software Architect **Company Name**: Dell **Location**: Round Rock, TX **Dates**: 06/2013 - 06/2014

Position Description, Duties, Responsibilities:
Participate in application modification and development of new applications to meet business needs. Provide full life-cycle project expertise. Project work focused on business applications and e-business solutions. Responsibilities included application integration and development using .NET including C#, ASP.Net, WinForms, MS Exchange, and Microsoft Sharepoint Portal Server.

Title: Summer Coding Fellowship **Company Name**: Apple **Location**: Cupertino, CA **Dates**: 05/2012 - 08/2012

Position Description, Duties, Responsibilities:
Built an automated framework on the Apple Maps Team for validating the internal pipeline that manages how different layers of maps data integrate using Python.

Education

Institution: Cornell University **Location**: Ithaca, NY **Dates**: 2009 - 2013 **Graduated?** Y

Level: BS (Bachelor of Science) **Subject/Major**: Computer Science

Relevant Coursework:
Systems Programming and Machine Organization, Privacy and Technology, Data Science I, Networks, Computing Hardware, Cloud Computing.

Additional Skills and Information

Skills: JS, Java, XPages, Flex / AIR, Processing, Git, Eclipse, HTML.

C.4 Sample Job Application: Salary History Not Asked + Candidate Volunteers

Tyler [redacted]

Application Details for Software Engineering position

*All sections are **required** except where noted. For candidates who are interviewed, all information entered below will be verified.*

Candidate Information

Candidate Id: 621 **Mailing Address**: [redacted] **City/State**: [redacted]

ZIP: [redacted] **Phone**: ([redacted]) [redacted]-[redacted] **Email**: [redacted] **URL**: http://[redacted]

Are you legally authorized to work in the US? Y **Are you over the age of 18?**: Y

Are you willing to relocate for this position? Y **Will you now (or in the future) require visa sponsorship?** N

Employment History (Last Three Jobs)

Title: Developer **Company Name**: Amazon **Location**: Seattle, WA **Dates**: 02/2014 - Present

Position Description, Duties, Responsibilities:
- Develop automated REST API test cases to ensure proper error handling.
- Conduct regression tests on internal and external products and services in order to successfully integrate new solutions to existing systems.
- Review and approve code releases from development and marketing departments. ensure thorough client policy compliance.

Title: Coder **Company Name**: Google **Location**: Mountain View, CA **Dates**: 05/2013 - 02/2014

Position Description, Duties, Responsibilities:
- Researched emerging technologies for database and network storage solutions by reviewing case studies and functionality to determine low-cost, but effective, models for supported environments.
- Provided leadership and decision making to impact infrastructure changes that included upgrading the Oracle database schema, applying new versions of Dart Enterprise, and implementing a virtualized hardware environment to reduce footprint and minimize data center presence.

Title: Software Development Trainee **Company Name**: GE Healthcare **Location**: Little Chalfont, UK **Dates**: 05/2012 - 08/2012

Position Description, Duties, Responsibilities:
Reduced waiting time to pull information from multiple systems - requests that used to take days, now only take minutes. Also worked closely with other IT professionals to design, test, and implement APIs in support of major ERP systems.

Education

Institution: University of Illinois at Urbana-Champagne **Location**: Champaign, IL **Dates**: 2009 - 2013 **Graduated?** Y

Level: BS (Bachelor of Science) **Subject/Major**: Computer Science

Relevant Coursework:
C++, Java, Microprocessor systems, Cryptography, Human-computer interface technology, Computer networks, and Large scale systems

Additional Skills and Information

Skilled in Python (Django), Java, Ruby on Rails, JavaScript (AngularJS, jQuery), SQL, PHP, HTML, CSS. I make about $125,000 per year right now (pre-bonus).

1

D Instructions (same for both the Told Experiment and Not Told Experiment Conditions)

Instructions

Thank you for your help screening our candidates. Please read these instructions carefully and completely before you begin this task.

1 About our Hiring Needs

We are interested in finding candidates for a full-stack software engineering position at a mid-sized software start-up company. Qualified candidates should have a working understanding of hardware systems infrastructure, creating and manipulating databases, writing back-end code in one or more languages (e.g., Ruby, Java, Python, C#), and writing front-end code in one or more languages (e.g., HTML, Javascript). Other responsibilities may include project management and technical documentation. Our company has locations in several cities throughout the United States and many of our software engineers work remotely; location will be determined in consultation with the candidate after an offer has been made.

Additional details about our opening are available in section 5.B.

2 Your Task

We will provide you with candidates' responses to our online job application form. We ask that you review this information and answer a few questions. In particular, we will ask you about:

- Whether we should interview the candidate
- What salary we should offer or accept if they pass our interview
- Additional questions about potential salary ranges

At this stage, we are interested in identifying worthy candidates. In that sense, we do not have a fixed number of positions so you should let us know about any candidate in our applicant pool that would be a good match for this position.

Software engineers currently at our firm make between $70,000 and $120,000. You should not feel constrained by our current range, and we welcome your own research about what candidates should be paid. We also offer benefits including health insurance, stock and a performance-based annual bonus. However, our questions for you today will be about the cash component (annual base salary) of compensation only.

2 Compensation for you, our recruiter, for this task

For your assistance with this task, you will be paid hourly (with a maximum of 2 hours allowable), plus a bonus. You can read the details about the bonus calculation in the appendix to these instructions, but we'll summarize it here:

1. We care about spending recruiting energy on candidates we're likely to hire -- candidates who will impress us in interviews and will accept our offers.

2. We care about the difference between what we pay candidates and the value they bring to our company. It's worth paying more for salaries, but only if they bring more value (and/or if they're more likely to accept). We want your decisions to consider value, cost, and probability of acceptance.

We will interview all candidates you suggest. We may also interview candidates you did not suggest upon recommendation from others at our company.

Please note, we do not negotiate salaries with candidates.

3 Your Feedback about the Candidates

We will provide an online form for you to fill in your evaluations to make it easier to work together without too much back-and-forth. There will be six sets of questions about the candidates themselves, and a few quick questions about yourself.

4 Additional Information

Ultimately our staff are very busy and not available to answer questions as you review these applications.

Please do not contact any of these candidates. We are asking you only to evaluate them and send us your private assessments. Someone from our staff will take the next step with the candidates. To prohibit you from contacting them, we have blacked out their contact information in the attached application forms.

Our hiring philosophy is to make interview decisions based on what is submitted. Therefore, please do not consult any information on individual candidates outside of the packets we send you. For example, do not look up the candidates on Google or LinkedIn.

5 Appendix

The remainder of this document includes:

- Some additional details about your bonus payment.
- Additional information about the job requirements for full-stack software engineer.

5.A Exact Formula for Calculating Your Bonus

We will calculate a bonus associated with each candidate you review according to the guidelines below, and then sum them up across all candidates and pay you the full sum in addition to your hourly rate. The bonuses will be paid after we have completed our interview and hiring decisions - approximately 45 days (or sooner) after your complete the task.

For candidates who are hired, we will examine their performance and trajectory about four weeks after the candidate starts work. We'll rate the newly hired candidate on three dimensions using the one-through-three scale outlined on the next page.

We will add up the candidate's three scores, for a total score ranging between 3 and 9. We then multiply that total score by five, and subtract [the candidate's salary / 100,000]. This is your bonus for each newly hired candidate.

Hired Candidate Bonus= 5 x (Technical Score + Innovation Score + Leadership Score) - Salary/100,000

As you know, you'll help set our workers' salaries through your feedback in this task.[1] This bonus gives you the incentive to find candidates who deliver a lot of value to our company above the salary we need to pay them.

For candidates who are NOT hired -- either because we don't make them an offer, or because they reject our offer -- your Hired Candidate Bonus for that candidate will be zero.

We will also subtract $5 from your overall bonus for everyone you suggest interviewing who isn't hired. This is to encourage you to be a little bit selective about forwarding candidates who have a realistic shot at joining our company. If we hire someone who you didn't suggest interviewing, we'll calculate the Hired Candidate Bonus as if you suggested interviewing that candidate. Also: If you suggest interviewing a candidate and the candidate declines to be interviewed, we would count this as a failed interview.

Please note: It would (in theory) be possible to earn a negative overall bonus. If this happens, we will set the overall bonus to $0.

[1] In one of our questions for you, we'll ask you what we should offer the candidate as a take-it-or-leave-it offer. For candidates who accept, we'll use that salary in the bonus calculation. We'll also ask you what to do if a candidate instead approaches us with a take-it-or-leave-it offer. If you guide us to accept those offers in some circumstances, then we'll use those salaries in the formula above.

Evaluation Dimensions

A. Technical Score
B. Innovation Score
C. Leadership Score

Examples of Performance in Each Dimension

A. Technical Score:

Rating 1 (Low): Gaining command of all core technologies and practices used in our firm's engineering team. Able to begin developing and productionizing low to moderate complexity modules.

Rating 2 (Middle): Reasonable command of core engineering systems. Shows comfort with owning reasonably high complexity modules.

Rating 3 (High): Responsible for driving, technically designing, implementing and productionizing high impact projects with the help of teams if needed. Can own and deliver on very large mission-critical projects that impact the company in a verifiable way.

B. Innovation Score:

Rating 1 (Low): Responsible for implementing specifications developed by senior engineers and product managers. Does not develop products.

Rating 2 (Middle): Develops incrementally innovative ideas that can be successfully patented. Does not take leadership of developing new products, features and lines of business.

Rating 3 (High): Develops patentable ideas that lead to breakthrough improvements. Comes up with ideas to expand their projects and may also have a reasonable free-hand in developing and executing on them.

C. Leadership Score:

Rating 1 (Low): Tech, design or architectural lead of a small team/project, but could not have direct reports.

Rating 2 (Middle): Be able to mentor engineers in the team, giving technical guidance, code reviews, and ultimately be able to take responsibility of delivering small projects end-to-end on production.

Rating 3 (High): Leads complex initiatives and technically drives teams towards implementing and productionizing them. Promotes professional growth and development inside and outside the team. Actively takes steps to increase technical excellence across the organization.

5.B Additional Information about Job Opening for a Full-Stack Software Engineer

The position of software engineer will involve work on a specific project critical to a start-up's needs with opportunities to change projects and teams as the software engineer grows. Engineers are required to be multifaceted, display successful leadership abilities, and be enthusiastic to tackle new and challenging problems.

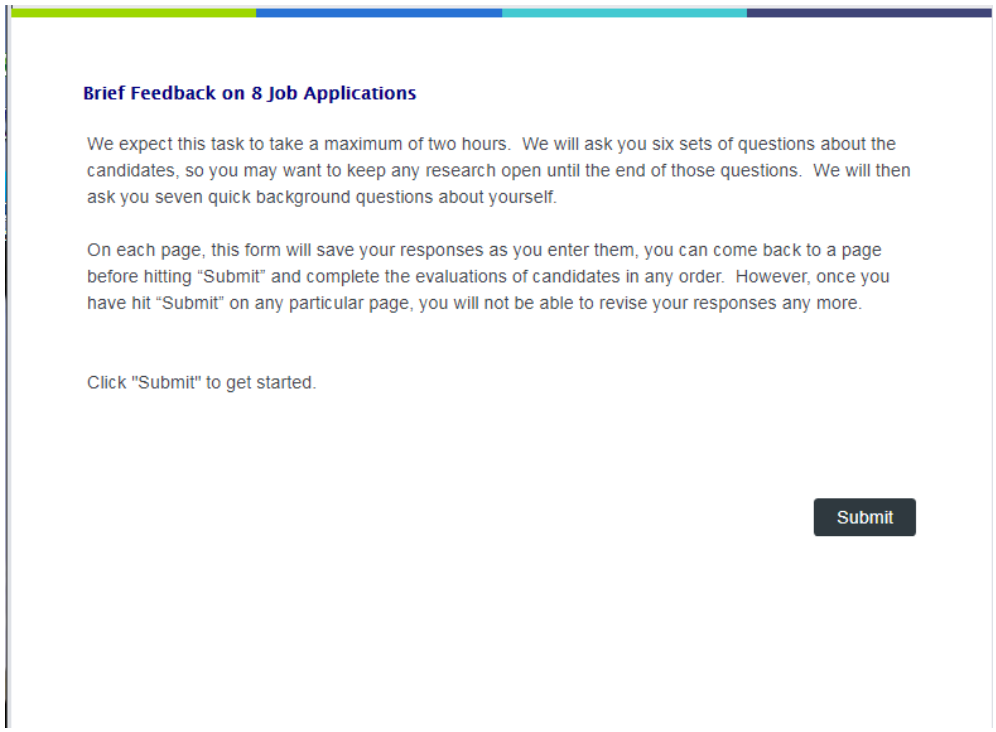
Responsibilities may include:

- Design, develop, test, deploy, maintain, and improve software
- Manage individual project priorities, deliverables, and deadlines
- Collaborate with other specialists in development teams
- Analyze and improve efficiency, scalability, and stability of various system resources

Minimum Qualifications:

- BA or BS degree in Computer Science or related technical field
- Experience with one or more general purpose programming languages including but not limited to: Java, C/C++, C#, Objective C, Python, JavaScript, or Go
- Experience working with two or more from the following: web application development, Unix/Linux environments, mobile application development, distributed and parallel systems, machine learning, information retrieval, natural language processing, networking, developing large software systems, and/or security software development
- Working proficiency and communication skills in verbal and written English

E Structured Evaluation (vary by branding between the Told Experiment and Not Told Experiment Conditions)



Brief Feedback on 8 Job Applications

Please review the candidate information in the packets we sent you. In the table below, tell us:

1. Which candidates do you suggest interviewing?
Then, for each candidate, assume he/she was interviewed and passed.

2. Suggest a take-it-or-leave-it annual base salary offer for each candidate.
Remember that it's worth paying more to make sure that higher quality candidates say yes. But we also are on a budget and do not want to overpay.

Please provide a salary offer even for candidates you do NOT suggest interviewing. Someone else at our company might suggest interviewing these candidates. If they pass, we will use your input for what salary to offer them.

For candidates you do NOT suggest interviewing, please enter the amount you think they should be offered were they to pass an interview - this may be helpful for us in the future. You may enter $0, but only if that is what you truly intend, otherwise please enter a non-zero value.

	Candidate's First name	Candidate's ID #	Interview Yes	Interview No	Your take-it-or-leave-it offer Amount In Dollars	Notes/Comments (Optional)
1			○	○		
2			○	○		
3			○	○		
4			○	○		
5			○	○		
6			○	○		
7			○	○		
8			○	○		

We will interview all candidates you suggest. We may also interview additional candidates. Interviewers will not know who suggested each candidate. We will consider your salary responses only after we finish interviewing and deciding which candidates to pursue.

Submit

Brief Feedback on 8 Job Applications

If you could make a single take it or leave it offer, which candidate would you want to make it to?

Notes/Comments (Optional)

As a reminder, here are your answers to the previous questions.

First Name	ID	Interview	Salary
Andrew	1111	Yes	150,000
Christoper	2222	No	150,000
Emily	3333	Yes	150,000
Jessica	4444	No	150,000
Joshua	5555	Yes	150,000
Samantha	6666	No	150,000
Sarah	7777	Yes	150,000
Tyler	8888	No	150,000

Submit

Brief Feedback on 8 Job Applications

Occasionally, candidates make us a take-it-of-leave-it offer (rather than us making one to them). Our policy is to take these proposals seriously. We will either accept or reject the candidate's offer, and the outcome is then final. We do not make counter offers or consider the candidate's follow-up offers. Candidates know this is our policy.

Please note: We do not feel that candidates who make us an offer are necessarily better or worse than those who wait for the offer.

We need your suggestions about how to respond in these settings. Please select the **maximum** offer we should accept for each candidate in this situation. For your reference, a table of the salary offers you suggested earlier can be found at the bottom of this page.

	Candidate's First name	Candidate's ID #	Maximum take-it-or-leave-it offer we should accept	Notes/Comments (Optional)
1	Andrew	1111		
2	Christoper	2222		
3	Emily	3333		
4	Jessica	4444		
5	Joshua	5555		
6	Samantha	6666		
7	Sarah	7777		
8	Tyler	8888		

As a reminder, here are your answers to some previous questions.

First Name	ID	Interview	Salary
Andrew	1111	Yes	150,000
Christoper	2222	No	150,000
Emily	3333	Yes	150,000
Jessica	4444	No	150,000
Joshua	5555	Yes	150,000
Samantha	6666	No	150,000
Sarah	7777	Yes	150,000
Tyler	8888	No	150,000

While we will interview all candidates you suggested, we may also interview additional candidates. For your responses, assume all candidates are interviewed and have passed. We will consider these responses only after we finish interviewing and deciding which candidates to pursue.

Submit

Brief Feedback on 8 Job Applications

Suppose we were interested in hiring each of the candidates.

Tell us:
A salary so low, they'd be only 5% likely to take it?
A salary they'd be just as happy taking or rejecting (50% likely to accept or reject).
A salary so high, we think they're very likely to take it? (95% likely to say yes)?

We are looking for 3 distinct salary values for each candidate. For example, for an entry level cashier job potential answers might be: there is a 5% chance they would take $7.25, there is a 50% chance they would take $9, and a 95% chance they would take $15. Use your best judgement.

	Candidate's First name	Candidate's ID #	Salary 5% likely to yes	Salary 50% likely to say yes	Salary 95% likely to say yes	Notes/Comments (Optional)
1	Andrew	1111				
2	Christoper	2222				
3	Emily	3333				
4	Jessica	4444				
5	Joshua	5555				
6	Samantha	6666				
7	Sarah	7777				
8	Tyler	8888				

While we will interview all candidates you suggested, we may also interview additional candidates. For your responses, assume all candidates are interviewed and have passed. We will consider these responses only after we finish interviewing and deciding which candidates to pursue.

Submit

Brief Feedback on 8 Job Applications

We're interested in your opinion of each candidate's competing offers - please enter your assessments of their potential competing offers below

	Candidate's First name	Candidate's ID #	Number of Competing Offers (not including current job)	Reason for Competing Offers	Notes/Comments (Optional)
1	Andrew	1111			
2	Christoper	2222			
3	Emily	3333			
4	Jessica	4444			
5	Joshua	5555			
6	Samantha	6666			
7	Sarah	7777			
8	Tyler	8888			

Submit

Brief Feedback on 8 Job Applications

Given your experience with this type of work, do you think that a job candidate like the ones you reviewed here would mis-report his/her most recent salary (even if we could verify past salary later)?

Yes, he/she may mis-report

No, he/she would not mis-report

Submit

Brief Feedback on 8 Job Applications

You said a job candidate might mis-report his/her most recent salary, if a candidate stated his/her current salary was $90,000, what do you think this candidates true most current salary is?

Submit

Brief Feedback on 8 Job Applications

Thank you for your work. We just have a few very short additional questions we would like you to answer

Upwork Profile URL:

How long have you been doing this type of work?

- Less than 3 months
- 3 months to 1 year
- 1 to 3 years
- 3 to 10 years
- Over 10 years

How often do you provide salary input during hiring?

- Never
- Sometimes
- About half the time
- Most of the time
- Always

How did you make judgments on the salary related questions? Please check all that apply.

- Used my previous experience with salaries in this setting
- Looked up salaries on a website like payscale.com, glassdoor.com or others
- Spoke with others who are familiar with salaries for software engineers
- Other

Submit

Brief Feedback on 8 Job Applications

The questions below are optional questions on background for statistical purposes.

I identify my gender as:

- Male
- Female
- Other

I identify my ethnicity as (select as many as apply)

- Asian
- Black/African
- Caucasian
- Hispanic/Latin American
- Native American
- Pacific Islander
- Other

Submit

Brief Feedback on 8 Job Applications

Do you have any suggestions to improvements in how we solicit advice from recruiters? This may include technical issues you had, or bigger picture things like other questions we should ask about evaluating candidates?

Submit

Brief Feedback on 8 Job Applications

We thank you for your time spent taking this survey.
Your response has been recorded.

Note: We also asked one additional question at the very end to those who were told they were part of an experiment which was "What do you believe this academic study was trying to explore?" We did not ask this question of those who were *not* told they were in a study as it would have clued them into being in a study. A total of six respondents (10% of those in the told you're in an experiment condition) mentioned gender, prejudices, or biases.

References

Agan, Amanda, Bo Cowgill, and Laura Gee, "Salary History and Employer Demand: Evidence from a Two-Sided Audit," *NBER Working Paper 29460*, 2022.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American economic review*, 2004, *94* (4), 991–1013.

Charness, Gary, Anya Samek, and Jeroen van de Ven, "What is considered deception in experimental economics?," *Experimental Economics*, 2022, *25* (2), 385–412.

Cowgill, Bo and Patryk Perkowski, "Delegation in Hiring: Evidence from a Two-Sided Audit," *Journal of Political Economy: Microeconomics*, forthcoming.

Gaddis, S Michael, *Audit studies: Behind the scenes with theory, method, and nuance*, Vol. 14, Springer, 2018.

Hamermesh, Daniel, "Are fake resumes ethical for academic research," *Freakonomics Blog*, 2012.

Oyer, Paul and Scott Schaefer, "Personnel Economics: Hiring and Incentives," *Handbook of Labor Economics*, 2011, *4*, 1769–1823.