Inequality as an Incentive

Jeffrey V. Butler
University of California, Merced

Eric Cardella
Texas Tech University *

This version: *January 9, 2020*

Abstract

We study the non-pecuniary incentives of salient inequality. In a laboratory experiment involving over 420 participants we randomly vary piece-rate wage levels for performance on real-effort tasks. Across treatments we exogenously vary whether the task is ability- or effort-intensive as well as the information participants receive about wage and income inequality. We find that salient wage inequality on an initial ability-intensive task increases effort provision on a subsequent task but, at the same time, reduces charitable donations. When the initial task is effort-intensive, we find no significant effects of wage inequality. Salient income inequality has little effect in either environment. We provide evidence using post-experiment survey data about a particular mechanism consistent with the observed behavioral patterns, which relies on contextual variation in a meritocracy narrative affecting Just World Beliefs.

JEL Classification:

Keywords: Inequality; Incentives; Just World Beliefs

*The authors thank George Akerlof, Yan Chen, Dan Friedman, Laura Giuliano, Einav Hart, Karla Hoff and Alessandro Santuoso and Maggie Switek, as well as conference participants at the 2019 Bay Area Behavioral and Experimental Workshop (BABEEW), ERINN 2019, the 2019 Texas Experimental Association Symposium, WEAI 2019 and seminar participants at George Mason University, the University of California, Santa Cruz, the University of Pennsylvania and the University of California, Merced for many helpful comments which have improved the paper.

1 Introduction

Inequality has re-emerged as a hotly debated topic in both the popular press and in academic discourse. While much of the attention has been focused on aggregate trends, such as the share of income going to the top one percent, inequality at a more personal level also warrants attention. One point of departure is to recognize that at the employee level traditional economic theories often prescribe quite a lot of earnings inequality even across identical workers taking identical actions.

As one prominent example, consider the the canonical principal-agent setting in which an employer prefers employees to exert costly effort, but where effort is not perfectly observable. Optimal contracts in this situation tie earnings to outcomes, which are observable but only stochastically related to effort. Consequently, identical employees choosing identical effort levels may end up receiving quite different earnings.[1] Tying earnings to relative performance, e.g., tournament incentives, is another common theoretically justified practice (e.g., Gibbons 1987; Hopkins and Kornienko 2009, 2010; Lazear 2000; Stiglitz 1975) that can obviously generate substantial horizontal earnings inequality.[2]

The traditional economic theories justifying many of these payment structures miss an important point, however, by ignoring the fact that they may naturally reveal earnings inequality to employees.[3] The importance of this omission is highlighted by a growing body of evidence documenting that inequality revelation itself may substantially affect behavior, a pattern which we will refer to as inequality entailing *non-pecuniary* incentives. Several recent studies have carefully documented that exogenously revealing earnings inequality induces backlash, particularly among those disadvantaged by inequality, reducing employee morale, productivity and retention (Breza, Kaur, and Shamdasani 2018; Card, Mas, Moretti, and Saez 2012; Dube, Giuliano, and Leonard 2019; Sseruyange and Bulte 2019). Others have shown that this backlash can be ameliorated through thin justifications such as implied or observed productivity differences, sometimes even on unre-

[1] When employees are not identical, optimal contracts in this setting may entail *ex ante* wage inequality as well in order to account for variation in risk tolerance or effort costs. This may also induce *ex post* earnings inequality through employees' endogenous responses to their (different) optimal wages.

[2] Other theoretical justifications include arguments from efficiency, where inequality permits high-ability workers to credibly signal their ability (Spence 1973). Vertical inequality within a firm across different job types may also be justified theoretically through its effects on beliefs about expected future returns to current effort (e.g., promotions) (Cullen and Perez-Truglia 2018). Yet another justification is due to Rosen (1982, p. 311): "Assigning persons of superior talent to top positions increases productivity by more than the increments of their abilities because greater talent filters through the entire firm by a recursive chain of command technology. These multiplicative effects support enormous rewards for top level management in large organizations." For recent overviews of the literature on incentives in personnel economics and in economics more generally, see Lazear and Oyer (2007); Prendergast (1999) and Lazear (2018).

[3] In the example above, if outcomes are visible to employees as they are to employers and these outcomes vary across employees, the resulting earnings inequality might be easily inferred. A situation fitting this description might be when compensation entails stock options with different publicly known vesting dates (Edmans and Gabaix 2016, see, e.g.,). Tournament incentives also obviously reveal inequality.

lated tasks (Bracha, Gneezy, and Loewenstein 2015; Breza et al. 2018).[4] Other studies document positive reactions to wage inequality (Hart and Piff 2018; Huet-Vaughn 2017).[5]

Because morale, productivity and turnover are crucial components of firm performance, a thorough understanding of both the pecuniary and the non-pecuniary incentives of inequality is crucial to assess the optimality of particular payment structures. In contrast to its pecuniary incentives, however, the non-pecuniary incentives of salient inequality are not well understood. While fairness notions such as Equity Theory (Adams 1963) are widely thought to be the driving force, the varying reactions to inequality revelation documented in the literature suggest that contextual factors play an important mediating role in the *perception* of fairness, which is ultimately what affects behavior.

Utilizing a real-effort laboratory experiment, we examine how participant workers respond after having been exposed to salient wage inequality. In our design, participants first complete an incentivized real-effort task under either a high or low piece-rate wage scheme. Importantly, we vary the context of this task such that participants either perceive the work as being effort-intensive or ability-intensive. After performing this initial task, we vary across treatments in a 2x2 factorial design whether participants are informed of: (i) initial-task wage inequality; or (ii) the (endogenous) earnings inequality within their session on the initial task. Participants then complete a second real-effort task where pay is based on self-reported productivity. Because this task has both a quantity and quality dimension, participants can cheat in two distinct ways: by outright lying about the number of units they produced; or, more subtly, by producing lower quality units. Finally, participants are presented with an opportunity to credibly donate a portion of their earnings to charity. Our design enables us to identify how salient wage inequality impacts future work ethic and other-regarding behavior as well as how these reactions vary with the ability-intensiveness of the work employees perform.

As a preview of our results, we find that when task performance depends primarily on ability rather than effort, revealing wage inequality increases the quantity and quality of output on the subsequent real-effort task among those *disadvantaged* by the inequality. Revealing wage inequality also reduces subsequent other-regarding behavior among the disadvantaged. Among the advantaged workers, revealing wage inequality has little effect on work ethic or on other-regard. Completing the picture, we find no effect of revealing wage inequality when work is effort-intensive, nor do we find any effect of revealing earnings inequality in any context.

[4] In Bracha et al. (2015), wages determined by the experimenters' "evaluation," using unspecified criteria, of a pre-experiment essay drove out wage backlash; in Breza et al. (2018), backlash is lower when heterogeneity in productivity is more pronounced and more easily observed.

[5] Hart and Piff (2018), in an experimental study involving deception, show that revealing earnings inequality increases contributions in a public goods game by those at the median of the earnings distribution. Huet-Vaughn (2017) finds that revealing earnings inequality associated with performance on a tedious real-effort task increases productivity among Amazon MechanicalTurk workers.

The particular combination of positive and negative non-pecuniary incentives we document is seemingly inconsistent with a straightforward Equity Theory explanation, as it suggests wage inequality is simultaneously perceived as inequitable (reducing other-regard) and as equitable (increasing effort), *by the disadvantaged*. The pattern is consistent, however, with a different underlying mechanism. Heuristically, a strong work ethic coupled with an emphasis on self-reliance is reminiscent of the "American/*laissez faire*" equilibrium detailed in Benabou and Tirole (2006), who attribute the pattern to strong Just World Beliefs (Lerner 2013, 1965), hereafter JWBs. A well documented manifestation of motivated beliefs (Bénabou and Tirole 2016), JWBs refer to the universal tendency to believe that people get what they deserve. Building on Butler (2016), we hypothesize that wage inequality in an ability-intensive context gives rise to a meritocracy narrative which permits subconscious justification of wage inequality as fair, so that revealing wage inequality actually strengthens JWBs and has predictable effects on subsequent behavior.[6] We investigate this mechanism directly using attitudinal data collected on a post-experiment survey and find patterns largely consistent with our hypothesis.

Our study makes several contributions. Most directly, we contribute to the literature on the non-pecuniary incentives of salient inequality (Abeler, Altmann, Kube, and Wibral 2010; Akerlof and Yellen 1990; Angelova, Güth, and Kocher 2012; Bartling and Von Siemens 2011; Bolton and Werner 2016; Bracha et al. 2015; Breza et al. 2018; Butler 2016; Card et al. 2012; Charness, Cobo-Reyes, Lacomba, Lagos, and Pérez 2016; Charness and Kuhn 2007; Clark, Masclet, and Villeval 2010; Cohn, Fehr, Herrmann, and Schneider 2014; Dube et al. 2019; Gächter and Thöni 2010; Gill, Prowse, and Vlassopoulos 2013; Godechot and Senik 2015; Greiner, Ockenfels, and Werner 2011; Gross, Guo, and Charness 2015; Hennig-Schmidt, Sadrieh, and Rockenbach 2010; Nosenzo 2013; Sseruyange and Bulte 2019). In this growing literature, we are among the first to consider how the behavioral consequences of inequality may depend on the content of work, i.e., whether work is ability-intensive or effort-intensive.[7] This is an important distinction, as personal and

[6] In a series of on-line and laboratory experiments using students at UC Berkeley as well as Amazon MechanicalTurk workers, Butler (2016) shows that ability-intensive environments cause people to distort their beliefs about the basis of inequality, perceiving even *one's own* disadvantage to be deserved. To explain these beliefs patterns, Butler conjectured without direct evidence, that ability-intensive environments bring to mind a narrative of meritocracy in which inequality is not only consistent with fairness, but actually *required by* fairness – think of college grading scales or Olympic awards ceremonies, quintessentially meritocratic environments where *equal* rewards would be unfair. This narrative allows individuals to avoid the cognitive dissonance that would arise from simultaneously holding strong JWBs and observing unjustifiable inequality by subconsciously distorting their beliefs about the basis of inequality. We take this logic one step further and conjecture that objectively unjustifiable inequality may actually be transformed into additional "evidence" in support of a just world, strengthening JWBs and producing subsequent behavior reflecting these stronger JWBs.

[7] The only other papers we are aware of are Butler (2016) and Hart and Piff (2018). Hart and Piff (2018) tell subjects that experimental earnings inequality is based on either: i) performance on a preferences survey, and hence ostensibly random; or ii) performance on a "knowledge" quiz, which in our terminology would be ostensibly ability-intensive. Their procedure involves deception. Still, they find (as we do) positive non-pecuniary incentives of salient

societal economic success increasingly depends on skill- or ability-intensive employment sectors. If the consequences of inequality differ qualitatively and predictably across this dimension, from the firm's perspective the optimal degree of inequality revealed naturally through pay structures or concealed through formal or informal pay secrecy policies may also vary across this dimension. Moreover, much of the literature in this domain focuses on how ex-ante revelation of pay inequality impacts effort provision in the designated task, whereas we focus on how ex-post revelation of pay inequality impacts behavior in subsequent tasks.[8]

Secondly, we contribute to the nascent literature on motivated beliefs in economics. Researchers have long understood that in order to avoid cognitive dissonance (Festinger 1957) individuals may subconsciously color their beliefs and this process may alter economic incentives (Akerlof and Dickens 1982; Bénabou and Tirole 2016). However, evidence on the determinants and consequences of motivated beliefs is scant. A primary exception is Di Tella, Galiant, and Schargrodsky (2007), who exploit a natural experiment to show that randomly assigned land wealth shifts attitudes towards "market beliefs," such as the belief that individuals can succeed through their own efforts. We contribute to this literature by outlining a novel mechanism through which the economic environment can interact with subconscious beliefs-formation processes and readily accessible culturally determined narratives (Collier 2016; Shiller 2017) to affect motivated beliefs and, in turn, behavior. We provide evidence on both the mechanism and the behavior it implies.

We also contribute methodologically. We construct all of our tasks to be ecologically valid, i.e., appropriate for, and familiar to, our student population: multiple-choice quizzes, Scantron bubbling and donating to well-known charities.[9] The purpose of this design feature is to provide properly incentivized evidence on pro- and anti-social behavior that complements the large body of experimental literature typically featuring decontextualized and unfamiliar tasks such as (abstract) dictator games or dice-rolling for money (Abeler, Nosenzo, and Raymond 2016; Charness and Rabin 2002; Dufwenberg and Dufwenberg 2018; Fischbacher and Föllmi-Heusi 2013; Gneezy, Kajackaite, and Sobel 2018; Mazar, Amir, and Ariely 2008). By comparing behavior in our experiment with behavior from conceptually related experiments, we generate evidence on the robustness of previous

inequality. Considered together, our papers provide evidence on the robustness of our findings both to the exact set of tasks involved and to the use of deception.

[8]One recent exception is Sseruyange and Bulte (2019), who conduct a field experiment in Uganda and find that piece-rate wage inequality on an initial bean-sorting task reduces effort provision in a subsequent non-incentivized bean-sorting task.

[9]Throughout the course of a typical day, students may proofread their own or others' work, take an in-class quiz, possibly involving bubbling Scantrons, or donate to well-known charities. Classroom quizzes often have an ability component, which we mimic by using questions from Raven's Advanced Progressive Matrices – a commonly used measure of general intelligence. Highlighting this connection, economist Greg Mankiw famously quipped "... math courses are one long IQ test" (Mankiw 2006). Many participants may have even taken a Raven's-like test, as many similar (unofficial) on-line tests purport to measure intelligence.

findings to arguably more realistic tasks, even if they still take place in a laboratory setting.

Before moving on, it is worth mentioning that the potential implications of our findings are large, subject to the typical caveat with experimental studies about external validity. First and foremost, our findings may provide one explanation for the concomitant growth in the skills-intensive sectors of modern economies and wage inequality. Firms in skills-intensive sectors that implement more transparently unequal pay structures may actually be at an advantage, suffering less from the inefficiencies of moral hazard, compared to firms implementing less visible wage inequality in the same sectors. Our findings may also have implications for particular aspects of corporate culture. Within firms, employees' *perceptions* of their work may be influenced by features under the firm's control, such as job descriptions, titles or even the firm's mission statement. Firms successfully conveying that performance depends heavily on skill or ability may again be able to ameliorate inefficiencies associated with imperfect observability of effort, yielding an additional channel through which corporate culture may affect firm performance.

Lastly, our findings can contribute to the recent discussions regarding wage transparency within organizations. Wage transparency within organizations remains a hot button issue in the popular press and has even been the target of recent policy reform (e.g., mandatory pay-ratio reporting under Dodd-Frank).[10] While some firms have embraced the idea of making salary information among its employees publicly available, the view on the merits of wage transparency remains polarized. Anecdotal evidence often supports the claim that pay transparency is a good thing, as it can result in higher productivity. However, some experts claim that wage transparency can result in lower productivity especially among the relatively underpaid subset of employees (see recent articles by Cooney (2018) and Conlin (2018)). Not surprisingly, there has been little formal empirical evidence on the causal impacts of wage transparency within firms on employee behaviors, given data limitations and identification challenges (c.f., Cullen and Perez-Truglia 2018). Importantly, the ramifications of policies aimed at increasing wage transparency hinge crucially on how employees might respond upon observing a (presumably) unequal pay distribution. Our results suggest the response to wage transparency might hinge crucially on the type of work employees are engaging in. Specifically, wage transparency among firms with (perceived) ability-intensive jobs (e.g., tech sector) can, indeed, be beneficial by boosting productivity and reducing shirking, especially among the subset of relatively low paid employees. As such, our study can contribute more broadly to the ongoing discussions of the merits of wage transparency.

The remainder of the paper proceeds as follows. First, we present our experimental design in

[10]By way of example, several recent articles on the merits of pay transparency have appeared across a variety of reputable business publication outlets including (but not limited too): Time (Cooney 2018), Fortune (Fisher 2015), The Wall Street Journal (Shellenbarger 2016), and Business News Daily (Conlin 2018).

detail before formally stating several behavioral hypotheses in the subsequent section. In section 4 we provide empirical results on these hypotheses. In the penultimate section, we investigate the underlying mechanism relating inequality to behavior using a direct measure of JWBs collected on the post-experiment survey. In the final section we summarize and conclude.

2 Experimental Design and Procedures

We conducted a laboratory experiment involving a real-effort work task followed by subsequent opportunities to cheat and to donate to charity. All experimental sessions were conducted at the Rawls College of Business at Texas Tech University in the Spring and Fall of 2018 and were programmed in oTree (Chen, Schonger, and Wickens 2016). All participants were recruited through a college maintained subject pool. In total, we conducted 31 experimental session in which 423 individuals participated (44% were female; average age was about 21). Our experiment consisted of eight treatments and two piece-rate pay schemes implemented using a between-subjects design. Sessions lasted approximately one hour and average compensation was $15.01. Participant instructions for all parts of the experiment are provided in an Instructions Appendix.

There were three decision-making phases in our experiment plus an information revelation phase. Phases 1 and 2 involved real-effort tasks. In Phase 3, participants were provided an opportunity to donate to charity. Phase 1.5, which occurred after the Phase 1 task was completed but before the Phase 2 task began, was a purely informational phase. Specifically, in Phase 1.5 all participants learned of their own Phase 1 performance and experimental income while some participants, depending on treatment, received additional information about relative wages and incomes. We describe each of these phases in detail below and summarize the timeline of the experiment in Figure 1.

2.1 Phase 1: Initial Real-Effort Task

To induce wage and income inequality, participants first completed a real-effort task for a specific piece-rate wage. As part of the experimental design, we consider two distinct tasks, which differ in the extent to which performance credibly depends on skill or ability versus effort alone.

The first task, which we refer to as the ability-intensive (AI) task, consists of 48 Raven's Advanced Progressive Matrices. The second, effort-intensive (EI) task, consisted of 48 sentences in which participants counted typographical errors. Each participant performed only one of these tasks and did not know of the existence of the task they were not assigned. Random assignment was accomplished across sessions, i.e., each experimental session featured only one of the tasks.

The Raven's matrices are multiple choice questions in which the respondent selects a picture

that best completes a given pattern from among eight alternatives. Raven's matrices are designed to measure the test taker's reasoning ability, which is considered to be an important component of general intelligence. We convey this interpretation to our participants in the instructions preceding the AI task. We therefore consider it a plausible assumption that participants perceive the task as being ability-intensive.

We constructed the EI task to be as parallel as possible to the AI task. The EI task consists of a sequence of 48 multiple choice questions, each with eight possible responses. Each of these questions asked the respondent to count the number of typographical errors that appeared in a particular sentence, with valid responses ranging from "none" to "7 or more." We deliberately constructed the sentences to contain only typos that were easily recognizable: duplicated words, missing words, or numbers that replaced letters. Our reasoning was that spotting these kinds of errors requires little (reading) ability but a reasonable amount of effort. Consequently, our maintained assumption is that performance on the EI task depends primarily on effort and that, moreover, participants perceived this to be the case.

We randomized tasks across sessions because, despite our efforts to make them as parallel as possible, they appear quite different on a computer screen. The AI task involves a grid of images, while the EI task features primarily text. Therefore, across-session randomization was intended to minimize the possibility of participants realizing there were different tasks and hence different treatments.

To induce explicit wage inequality we implemented two different piece-rate pay schemes:

Low Wage: \$0.20 for each correct response; \$0.00 for each incorrect response.

High Wage: \$0.30 for each correct response; \$0.10 for each incorrect response.

Randomization into pay scheme occurred within-session, with each participant being equally likely to be assigned each piece-rate wage. During Phase 1, participants only knew about their own piece-rate wage and did not know of the existence of the wage they were not assigned. This eliminates any scope for relative pay concerns to affect Phase 1 behavior. Additionally, by keeping marginal (monetary) incentives for correctly answering a question constant across the two pay schemes (\$0.20), we mitigate the possibility of differences in performance due solely to monetary incentives (see Butler 2016).

Participants had exactly 10 minutes to answer as many of the 48 questions comprising the Phase 1 task as they could. A timer appeared across the top of the computer screen during the task. Prior to beginning the task, each participant was provided a description of the task they were about to complete and informed of their own piece-rate wage as well as the total number of

questions possible. They were also informed that the task would be timed and that there would be a 10-minute time limit.

2.2 Phase 1.5: Information Revelation

After completing the 10-minute Phase 1 task, participants were informed of their own performance and their own income from the task.[11] In addition to these basic pieces of information, we exogenously varied the information participants received about the existence of alternative pay schemes and the distribution of income in their session.

Within each session each participant was equally likely be informed of the existence of the two possible piece-rate wages or not. Conditional on being informed of the multiple pay schemes, participants were told they were uniformly randomly assigned. We think of revealing the existence of different piece-rate wages as implementing visible wage inequality, which we denote by the acronym **VW**. Concealing this information implements invisible wage inequality, denoted **IW**.

The other domain of information we manipulated was relative income. Specifically, in the visible income inequality (**VI**) condition participants were provided with a binned relative frequency chart describing the distribution of experimental income in their specific session. In the invisible income inequality (**II**) condition participants received no such information. Through this manipulation we reveal comprehensive information about the income inequality within a session, allowing participants to locate themselves within the session's income distribution. Because the VI condition involves a large chart appearing on participants' screens, to minimize the possibility of participants discovering there were multiple treatments, we varied this factor only across sessions. That is to say, all participants in a particular session were assigned to either the VI condition or the II condition.

Because information about wage and income inequality was provided to participants only after they completed the Phase 1 task, effort provision in Phase 1 cannot be affected by these manipulations. This feature enables us to cleanly identify how knowledge of wage and income inequality impact subsequent behavior, including anti-social behavior (lying and shirking) and pro-social behavior (charitable donations). Because the experiment is identical for all participants after Phase 1.5, we can summarize our experimental design as consisting of a 2 (wage inequality visibility) x 2 (income inequality visibility) x 2 (task type) full factorial design with eight distinct treatments (Table 1). For ease of exposition, we refer to each treatment using the format [VW/IW]-[VI/II]-

[11] We inserted filler questions between Phase 1 and Phase 1.5 which we do not analyze, but simply control for. Participants received one of three possible sets of questions, with the sets being randomly assigned at the individual level. Specifically, they were asked to state their political party or gender and then describe briefly an event which made them feel connected to their gender or party. The third possible set of questions asked them about the primary medium through which they receive television programming (e.g., cable or the internet), and to briefly describe their reason for selecting that medium.

Table 1: *Summary of Treatments*

Treatment Name	Task	Visible Wage Inequality	Visible Income Inequality	N
IW-II-EI	Proofreading	No	No	54
IW-II-AI	Raven's Matrices	No	No	56
VW-II-EI	Proofreading	Yes	No	53
VW-II-AI	Raven's Matrices	Yes	No	58
IW-VI-EI	Proofreading	No	Yes	52
IW-VI-AI	Raven's Matrices	No	Yes	49
VW-VI-EI	Proofreading	Yes	Yes	53
VW-VI-AI	Raven's Matrices	Yes	Yes	48

Notes: **[1]** Raven's matrices refer to Raven's Advanced Progessive matrices, commonly thought to be a culture- and language-free measure of general intelligence.measure. **[2]** For the proofreading task, we randomly inserted simple forms of typogaphical errors into sentences and participants had to count the number of typos appearing in each sentence. **[3]** Task format was constructed to be as similar as possible; each item of each task involved selecting the correct answer from among eight choices, and was approximately the same size and format on participants' computer screens.

[AI/EI]. As an example, VW-VI-AI refers to the treatment with visible wage inequality, visible income inequality and an ability-intensive Phase 1 task.

2.3 Phase 2: Scantron Task

The Phase 2 task consisted of a second real-effort task where participants could earn additional compensation. In the spirit of ecological validity, participants were asked to prepare multiple versions of a Scantron answer key, corresponding to different versions of a statistics exam. Participants were provided with a mock exam closely resembling an actual exam administered in a statistics class taught by one of the authors. Participants were also provided with a master answer key for the original (mock) exam, a sheet listing the fifteen desired answer key versions (labeled generically as Version A – Version O), and fifteen Scantrons with which to create these versions.[12] The Phase 2 task was again timed: participants had 15 minutes to complete as many answer key versions from the list as they could. They were instructed that at the end of the 15 minutes they would report how many Scantrons they completed and would be paid, based on this self-report, $1 for every completed Scantron.

After the 15 minutes had elapsed, participants placed all of their Scantrons, completed or not,

[12]To facilitate statistical identification, all participants received the same exam copy, master answer key and version variations. The versions were created by taking random permutations of question orderings on the master key. To enhance ecological validity, through opacity in our instructions we created an environment where participants likely perceived the task as regular economically valuable work (Falk and Ichino 2006). At the same time, to avoid deception participants were not told anything about the intended use of the answer keys they would be preparing. They were simply informed that they were to make answer keys for different versions of the exam based on the exam master key provided.

in a box at the back of the room. The box was not monitored by the (lone) experimenter. After returning to their carrels, each participant filled out an anonymized payment slip. It was on this slip that production was self-reported and upon which payment from Phase 2 was based. Because participants were explicitly informed we would pay them according to their self-reports, there were opportunities to engage in cheating. By inflating their self-reported number of completed Scantrons, participants could cheat in a very deliberate manner: lying for financial gain. A more subtle form of cheating was also possible, which we term "shirking." By completing versions less accurately, which presumably requires less time and effort, participants could earn a given amount of money with less effort or complete more Scantrons in the allotted time than if they filled out versions accurately.

Importantly, it was obvious that for logistical reasons that both forms of cheating – lying or shirking – were imperceptible by the experimenter during the session. Great pains were taken to ensure, and to convey implicitly to participants, that cheating could only be detected after sessions were completed and could never be attributed to a particular name, but rather only to the participant's carrel number, and that we could never match names to carrel numbers.[13] We argue that this provides essentially the same level of anonymity as the more frequently used dice-rolling paradigm (Abeler et al. 2016; Dufwenberg and Dufwenberg 2018; Fischbacher and Föllmi-Heusi 2013; Gneezy et al. 2018; Mazar et al. 2008) while still being familiar and appropriate for our subject pool and, importantly, allowing for precise identification of cheating.[14]

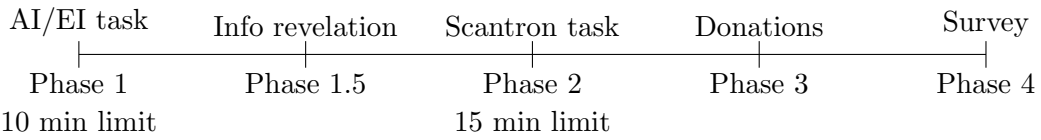
2.4 Phase 3: Charitable Giving Opportunity

In Phase 3, participants first received their cumulative earnings from the Phase 1 and Phase 2 tasks in cash in an envelope. They were then given the opportunity to anonymously donate any amount they wished to two specific charities by leaving some of their earnings in the envelope along with a slip specifying how to divide the cash between the charities. The two charitable giving options were: South Plains Food Bank, a local charity; or the American Red Cross, a national charity. We

[13] With only one experimenter in the room, it would have been logistically impossible at the time to examine self-reports for accuracy, which should have been apparent to participants. In particular, the experimenter did not approach the back of the room, where the completed Scantrons were delivered, while the participants were still in the room. Moreover, we paid using cash in envelopes marked with code numbers which could not be linked to participant names. We also did not have participants sign receipts.

[14] Our logic is as follows. In the typical dice-rolling experiment, lies can be identified at some aggregate level – say, the session level. Experimenters typically know who participated in which sessions, so that an experimenter conducting a dice-rolling experiment would know that John Smith certainly participated in a session where lying occurred. Consequently, the experimenter would know that with some probability John Smith told a lie. In our setting, we would know that somebody seated at carrel 33 certainly told a lie and that, with some probability, the person seated at carrel 33 was John Smith. Therefore, we would know that with some probability John Smith lied. Overall, therefore, both designs feature probabilistic personal identification of lying so that while there may be a difference in the probability with which liars are identifiable, this difference is a matter of degree rather than of kind.

Figure 1: *Timeline*



conveyed to participants that this was a credible donation and that they would receive by email at the conclusion of the study a link where they could view a receipt for the total amount of money donated to each charity (across all participants). Participants were asked to leave their envelope at their carrel when they left the experiment. As an added step to ensure anonymity, even participants who chose to donate nothing were asked to place a completed donation slip in their envelope and leave the envelope at their carrel. These procedures should ensure that donations were anonymous to the experimenters as well as to participants' peers.

After participants had completed filling their charitable donation envelopes, we conducted a time-preferences elicitation task. We do not analyze the task here. Instead it should be viewed as a distractor task, placing temporal and psychological distance between the potentially (un)ethical decisions in Phases 2 and 3 and the post-experiment survey.

2.5 Phase 4: Post-Experiment Survey

Before leaving their sessions, participants were asked to fill out a survey. The survey was anonymous and conducted using Qualtrics. We chose to use Qualtrics instead of programming the survey in oTree, as we did for the Phase 1 task, to introduce further mental separation between the various phases of the experiment. On the post-experiment survey we gathered general demographic data, various psychological measures and self-reported beliefs and attitudes on a variety of relevant topics. Importantly, interspersed among the various questions on the post-experiment survey was a direct measure of JWBs. Specifically, participants were asked to state how much they agreed with the statement "People generally get what they deserve." Responses were collected on a 7-point Likert scale ranging from 1 ="totally disagree" to 7 ="totally agree."

Well after each session was completed, we matched outcome measures from all phases of the experiment. Matching was accomplished by carrel number, which was automatically recorded for the computerized portions of the experiment (Phase 1 and the post-experiment survey). For the other phases of the experiment, participants noted their carrel numbers on materials they submitted. It is worth noting that it is not possible for the experimenters to match participants' names to carrel numbers or even payment amounts as, e.g., no receipts were collected.

3 Hypotheses

Having described the experiment in detail, we are now in a position to formally state our hypotheses. We begin by providing more detail on our conjectured mechanism relating context to behavior. We then state formally our hypotheses about how our experimental factors will affect behavior, followed by separate hypotheses about how attitudinal measures will be affected.

3.1 Behavioral hypotheses

Our first hypothesis constitutes a simple specification check as well as a confirmation of standard incentive theory. In the absence of non-pecuniary inequality incentives, performance should theoretically depend primarily on marginal monetary incentives. Because marginal monetary incentives are identical in our two pay schemes, \$0.20 for each additional correct answer, we expect no significant variation in Phase 1 task performance across pay schemes. Since, furthermore, neither wage nor income inequality was disclosed until after the Phase 1 task was completed, neither type of inequality should affect Phase 1 task performance.

Hypothesis 1: *For the AI and EI Phase 1 tasks considered separately, the average number of correct responses will not vary significantly across any of our experimental factors, including piece-rate wage.*

Our next three hypotheses consider the effects of disclosing wage inequality. Being completely randomly assigned and hence objectively unjustifiable, we expect this form of inequality – particularly among those disadvantaged by it – to be the most likely to engender cognitive dissonance absent countervailing cognitive processes. Among this group it is therefore the most likely that our hypothesized cognitive mechanism is induced, yielding our best chance for pronounced and predictable behavioral effects. We conjecture that these behavioral effects will be consistent with a strengthening of JWBs, reflecting a strong work ethic and an emphasis on self-reliance.

In our experimental design, work ethic may be reflected in two ways. The first way is to increase the number of Scantrons produced in Phase 2. Since the number of Scantrons produced does not directly determine earnings and is essentially unobservable, the quantity produced should rely on intrinsic effort incentives. The second way strong JWBs may be reflected in our experiment is by reducing shirking or, equivalently, increasing the quality of the Scantrons produced. Again, because quality entails essentially unobservable effort and has an even more tenuous relationship with earnings determination than the quantity of Scantrons produced, quality should also reflect intrinsic effort incentives. This leads to our second hypothesis.

Hypothesis 2: *Disclosing wage inequality when the Phase 1 task is ability-intensive will increase the quantity and quality of Scantrons produced in Phase 2. These positive non-pecuniary incentive effects will be strongest among those assigned to the low-wage condition.*

Strong JWBs also induce an emphasis on self-reliance, which may reduce other-regard. Heuristically, if (other) people generally get what they deserve, one's own incentives to take actions which affect others' outcomes are reduced. In our experiment, other-regard is reflected in Phase 3's charitable donations opportunity, implying our third hypothesis.

Hypothesis 3: *Disclosing wage inequality when the Phase 1 task is ability-intensive will reduce the proportion of experimental earnings donated to charity in Phase 3. This reduction will be strongest among those assigned to the low-wage condition.*

Our next hypothesis is more speculative, as previous research is silent on this aspect. It could be the case that strengthened JWBs discourage lying. Intuitively, if people get what they deserve and lying deserves to be punished, then JWBs provide an expected disincentive for lying.

Hypothesis 4: *Disclosing wage inequality when the Phase 1 task is ability-intensive will reduce the propensity to lie about Phase 2 Scantron production. This reduction will be most pronounced among those assigned to the low-wage condition.*

For our final behavioral hypothesis, we consider the effects of revealing income inequality. The existing literature on salient inequality largely ignores the distinction between wages and income. Conceptually, wage inequality is an *ex ante* measure of (mis)treatment reflecting unequal opportunities, while income inequality incorporates both this *ex-ante* (mis)treatment and individuals' endogenous responses to it. In particular, because wage inequality may be (partially) overcome through individuals' actions, some of the blame for income inequality may be subjectively shifted to the individual. The end result may be that income inequality is easier to objectively justify and consequently less threatening to JWBs so that it is less likely to give rise to the beliefs distortions associated with wage inequality. This should be true whether income inequality is revealed by itself or in conjunction with wage inequality.

Hypothesis 5: *Revealing income inequality in isolation will have no effect on behavior; revealing income inequality in addition to wage inequality will generally weaken the effects of wage inequality on behavior, i.e., produce less of a reduction in shirking and lying and less of a reduction in charitable donations than wage inequality alone.*

4 Results

4.1 Descriptive statistics

To provide an overview of the data, in Table 2 we report raw means of our primary outcome variables by treatment and pay scheme. For Phase 1, we report (experimental) income and the number of correct responses on the assigned task (score).[15] For Phase 2, we report the quantity and quality of Scantrons produced. Our quantity measure is the number of Scantrons submitted, irrespective of quality, which hypothetically takes values from 0 to 15. For our quality measure, for each participant we compute the average proportion of correctly bubbled items (out of 20) across all Scantrons he or she submitted.[16] This can be interpreted as the inverse of shirking – more (less) effort should increase (decrease) quality – and takes values from 0 to 1. Our measure of lying is an indicator variable taking the value of one whenever the participant's self-reported number of Scantrons completed does not match the actual number he or she submitted.[17] For charitable donations, for simplicity we report only the proportion of a participant's total experimental income that he or she donated to either charity.

From the raw means we can glean a few patterns. First of all, we were successful in generating substantial income inequality on the Phase 1 task across pay schemes. The income difference across pay schemes is highly significant within each of the eight treatments ($p < 0.001$ always; two-tailed t-test). Pooling all treatments, participants assigned the Low-Wage scheme earned \$4.90 from this task, while High-Wage participants earned \$8.97, or 83% more ($p < 0.001$; two-tailed t-test). These numbers suggest the Phase 1 earnings distribution was bi-modal, an impression that is confirmed in Figure 2 where we present kernel density estimates of Phase 1 earnings by Phase 1 task. The substantial income heterogeneity both within and across wage levels should aid our ability to detect behavioral reactions to the revelation of income inequality.

[15] To provide evidence that randomization yielded comparable samples across treatments, we conducted simple balance checks using the limited set of demographics we collected on the post-experiment survey. These demographics included age and gender as well as self-reported categorical measures of family income and GPA. For each of these variables separately we conducted a Chi-squared test with the null hypothesis of independence across treatments. Only the test associated with GPA ($p = 0.084$) was significant at near-conventional levels. Overall, randomization into treatments appears to have been generally successful. As a precaution, however, in our formal econometric estimates below we control for available demographics.

[16] More precisely, we restrict attention to Scantrons on which a participant bubbled at least one item, compute the proportion of correct answers for each of these Scantrons, and then take the average of these proportions for each participant.

[17] As with the more prevalent dice-in-cups procedure (Fischbacher and Föllmi-Heusi 2013; Mazar et al. 2008), we interpret this mismatch as a proxy for lying. Obviously some misreports could also be due to mistakes. Strengthening our case, the vast majority of misreports are over-reports, i.e., resulting in higher earnings rather than lower earnings. Similar to much of the literature, we find that participants who lie do not lie as much as they can. Lies in our experiment typically take the form of inflating one's report by one unit, rather than reporting the maximum possible of 15 — which few achieved or reported. Because of this pattern we lose little by using an indicator variable rather than the difference between self-reports and actual production as a measure of lying.

Secondly, considering Phase 1 task performance (score) we see that although average performance is generally higher in the effort-intensive task than in the ability-intensive task,[18] within each task there is little variation in performance by treatment. This appearance is confirmed by separate Chi-squared tests conducted on the pooled EI-session data ($p = 0.224$) and on the pooled AI-session data ($p = 0.568$). Testing for a relationship between pay scheme and Phase 1 task performance yields similar conclusions.[19] Consequently, our data pass an important specification check which formed the basis of our first hypothesis. On the whole, our data support the notion that marginal monetary incentives are decisive determinants of Phase 1 task performance and that participants did not anticipate the experimental manipulations to follow. This set of null findings permits the attribution of differences in subsequent outcome variables to the revelation of inequality.

Result 1: *Hypothesis 1 is supported. There was no significant variation in Phase 1 task performance by treatment or by piece-rate wage scheme.*

Overall, there are very few significant patterns which can be gleaned from the raw data alone. The average number of Scantrons produced in Phase 2 ranges narrowly from about seven to about eight. Pooling across all treatments, High-Wage participants produced marginally significantly more Scantrons than their Low-Wage counterparts (7.70 vs. 7.36; $p = 0.103$, two-sided t-test). Pooling over pay schemes, marginally significantly fewer Scantrons were produced following the AI Phase 1 task than following the EI task (7.36 in AI vs. 7.71 in EI; $p = 0.084$, two-sided t-test). A Chi-squared test on the quality of Scantrons submitted, i.e., the opposite of "shirking," reveals little variation across experimental factors ($p = 0.773$), as do similar tests for lying and for the proportion of experimental income donated to charity (lying: $p = 0.704$; donation proportion: $p = 0.522$).

Putting aside statistical significance for a moment, the raw means provide some preliminary indications that behavior following an ability-intensive Phase 1 task is qualitatively different. While it is typically Low-Wage participants that donate a larger proportion of their earnings to charity, this pattern is reversed when wage inequality is revealed in an ability-intensive setting (VW-IW-AI). With this in mind, we turn next to more formal econometric estimates.

4.2 Econometric estimates

While descriptive statistics are useful in providing an overview of the behavioral patterns in our data, the large number of treatments involved complicates our ability to discern anything more

[18] This would be consistent with, e.g., individuals exerting full effort in both tasks but the ability component holding back performance in the ability-intensive task.

[19] Pooling all AI-session, a Chi-squared test for a relationship between pay scheme and score is non-significant ($p = 0.209$). The same procedure for EI-session observations yields a p-value of 0.495. Moreover, testing by pay scheme within each of the eight treatments separately yields only one significant difference. The approximately four-item difference in the VW-VI-EI treatment by pay scheme yields $p = 0.014$ using a two-sided t-test.

detailed using just the raw averages. In particular, while a handful of individually significant relationships are present in the data, singling them out runs the risk of identifying false positives. More formal econometric estimates will lessen this concern and, at the same time, allow us to control for potentially important confounding factors such as unintended random variation in demographics or other observables across treatments.

For ease of exposition, because our experiment involves a large number experimental factors and potential interactions, we split the sample by Phase 1 task type and estimate our econometric models on each of the resulting subsamples separately. Since randomization into task type occurred only across sessions, this is a relatively clean split. No individual performing the AI task could have possibly affected any individual assigned to the EI task as, for instance, the relative performance information for AI-sessions never included any EI-session component. This route has the advantage of letting us report all of our experimental controls and their interactions, facilitating the interpretation of the estimated magnitudes.

In Table 3 we restrict observations to sessions featuring the AI Phase 1 task. Each column presents a separate OLS estimate whose dependent variable is listed in the column heading. As explanatory variables, we include indicator variables for high wage (HW), visible wage inequality (VW), visible income inequality (VI) as well as interactions among these variables. To account for the possibility of negative reactions related to one's own performance itself, such as from falling short of a subjective reference level of earnings or task performance, we control for earnings from the Phase 1 task (Phase 1 income) where possible. We also, where possible, control for the actual number of Scantrons individuals produced in Phase 2 (Phase 2 quantity). Finally, each column includes demographic controls, which we omit for readability.[20]

The coefficients associated with VW capture the main effect of revealing (only) wage inequality on Low-Wage participants' behavior, while HWxVW captures the additional effect on their High-Wage counterparts. The data suggest the main effects of revealing wage inequality on both the quantity and quality of Scantrons produced in Phase 2 among Low-Wage participants is positive, highly significant and substantial in magnitude. The one-unit increase among Low-Wage participants in the quantity of Scantrons produced constitutes 48.7% of one standard deviation of AI-session quantity, while the 7 percentage point increase in Scantron quality represents 46.6% of a standard deviation of AI-session quality. As a back of the envelope calculation, taking the point estimate (0.01) of the effect of Phase 1 earnings on Phase 2 quality provision at face value, the effect of revealing wage inequality on subsequent quality provision is equivalent to a $7 increase in earnings, which is 120% of the AI-session mean of Phase 1 earnings.

[20]These demographic controls are almost always non-significant. Exceptions to this rule are noted either in the text or in the table footnotes.

Considering next High-Wage participants, the coefficients on HWxVW exhibit opposite signs but similar magnitudes as compared to the coefficients on VW for both quality and quantity provision. This implies that revealing wage inequality by itself has little if any effect on High-Wage participants. Putting both of these observations together, we conclude that on balance the data are consistent with Hypothesis 2.

Result 2: *Hypothesis 2 finds support. Disclosing wage inequality when the Phase 1 task is ability-intensive is associated with a significant increase in both the quantity and quality of Scantrons produced in Phase 2 among Low-Wage participants and little effect on High-Wage participants.*

Turning next to other-regarding behavior, in the last column the coefficient on VW is negative, highly significant and substantial in magnitude. The implied 12 percentage point reduction in the proportion of experimental earnings donated to charity by Low-Wage participants constitutes 42.5% of one standard deviation of AI-session donation proportions and 56.6% of the unconditional AI-session mean. Considering the differential impact on High-Wage participants, the coefficient on HWxVW is opposite in sign, of about the same magnitude as the coefficient on VW, but non-significant. As with Scantron quantity and quality above, taken together these patterns suggest that the effects of revealing wage inequality on subsequent behavior are concentrated primarily upon Low-Wage participants.

Result 3: *Hypothesis 3 finds support. Disclosing wage inequality associated with an ability-intensive Phase 1 task significantly and substantially reduces charitable donations among Low-Wage participants. The estimated magnitude of this reduction is smaller among High-Wage participants.*

Finally, we turn to the more obviously anti-social behavioral opportunity afforded participants — the opportunity to lie for financial gain. Here, our data provide little support for an effect of wage inequality on subsequent anti-social behavior. The coefficient on VW associated with our measure of lying is positive and substantial in magnitude – about 35% of a standard deviation – but non-significant. Recall, our related hypothesis anticipated a significant *negative* effect. The additional effect on High-Wage participants, captured by the coefficient on HWxVW, implies a small negative effect but is again non-significant. In fact, none of our experimental factors nor (unreported) demographics has a significant relationship with lying.[21]

[21]The closest to significant at conventional levels is GPA. Our indicator variable for having a GPA above 3.50 is marginally significant ($p = 0.058$) and has an associated coefficient of 0.12, indicating students who report having a high GPA are also more likely to lie.

Result 4: *Hypothesis 4 finds no support. Disclosing wage inequality associated with an ability-intensive Phase 1 task has no significant effect on lying in Phase 2.*

We turn next to income inequality in an ability-intensive environment. Focusing on the disadvantaged first, we find little evidence for any direct effect of revealing income inequality on any of our outcome variables, as measured by the coefficients on VI in columns 1-5. Revealing income inequality in addition to wage inequality, on the other hand, tends to mute the effect of revealing wage inequality alone: the coefficients associated with VWxVI are always opposite in sign and roughly equivalent in magnitude to the coefficients associated with VW but rarely significant. The lone exception to this lack of significance appears in column 2, where the negative and significant coefficient on VWxVI suggests that revealing income inequality in addition to wage inequality essentially completely erases the Scantron quantity gains associated with revealing wage inequality alone among Low-Wage participants.

Among High-Wage participants, the picture is slightly more complicated but follows the same general pattern. The only significant difference in the effect of revealing income inequality by itself on High-Wage participants compared to Low-Wage participants is with respect to the quantity of Scantrons produced. The large, negative and highly significant coefficient on HWxVI suggests an overall negative effect of revealing visible income inequality on High-Wage workers (VI + HWxVI = 0.75 - 1.74 = 0.99). The estimated extra effects of revealing income inequality in addition to wage inequality on High-Wage participants (HWxVWxVI) are typically non-significant. The exception is for the quantity of Scantrons produced, where the sum of the relevant coefficients VI, HWxVI, VWxVI and HwxVWxVI suggests an approximately null effect of -0.22 Scantrons.

Overall, the data suggest that revealing income inequality by itself typically has no direct effect on subsequent behavior. On the other hand, revealing income inequality in addition to wage inequality tends to mute the behavioral responses associated with revealing wage inequality alone.

Result 5: *Hypothesis 5 finds support in our data. Revealing income inequality in isolation never has a significant effect on behavior. Revealing income inequality in addition to wage inequality generally provides countervailing effects on both behavior and on JWBs among Low-Wage participants.*

Before moving on, for completeness we estimate the same econometric models restricted to sessions featuring an effort-intensive Phase 1 task and report the results in Table 4. The picture that emerges is quite easy to summarize. There is a total lack of any significant effect of any of our experimental factors or their interactions on the behaviors we consider. For example, the main

effects of revealing wage inequality to Low-Wage participants, as captured by the coefficients on VW, are small in magnitude and non-significant for all of our outcome measures.

Summing up, we conclude that our data support the view that reactions to wage and income inequality are context-dependent. In particular, they depend heavily on the work environment, whether it is perceived as ability-intensive or effort-intensive. Moreover, the reactions that we document in the ability-intensive environment as well as the lack of reactions we document in an effort-intensive environment are largely in line with our hypotheses. Our hypotheses build directly on prior research suggesting an interaction between contextually determined narratives and subconscious psychological processes. In the next section, we investigate whether our data can provide any supporting evidence on this hypothesized mechanism.

5 The underlying mechanism: meritocracy narrative, cognitive dissonance and JWBs

Recall that we conjectured a particular mechanism relating our specific experimental factors to the subsequent behaviors we observed. We hypothesized that because ability-intensive environments are reminiscent enough of quintessentially meritocratic environments they automatically bring to mind a meritocracy narrative. Because of this, revealing objectively unjustifiable inequality in ability-intensive environments triggers subconscious cognitive processes (Butler 2016) that ultimately strengthen JWBs, particularly among the disadvantaged. These strengthened JWBs have predictable effects (Benabou and Tirole 2006) on subsequent behavior. On the other hand, effort-intensive environments do not naturally give rise to a meritocracy narrative so that the cognitive processes transforming objectively unjustifiable inequality into evidence in favor of a just world are not triggered. Consequently, JWBs are not strengthened, and subsequent behavior is less predictable.

Important supporting evidence for this mechanism would be if we find that JWBs are strengthened by wage inequality revelation in our ability-intensive context but not in our effort-intensive context. Before examining JWBs directly, however, we stress that our analyses here are subject to a very important caveat. Since we only measure JWBs at the end of the experiment and participants have made quite a few decisions between being exposed to our experimental factors and reporting their attitudes, we cannot know whether it is our experimental factors that directly cause variation (or its absence) in measured JWBs, or rather whether variation is mediated by intervening behavior. Nevertheless, examining the relationships between our experimental factors and self-reported JWBs may provide suggestive evidence on the mechanism relating salient inequality to subsequent behavior.

For our measure of JWBs, recall that we asked participants how much they agreed with the statement "people generally get what they deserve" and collected responses on a 7-point Likert scale ranging from 1="totally disagree" to 7="totally agree." In Table 5 we report OLS estimates using this measure of JWBs as the dependent variable. For ease of exposition, as in our analysis above we split our data by the Phase 1 task. In the first column we report a minimal specification for the ability-intensive task data, not controlling for Phase 1 income or Phase 2 quantity. In the second and third columns we add in additional controls sequentially. In columns 4-6 we estimate the same specifications on the effort-intensive task data.

The *least* objectively justifiable form of inequality in our experiment is wage inequality, which participants are explicitly told is uniformly randomly assigned. Moreover, being disadvantaged by this inequality should be particularly difficult to square with a belief that the world is fair. Consequently, we would expect cognitive dissonance to be a particularly acute threat among this group of participants and, hence, for our conjectured mechanism to be the most likely to have bite among them. Overall, this suggests that we might expect the JWB-strengthening effect of revealed wage inequality to be at its peak among the disadvantaged in our ability-intensive sessions.

Considering the first three columns of Table 5, where we restrict to the ability-intensive sessions only, our data are largely consistent with this implication. The coefficients associated with VW represent the effect of revealing wage inequality on those assigned a low wage in an ability-intensive environment. This coefficient is positive, indicating a strengthening of JWBs, highly significant, and substantial in magnitude. The coefficient of 0.41 represents almost one third (29%) of a standard deviation of the unconditional JWBs measure, restricted to the AI session data. Moreover, comparing columns 1-3, this estimate is not affected by adding additional controls. In fact, none of the other controls are statistically significant except Phase 2 quantity.[22]

Turning next to the EI data (columns 4-6), we see qualitatively different patterns as would be expected if the effort-intensive context does not give rise to the same subconscious processes as the ability-intensive environment. In particular, visible wage inequality actually undermines JWBs among disadvantaged participants as evidenced by the by negative and significant coefficients on VW. This is what would be expected from observing unjustifiable inequality that is not subconsciously distorted, as it should (objectively) represent evidence against fairness. As in the AI data, the estimated coefficients on VW are robust to the inclusion of additional controls: moving from column 4 to columns 5 and 6 does not change the magnitude, sign or significance of the estimated effect of revealing wage inequality. Interestingly, consistent with one version of the Fundamental

[22]The estimated coefficient on this variable (column 3) is positive and significant, suggesting there might be a common factor explaining JWBs and productivity — at least following the AI task. It might also obviously reflect reverse-causation through our hypothesized mechanism.

Attribution Error in which people automatically attribute positive developments to their own traits or behaviors, the negative effect of revealing wage inequality is completely erased among the advantaged — the coefficient on HWxVW is opposite in sign, highly significant, and approximately equal in magnitude to the coefficient on VW. Apparently, being randomly assigned a high wage is not interpreted as additional evidence against, or in favor of for that matter, a just world. None of the unreported demographic controls are significant.

In passing, it should also be noted that there is never a significant association between revealed income inequality and JWBs. This is consistent with our conjecture that income inequality is easier to justify by other means, not the least of which is to simply attribute income inequality to individuals' responses to *ex ante* (wage) inequality rather than to an unjust world. Such victim-blaming is characteristic of the psychological literature on JWBs.

Overall, the patterns in the data are consistent with our hypothesized mechanism. Observing objectively unjustifiable and disadvantageous inequality is associated with stronger JWBs in an ability-intensive context. In stark contrast, in an effort-intensive context it is JWBs that give way among the disadvantaged.

6 Discussion and Concluding Remarks

Summing up, our study documents that the (subsequent) behavioral responses to inequality revelation vary qualitatively with the type of inequality revealed, *ex-ante* (wage) or *ex-post* (income), and with the context in which inequality occurred. We focused particularly on the reactions of those disadvantaged by inequality and found that the non-pecuniary incentives of wage inequality complement its pecuniary incentives in ability-intensive contexts, ameliorating moral hazard. However, these non-pecuniary incentives also discourage other-regard among the same population. Analyses utilizing a direct measure of JWBs collected on a post-experiment survey largely supported our novel conjecture about the mechanism underlying these behavioral patterns.

Taken at face value, our study suggests the non-pecuniary incentives of inequality vary substantially, qualitatively and, most importantly, predictably with context. As a consequence, these non-pecuniary incentives should properly be incorporated into economists' formulations of optimal contracts (cf Koszegi 2014).

Not only do firms face important decisions about the optimal compensation structure, but a more recent debate has sprung up about the optimal degree of wage transparency within organizations. Namely, should employee-level wage and earnings data be publicly shared among employees? Proponents tout its ability to potentially boost productivity. The results from our study support this assertion, while also providing evidence of another benefit of wage transparency – reduced

shirking. However, the results from our study also suggest a more nuanced evaluation of wage transparency might be needed. In particular, the potential benefits from revealing a (presumably) unequal pay distribution among employees is likely to be context dependent, and may only accrue within organizations where the work is perceived to be more ability intensive. That said, our findings appear to be consistent with the recent anecdotal cases of tech firms and internet start-ups having reported success with moving to more pay transparency.[23]

External validity is a concern, as always with laboratory experiments, and ours is no exception. As a modicum of reassurance, it is worth mentioning that crucial features of the mechanism we conjecture have been documented in a wide array of settings. Evidence for the cognitive processes at work has appeared across multiple disciplines, including economics, across a long time horizon and across various settings: in laboratory experiments using student subject pools; in on-line experiments using working adults; in field experiments; and also in observational data (inter alia, Benabou and Tirole 2006; Bénabou and Tirole 2016; Butler 2016; Hart and Piff 2018; Lerner 2013, 1965).

Even so, we acknowledge that features of our setting — our subject pool consisted of business school undergraduates at a university in west Texas — may have been particularly conducive to inducing a meritocracy narrative. We leave for future research whether this narrative, and its related behavioral consequences, is readily induced among other populations. Another important question for future research is how this narrative can be induced. In particular, it is an open question whether the actual content of work is important or, rather, whether the perception of ability-intensiveness can be influenced through job descriptions, mission statements or other aspects of corporate culture.

References

Abeler, J., S. Altmann, S. Kube, and M. Wibral (2010): "Gift exchange and workers' fairness concerns: when equality is unfair," *Journal of the European Economic Association*, 8, 1299–1324.

Abeler, J., D. Nosenzo, and C. Raymond (2016): "Preferences for truth-telling," .

Adams, J. S. (1963): "Towards an understanding of inequity." *The Journal of Abnormal and Social Psychology*, 67, 422.

Akerlof, G. A. and W. T. Dickens (1982): "The Economic Consequences of Cognitive Dissonance," *The American Economic Review*, 72, 307–319.

[23]For example, see the recent popular press articles by (Burkus 2016; McLaren 2019; Weller 2017).

Akerlof, G. A. and J. L. Yellen (1990): "The fair wage-effort hypothesis and unemployment," *The Quarterly Journal of Economics*, 105, 255–283.

Angelova, V., W. Güth, and M. G. Kocher (2012): "Co-employment of permanently and temporarily employed agents," *Labour Economics*, 19, 48–58.

Bartling, B. and F. A. Von Siemens (2011): "Wage inequality and team production: An experimental analysis," *Journal of Economic Psychology*, 32, 1–16.

Benabou, R. and J. Tirole (2006): "Belief in a just world and redistributive politics," *The Quarterly journal of economics*, 121, 699–746.

Bénabou, R. and J. Tirole (2016): "Mindful economics: The production, consumption, and value of beliefs," *Journal of Economic Perspectives*, 30, 141–64.

Bolton, G. and P. Werner (2016): "The influence of potential on wages and effort," *Experimental Economics*, 19, 535–561.

Bracha, A., U. Gneezy, and G. Loewenstein (2015): "Relative pay and labor supply," *Journal of labor economics*, 33, 297–315.

Breza, E., S. Kaur, and Y. Shamdasani (2018): "The morale effects of pay inequality," *The Quarterly Journal of Economics*, 133, 611–663.

Burkus, D. (2016): "Why do we keep salaries secret?" *Forbes*.

Butler, J. V. (2016): "Inequality and relative ability beliefs," *The Economic Journal*, 126, 907–948.

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

Charness, G., R. Cobo-Reyes, J. A. Lacomba, F. Lagos, and J. M. Pérez (2016): "Social comparisons in wage delegation: Experimental evidence," *Experimental Economics*, 19, 433–459.

Charness, G. and P. Kuhn (2007): "Does pay inequality affect worker effort? Experimental evidence," *Journal of labor economics*, 25, 693–723.

Charness, G. and M. Rabin (2002): "Understanding social preferences with simple tests," *The Quarterly Journal of Economics*, 117, 817–869.

Chen, D. L., M. Schonger, and C. Wickens (2016): "oTree–An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 9, 88–97.

Clark, A. E., D. Masclet, and M. C. Villeval (2010): "Effort and comparison income: Experimental and survey evidence," *ILR Review*, 63, 407–426.

Cohn, A., E. Fehr, B. Herrmann, and F. Schneider (2014): "Social comparison and effort provision: Evidence from a field experiment," *Journal of the European Economic Association*, 12, 877–898.

Collier, P. (2016): "The cultural foundations of economic failure: A conceptual toolkit," *Journal of Economic Behavior & Organization*, 126, 5–24.

Conlin, B. (2018): "The Pros and Cons of Salary Transparency," *Business News Daily*.

Cooney, S. (2018): "Should you Share your Salary with Co-Workers? Here's what Experts Say," *Time Magazine*.

Cullen, Z. and R. Perez-Truglia (2018): "How much does your boss make? the effects of salary comparisons," *NBER Working Paper no. 24841*.

Di Tella, R., S. Galiant, and E. Schargrodsky (2007): "The formation of beliefs: evidence from the allocation of land titles to squatters," *The Quarterly Journal of Economics*, 122, 209–241.

Dube, A., L. Giuliano, and J. Leonard (2019): "Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior," *American Economic Review*, 109, 620–663.

Dufwenberg, M. and M. A. Dufwenberg (2018): "Lies in disguise–A theoretical analysis of cheating," *Journal of Economic Theory*, 175, 248–264.

Edmans, A. and X. Gabaix (2016): "Executive compensation: A modern primer," *Journal of Economic literature*, 54, 1232–87.

Falk, A. and A. Ichino (2006): "Clean evidence on peer effects," *Journal of labor economics*, 24, 39–57.

Festinger, L. (1957): *A theory of cognitive dissonance*, vol. 2, Stanford university press.

Fischbacher, U. and F. Föllmi-Heusi (2013): "Lies in disguise–an experimental study on cheating," *Journal of the European Economic Association*, 11, 525–547.

Fisher, A. (2015): "How Pay Transparency can Keep People from Quitting," *Fortune*.

Gächter, S. and C. Thöni (2010): "Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis," *Journal of Economic Behavior & Organization*, 76, 531–543.

Gibbons, R. (1987): "Piece-rate incentive schemes," *Journal of Labor Economics*, 5, 413–429.

Gill, D., V. Prowse, and M. Vlassopoulos (2013): "Cheating in the workplace: An experimental study of the impact of bonuses and productivity," *Journal of Economic Behavior & Organization*, 96, 120–134.

Gneezy, U., A. Kajackaite, and J. Sobel (2018): "Lying Aversion and the Size of the Lie," *American Economic Review*, 108, 419–53.

Godechot, O. and C. Senik (2015): "Wage comparisons in and out of the firm. Evidence from a matched employer–employee French database," *Journal of Economic Behavior & Organization*, 117, 395–410.

Greiner, B., A. Ockenfels, and P. Werner (2011): "Wage transparency and performance: A real-effort experiment," *Economics Letters*, 111, 236–238.

Gross, T., C. Guo, and G. Charness (2015): "Merit pay and wage compression with productivity differences and uncertainty," *Journal of Economic Behavior & Organization*, 117, 233–247.

Hart, E. and P. K. Piff (2018): "When Inequality is Good for Groups: The Moderating Role of Fairness," *SSRN Working Paper*.

Hennig-Schmidt, H., A. Sadrieh, and B. Rockenbach (2010): "In search of workers' real effort reciprocity–a field and a laboratory experiment," *Journal of the European Economic Association*, 8, 817–837.

Hopkins, E. and T. Kornienko (2009): "Status, Affluence, and inequality: Rank-based comparisons in games of status," *Games and Economic Behavior*, 67, 552–568.

——— (2010): "Which inequality? The inequality of endowments versus the inequality of rewards," *American Economic Journal: Microeconomics*, 2, 106–37.

Huet-Vaughn, E. (2017): "Do social comparisons motivate workers? A field experiment on relative earnings, labor supply and the inhibitory effect of pay inequality," Tech. rep., Working Paper.

Koszegi, B. (2014): "Behavioral contract theory," *Journal of Economic Literature*, 52, 1075–1118.

Lazear, E. P. (2000): "The power of incentives," *American Economic Review*, 90, 410–414.

——— (2018): "Compensation and Incentives in the Workplace," *Journal of Economic Perspectives*, 32, 195–214.

Lazear, E. P. and P. Oyer (2007): "Personnel economics," Tech. rep., National Bureau of economic research.

Lerner, M. (2013): *The Belief in a Just World: A Fundamental Delusion*, Springer Science & Business Media.

Lerner, M. J. (1965): "Evaluation of performance as a function of performer's reward and attractiveness." *Journal of Personality and Social Psychology*, 1, 355.

Mankiw, G. (2006): "Why Aspiring Economists Need Math," http://gregmankiw.blogspot.com/2006/09/why-aspiring-economists-need-math.html, accessed: 2019-07-26.

Mazar, N., O. Amir, and D. Ariely (2008): "The dishonesty of honest people: A theory of self-concept maintenance," *Journal of marketing research*, 45, 633–644.

McLaren, S. (2019): "Why These 3 Companies Are Sharing How Much Their Employees Make," *LinkedIn Talent Blog*.

Nosenzo, D. (2013): "Pay secrecy and effort provision," *Economic inquiry*, 51, 1779–1794.

Prendergast, C. (1999): "The provision of incentives in firms," *Journal of economic literature*, 37, 7–63.

Rosen, S. (1982): "Authority, control, and the distribution of earnings," *The Bell Journal of Economics*, 311–323.

Shellenbarger, S. (2016): "Open Salaries: The Good, the Bad, and the Awkward," *Wall Street Journal*.

Shiller, R. J. (2017): "Narrative economics," *American Economic Review*, 107, 967–1004.

Spence, M. (1973): "Job Market Signaling," *The Quarterly Journal of Economics*, 87, 355–374.

Sseruyange, J. and E. Bulte (2019): "Wage Differentials and Workers' Effort: Experimental Evidence from Uganda," *Oxford Bulletin of Economics and Statistics*.

Stiglitz, J. E. (1975): "Incentives, risk, and information: notes towards a theory of hierarchy," *The Bell Journal of Economics*, 552–579.

Weller, C. (2017): "A CEO who makes everyone's salary transparent says people are more productive than ever," *Business Insider*.

Tables and Figures

Tables

Table 2: *Descriptive Statistics*

Treatment	Pay scheme	N	Phase 1		Phase 2			Phase 3
			income	score	quantity	quality	lying	donations
IW-II-EI	Low Wage	27	5.76	28.81	7.67	0.99	0.04	0.17
	High Wage	27	10.06	28.07	7.67	0.98	0.04	0.14
IW-II-AI	Low Wage	32	4.18	20.91	6.91	0.88	0.16	0.30
	High Wage	24	7.29	19.38	7.75	0.92	0.13	0.24
VW-II-EI	Low Wage	26	6.00	30.00	7.88	0.96	0.08	0.23
	High Wage	27	10.60	29.89	8.26	0.96	0.11	0.15
VW-II-AI	Low Wage	24	4.03	20.17	7.75	0.96	0.25	0.16
	High Wage	34	7.70	19.03	7.74	0.96	0.09	0.22
IW-VI-EI	Low Wage	22	5.74	28.68	7.59	0.97	0.09	0.28
	High Wage	30	10.30	29.07	7.93	0.95	0.03	0.12
IW-VI-AI	Low Wage	28	4.01	20.04	7.64	0.94	0.14	0.25
	High Wage	21	7.52	20.00	6.76	0.95	0.10	0.17
VW-VI-EI	Low Wage	30	5.53	27.67	6.73	0.96	0.10	0.26
	High Wage	23	10.66	31.30	8.09	0.94	0.04	0.13
VW-VI-AI	Low Wage	20	3.85	19.25	6.85	0.92	0.10	0.19
	High Wage	28	7.63	19.79	7.25	0.96	0.11	0.13

Notes: **[1]** Treatments are denoted using a triple (V/I)W-(V/I)I-(E/A)I, where the first component refers to whether W(age) inequality was visible or not, the second component refers to the visibility of I(ncome) inequality and the last component refers to the task type – A(bility) or E(ffort) intensive. **[2]** Columns 4-8 present averages of the variables listed in the column headings. "Phase 1 income" refers to participants' earnings from the phase 1 task only; "Phase 1 score" refers to the number of correct questions (out of a maximum of 48) on the phase 1 task; "Phase 2 quantity" refers to the number of scantrons submitted; "Phase 2 quality" refers to the average accuracy of submitted scantrons; "Phase 2 lying" is the proportion of participants who misreported the number of Scantrons they submitted; "Phase 3 donations" refers to the proportion of total experimental income donated to either of the charity options (American Red Cross and the South Plains Food Bank).

Table 3: *OLS Estimates AI only*

	Phase 1	Phase 2			Phase 3
	score	quantity	quality	lying	donations
HW	-1.23	0.71	-0.01	0.07	0.02
	(1.582)	(0.582)	(0.060)	(0.106)	(0.110)
VW	-0.51	1.00**	0.07**	0.12	-0.12**
	(1.088)	(0.417)	(0.033)	(0.117)	(0.051)
VI	-0.62	0.75	0.05	-0.01	-0.03
	(1.382)	(0.513)	(0.037)	(0.076)	(0.096)
HWxVW	0.35	-1.17***	-0.04	-0.16	0.11
	(2.360)	(0.340)	(0.042)	(0.113)	(0.082)
HWxVI	1.04	-1.74**	-0.01	-0.04	-0.10
	(2.236)	(0.634)	(0.044)	(0.113)	(0.131)
VWxVI	0.21	-1.47**	-0.08	-0.13	0.03
	(1.553)	(0.613)	(0.056)	(0.149)	(0.122)
HWxVWxVI	-0.56	2.24**	0.05	0.19	-0.01
	(2.915)	(0.948)	(0.087)	(0.172)	(0.142)
Phase 1 income		0.07	0.01	-0.02	-0.01
		(0.081)	(0.008)	(0.022)	(0.020)
Phase 2 quantity			0.03**	-0.01	-0.03**
			(0.011)	(0.014)	(0.012)
Constant	32.36***	7.17***	0.71***	0.33	0.27
	(4.051)	(1.156)	(0.161)	(0.262)	(0.243)
Observations	207	207	207	207	207
R-squared	0.122	0.162	0.190	0.074	0.123

Notes: **[1]** Each column reports an OLS regression restricted to observations from sessions featuring the AI Phase 1 task with the dependent variable labeled in the column heading. "Phase 1 income" is participants' earnings from the the Phase 1 task alone; "Phase 2 quantity" refers to the number of scantrons submitted; "Phase 2 quality" refers to the average accuracy of submitted scantrons; "Phase 2 lying" is the proportion of participants who misreported the number of Scantrons they submitted; "Phase 3 donations" refers to the total amount donated across both charity options (American Red Cross and the South Plains Food Bank); "JWB" is the participant's response to the Just World Beliefs question. **[2]** Experimental controls are as follows: *HW* is an indicator variable for the High-Wage pay scheme, *VW* is an indicator for visible pay inequality, *VI* is an indicator variable visible (experimental) income inequality. **[3]** Included in each estimate, but not reported for readability, are additional controls: gender, age, high- and low-GPA dummies, high- and low-family-income dummies, as well as dummies for the type of filler questions inserted between stages 1 and 1.5, which were randomly assigned and asked about gender, political affiliation or a purchasing decision. **[4]** Robust standard errors clustered by session appear in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4: *OLS Estimates EI Only*

	Phase 1	Phase 2			Phase 3
	score	quantity	quality	lying	donations
HW	-1.09	-1.17	-0.06	0.10	0.03
	(0.888)	(0.701)	(0.037)	(0.083)	(0.081)
VW	1.30	0.16	-0.03	0.05	0.08
	(1.214)	(0.613)	(0.033)	(0.046)	(0.069)
VI	-0.05	-0.09	-0.02	0.07	0.12
	(1.429)	(0.440)	(0.014)	(0.071)	(0.084)
HWxVW	0.44	0.31	0.01	0.02	-0.04
	(1.989)	(0.675)	(0.019)	(0.089)	(0.053)
HWxVI	1.52	0.50	-0.01	-0.05	-0.10
	(1.796)	(0.755)	(0.019)	(0.061)	(0.104)
VWxVI	-2.50	-0.96	0.01	-0.07	-0.14
	(1.796)	(0.909)	(0.046)	(0.116)	(0.116)
HWxVWxVI	2.37	0.53	-0.01	-0.01	0.13
	(2.614)	(1.020)	(0.045)	(0.133)	(0.126)
Phase 1 income		0.23	0.01	-0.02	-0.02
		(0.156)	(0.009)	(0.015)	(0.014)
Phase 2 quantity			-0.01	-0.00	-0.02***
			(0.008)	(0.006)	(0.006)
Constant	31.90***	4.23**	0.98***	0.33**	0.18
	(2.823)	(1.442)	(0.099)	(0.141)	(0.222)
Observations	212	212	212	212	212
R-squared	0.089	0.118	0.056	0.063	0.143

Notes: **[1]** Each column reports an OLS regression restricted to observations from sessions featuring the EI Phase 1 task with the dependent variable labeled in the column heading. "Phase 1 income" is participants' earnings from the the Phase 1 task alone; "Phase 2 quantity" refers to the number of scantrons submitted; "Phase 2 quality" refers to the average accuracy of submitted scantrons; "Phase 2 lying" is the proportion of participants who misreported the number of Scantrons they submitted; "Phase 3 donations" refers to the total amount donated across both charity options (American Red Cross and the South Plains Food Bank); "JWB" is the participant's response to the Just World Beliefs question. **[2]** Experimental controls are as follows: HW is an indicator variable for the High-Wage pay scheme, VW is an indicator for visible pay inequality, VI is an indicator variable visible (experimental) income inequality. **[3]** Included in each estimate, but not reported for readability, are additional controls: gender, age, high- and low-GPA dummies, high- and low-family-income dummies, as well as dummies for the type of filler questions inserted between stages 1 and 1.5, which were randomly assigned and asked about gender, political affiliation or a purchasing decision. **[4]** Robust standard errors clustered by session appear in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5: *JWBs Estimates*

	AI only			EI only		
HW	-0.14	-0.14	-0.14	-0.09	-0.09	-0.09
	(0.259)	(0.259)	(0.259)	(0.437)	(0.437)	(0.437)
VW	0.41***	0.41***	0.41***	-0.97**	-0.97**	-0.97**
	(0.122)	(0.122)	(0.122)	(0.399)	(0.399)	(0.399)
VI	0.29	0.29	0.29	-0.58	-0.58	-0.58
	(0.205)	(0.205)	(0.205)	(0.631)	(0.631)	(0.631)
HWxVW	0.21	0.21	0.21	1.02***	1.02***	1.02***
	(0.469)	(0.469)	(0.469)	(0.297)	(0.297)	(0.297)
HWxVI	0.15	0.15	0.15	0.51	0.51	0.51
	(0.292)	(0.292)	(0.292)	(0.629)	(0.629)	(0.629)
VWxVI	-0.22	-0.22	-0.22	1.18	1.18	1.18
	(0.389)	(0.389)	(0.389)	(0.689)	(0.689)	(0.689)
HWxVWxVI	-0.84	-0.84	-0.84	-1.32	-1.32	-1.32
	(0.710)	(0.710)	(0.710)	(0.869)	(0.869)	(0.869)
Phase 1 income		0.08	0.08		-0.17**	-0.17**
		(0.052)	(0.052)		(0.081)	(0.081)
Phase 2 quantity			0.12**			-0.03
			(0.042)			(0.030)
Constant	0.37	0.37	0.37	5.52***	5.52***	5.52***
	(1.349)	(1.349)	(1.349)	(0.850)	(0.850)	(0.850)
Observations	207	207	207	209	209	209
R-squared	0.117	0.117	0.117	0.161	0.161	0.161

Notes: [**1**] Each column reports an OLS regression restricted to observations from either AI or EI sessions. [**2**] The dependent variable in each column is our measure of *Just World Beliefs*, which is a participant's degree of agreement with the statement "People generally get what they deserve." Valid responses range from 1="totally disagree" to 7="totally agree." Independent variables are as follows: "Phase 1 income" is participants' earnings from the the Phase 1 task alone; "Phase 2 quantity" refers to the number of scantrons submitted; "Phase 2 quality" refers to the average accuracy of submitted scantrons; "Phase 2 lying" is the proportion of participants who misreported the number of Scantrons they submitted; "Phase 3 donations" refers to the total amount donated across both charity options (American Red Cross and the South Plains Food Bank); "JWB" is the participant's response to the Just World Beliefs question. [**3**] Experimental controls are as follows: *HW* is an indicator variable for the High-Wage pay scheme, *VW* is an indicator for visible pay inequality, *VI* is an indicator variable visible (experimental) income inequality. [**4**] Included in each estimate, but not reported for readability, are additional controls: gender, age, high- and low-GPA dummies, high- and low-family-income dummies, as well as dummies for the type of filler questions inserted between stages 1 and 1.5, which were randomly assigned and asked about gender, political affiliation or a purchasing decision. [**4**] Robust standard errors clustered by session appear in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figures

Figure 2: *Kernel Density Estimates of Phase 1 Earnings*

