EIEF WORKING PAPER SERIES

Einaudi Institute for Economics and Finance

EIEF Working Paper 05/13
March 2013

Inequality and Relative Ability Beliefs

by

Jeffrey V. Butler

(EIEF)

Inequality and Relative Ability Beliefs*

Jeffrey V. Butler

Einaudi Institute for Economics and Finance, Rome, Italy

jeff.butler@eief.it

Abstract: In this study I present experimental evidence of a novel channel yielding inequality persistence. In an initial experiment, results suggest that individuals respond to salient inequality by adjusting their performance beliefs to justify the inequality. Subsequent experiments reveal: i) that it is beliefs about relative ability—an ostensibly stable trait—rather than effort provision that respond to inequality; and that ii) unequal pay in an initial task affects willingness to compete on a subsequent task for male participants. Taken together, these patterns may cause inequality to become self-perpetuating. I conclude by discussing some implications of these findings.

JEL Classifications: D84, J71, Z13, A12, D03

Keywords: Confidence, Experiment, Inequality, Identity, Discrimination, Just World Beliefs

*I owe a great debt of gratitude to George A. Akerlof, Stefano DellaVigna, Luigi Guiso, Shachar Kariv and Georg Weiszäcker for helpful comments. I would also like to thank seminar participants at LUISS Guido Carli University, the European University Institute, Southern Methodist University, University of Arizona and conference participants at the 2012 North-American ESA Conference, Tuscon. I am also grateful to to the National Science Foundation under Research Grant SES 04-17871 and UC Berkeley's X-lab for financial support.

1 Introduction

Given the rising tide of wealth and income inequality in the US and around the world, the causes and consequences of inequality are once again occupying the attention of many economists. If, as seems likely, inequality and disadvantage persist for reasons not solely due to differences in fundamental traits or preferences (e.g., ability or disutility of effort), then understanding what causes disadvantage to persist is of interest on both moral and efficiency grounds. Recent research has highlighted group-level variation in the willingness to compete as a potential explanation for persistent, durable, inequality (Hoff and Pandey, 2006; Niederle and Vesterlund, 2007; Afridi, Li and Ren, 2012). The logic is straightforward: many pathways to upward mobility (lucrative careers, prestigious colleges) feature highly competitive environments, so that shying away from competition restricts one's opportunities to succeed. If disadvantaged groups are also, for whatever reason, less willing to compete, then disadvantage and inequality are perpetuated.

While current research documents diminished performance and willingness to compete among real-world historically disadvantaged groups—particularly when group affiliation is made salient (caste, gender or Chinese *hukou*)—the mechanisms through which disadvantage undermines competitiveness is not well understood. This is largely due to the methodology of existing studies, which rely on examining differences in competitiveness among *real-world* groups. Specifically, while varying the salience of *real-world* group affiliations has the advantage of being directly applicable to actual people groups in particular cultures, it has one major drawback: lack of control over what is being varied by making existing divisions salient. Does making group affiliation salient activate preferences to conform to group-based stereotypes? Does emphasizing gender, caste or hukou evoke culturally transmitted beliefs about the inferiority of particular groups? More fundamentally, the question arises as to whether salient inequality impacts behavior indirectly, by modifying beliefs upon which economic incentives rely, or directly by changing preferences—or a combination of both? Understanding the mechanisms through which inequality affects behavior is important, as it is a necessary prerequisite to constructing policies effective at counteracting the negative consequences of inequality.

An alternative, complementary, approach is to use laboratory experiments to create artificial groups and directly manipulate group-level disadvantage. While this approach lacks the realism attainable using pre-existing real world group boundaries, it has the advantage of allowing the researcher control over exactly which form inequality takes. This additional level of control potentially permits a finer look at the mechanisms translating disadvantage into willingness to compete. While many forms of inequality would be potentially interesting to study, the scope for treating some participants badly is limited by both informed consent and practicality. Consequently, in this paper I employ laboratory experiments to implement a form of inequality of particular interest to economists — unequal earnings. By paying two artificially created groups differently for performing the same task, I seek to shed light from a different angle than existing studies on the relationship between inequality and competitiveness.

I take as a starting point the observation that, while many years of persistent inequality may, through several processes, obviously create a true correlation between social position and fundamental attributes such as competitiveness, such long-run effects cannot be the whole story. For example, Afridi, Li and Ren (2012) replicate the findings of Hoff and Pandey (2006) using unequal groups created only decades ago. This leads me to focus on beliefs as the prime candidate for a mechanism connecting willingness to compete and disadvantage.

For testable hypotheses, I draw on research in psychology and sociology demonstrating that people tend to believe there is a fit between a person's outcomes and his or worthiness. A striking example of this tendency is reported in a seminal study by Lerner (1965) where participants observed two other participants working at an anagram task. Observers were informed that one of these other two workers would be randomly chosen to be paid. He found that "...once the outcome was known to the observers they tended to persuade themselves that the person who had been awarded the money by chance had really earned it, after all." While there is considerable evidence of this phenomenon when it comes to beliefs about *others'* worthiness—"Belief in a Just World"—relatively little is known about how strong this tendency is when the beliefs in question involve *one's own* worthiness or merit. If Just World beliefs continue to exert substantial influence even with respect to one's own merit, this provides a channel through which initial inequality may affect competitiveness and perpetuate itself: by undermining (enhancing) the relative ability beliefs of the initially disadvantaged (advantaged) and reducing (increasing) the ex-ante expected value of rewards depending on ability-based competition.

Whether Just World beliefs impact assessments involving an individual's own abilities is not *a priori* obvious, as studies focusing on *over*confidence provide evidence for various ego-preserving biases which might dampen or even reverse the effects of Just World beliefs (Eil and Rao, 2010; Grossman and Owens, 2012). Consequently, the current inquiry tests two main hypotheses. The first testable hypothesis is that inequality affects beliefs about one's own (relative) ability in a way consistent with the Belief in a Just World phenomenon.

Hypothesis 1: *Disadvantage decreases relative ability beliefs, while advantage increases relative ability beliefs.*

The second hypothesis is that inequality exerts sufficiently strong and persistent influence on beliefs to color an individual's willingness to compete on future tasks, even when subsequent tasks do not involve pay inequality.

Hypothesis 2: *Initial disadvantage reduces subsequent willingness to compete irrespective of pay inequality.*

I test these hypotheses using three separate experiments. As an initial test of Hypothesis 1, in Experiment 1 participants are randomly assigned to one of two groups—one group will be paid relatively well for performance on a task, while the other group will be paid relatively poorly for performance on the same task. The task is designed to require some analytical ability. After learning their pay group assignments, participants complete the task and earnings are determined. Finally, participants' beliefs about their *relative* task performance are elicited in an incentive compatible manner. A crucial component of the pay structure implemented is that marginal performance incentives are held constant across pay groups. Standard economic theory predicts this should result in equal performance across pay groups, which is a feature essential to identifying the direct effect of inequality on beliefs—otherwise performance beliefs could vary simply because performance varies.

To dig deeper into the mechanisms relating inequality to beliefs, and to ultimately test Hypothesis 2, Experiments 2 and 3 feature the same form of pay inequality as in Experiment 1—unequal pay across artificially-induced groups, with marginal performance incentives held constant. Differently from Experiment 1, both Experiment 2 and Experiment 3 feature random task assignment—participants are assigned either an ability-intensive task or an effort-intensive task. Randomly assigning the ability-intensiveness of tasks allows me to address important confounds in the attribution of belief variation to inequality *per se* present in Experiment 1.[2] Experiment 3 differs from Experiment 2 by the addition of subsequent phase in which participants choose whether to apply piece-rate or tournament incentives to a second task. The incentive structure choice on this subsequent task allows me to test directly whether previous inequality affects individuals' subsequent willingness to compete (Hypothesis 2). More generally, all three experiments together provide evidence of the robustness of the effect of inequality on beliefs.

As an overview of the results, I find that *mere inequality* affects relative performance beliefs without affecting actual performance, suggesting inequality changes beliefs directly and does not alter effort provision. This distinction is further supported in Experiments 2 and 3: inequality affects beliefs only when task performance relies heavily on ability. Since ability is likely a persistent trait, this provides a rationale to expect the effects of past inequality on retrospective performance beliefs to affect subsequent decisions about whether to shy away from competiton. This latter pattern would provide the scope nec-

[2] For example, it could be that low pay simply "demoralizes" some participants, causing them to expect low performance because they put in low effort. If so, one would expect demoralization to affect relative performance beliefs—perhaps even moreso—on a task where performance depends more on effort than on ability.

essary for mere inequality to become self-perpetuating. Examining this final link directly, I find strong, but gender-specific, patterns. Past inequality affects male participants' subsequent willingness to compete on the ability-intensive task, but has no effect on female participant's competition decision. At the same time, my data suggest the lack of an effect of past inequality on subsequent competitiveness among female participants is not due to a fundamental difference in competitive preferences: on the effort-intensive task, where beliefs are not affected by inequality, women are no less likely than men to select into competition.

Taken together, all three experiments reported in this study provide evidence for a novel mechanism perpetuating inequality: initial inequality colors beliefs about one's own ability relative to others, lowering the ex-ante expected return to courses of action which require, at some point, ability-based competition. Since ability-based competition is a feature of many paths to upward mobility, initial inequality may become persistent inequality.

The remainder of the paper is organized as follows. First, I discuss closely related literature. In Section 3, the design of Experiment 1 is presented in detail and results are presented subsequently. In Section 4, the design of Experiment 2 is presented, followed by results. In Section 5, Experiment 3's design is discussed, again followed by results. Section 6 summarizes and concludes.

2 Closely Related Literature

An influential handful of studies suggest that historically disadvantaged groups exhibit lower willingness to compete. Gneezy, Niederle and Rustichini (2003) present experimental evidence that women perform worse than men under tournament incentives, particularly when members of both genders are present simultaneously.[3] One reason for this mixed-gender-session effect could be that the presence of both males and females makes gender a salient group boundary. Accordingly, Hoff and Pandey (2006) find that explicitly varying the salience of another unequal group boundary—Indian caste affiliiation—has similar

[3] It is generally agreed that women have a long history of being disadvantaged both *de jure* and *de facto* in many countries.

effects: members of low caste groups underperform their high caste counterparts under tournament incentives, but only when caste is made salient. Afridi, Li and Ren (2012) find similar results when manipulating the salience of another unequal group bounday—Chinese *hukou* system. A potential explanation unifiying these results is that making group affiliations salient evokes memories of unequal treatment among the disadvantaged, undermining current beliefs about relative ability and performance and thereby weakening tournament incentives among disadvantaged groups. This chain involves several links, of course, which the studies mentioned investigate only indirectly. Besides using experimentally-induced unequal groups rather than real world groups, the most important way the current study differs from those just mentioned is by eliciting beliefs to examine directly the beliefs link in the causal chain outlined above.

Another set of closely related studies are Niederle and Vesterlund (2007) and Gneezy, Leonard and List (2009), both of which use experiments with monetary incentives. The former study demonstrates a pronounced gender effect among a typical US experimental participant population: women in this population are half as likely as their male counterparts to select tournament-based compensation over piece-rate pay and, moreover, even though they perform no worse on average, women also report believing they perform relatively worse. In the latter studies the authors ask whether the interaction between gender and competitiveness is culturally dependent. They conduct identical experiments in two societies. One society is quintessentially patriarchial (the Maasai in Tanzania) while the other is matrilineal (the Khasi in India). The authors show that competitiveness patterns flip across these two societies. Among the patriarchal Maasai, men are more likely to select tournament incentives than women, whereas among the matrilineal Khasi, it is women who are more competitive. Taken together, these studies provide suggestive evidence that disadvantage undermines willingness to compete. Still, despite being incredibly well designed and executed the evidence they present remains only suggestive on this point. The latter study provides the most direct evidence, but since many things may differ between the men and women in the societies involved besides gender roles, attributing the differences to a single factor like unequal treatment is difficult.[4] Differently from these studies, the current inquiry randomly assigns unequal treatment, making identification less problematic.

Outside of economics, the current study builds on two closely veins of research in social psychology and sociology: Belief in a Just World (Lerner, 1965) and System Justification Theory (see, e.g., Jost et al., 2003). A central hypothesis of both theories is that a basic human need for justice coupled with an aversion to cognitive dissonance causes individuals to rationalize an unequal *status quo* by choosing to believe existing inequality is deserved. Lerner (1965), in a seminal contribution to these literatures, provides non-incentivized experimental evidence that third-party observers rated workers who were randomly chosen to be paid as more deserving of pay than those workers who were randomly chosen not to be paid. Recognition of this type of phenomenon dates back to at least Allport (1954), while more recent social psychological research suggests that the phenomenon is fundamental by demonstrating its prevalence across multiple cultures and nations (Cuddy, *et al*, 2009; Fiske, Cuddy and Glick, 2007; Caprariello, Cuddy and Fiske, 2009).[5] Whether beliefs about oneself are affected by inequality in a manner similar to the way beliefs about others or other groups vary in Just World/System Justification research is an important open question whose answer is not obvious. Given the large literature documenting *over*confidence, one might *a priori* expect myriad psychological mechanisms serving to protect self image and self esteem motives to an overwhelmingly countervailing force. Differently from these veins of research, the current study uses an incentive compatible belief elicitation mechanism in experiments with monetary incentives to investigate whether beliefs about one's *own* relative ability are affected by randomly-assigned initial inequality, and whether the effect, if any, persists to affect future decisions.

A less directly related strand of research that also seeks to explain why disadvantage may persist is "stereotype threat" (Steele and Aronson, 1995). In this view, a preference

[4]The direction of causality may even be an issue, as it could be that societies whose women are more competitive—perhaps because of a competitiveness trait or gene—are more likely to be matrilineal or matriarchal.

[5]For a discussion of cognitive dissonance as well as other economically-relevant implications, see, e.g., Akerlof and Dickens (1982)

to conform to group-based stereotypes leads underprivileged groups about whom negative stereotypes exist to underperform and, consequently, perhaps to rationally opt out of competitive environments. For example, Steele and Aronson show that making race salient causes African-Americans to underperform others on a Scholastic Aptitude Test. However, disadvantage and negative stereotypes do not perfectly overlap. For some underprivileged minority groups (e.g., some Asian immigrant groups) positive stereotypes exist. Making race or ethnicity salient for such groups has been shown to *increase* performance. Differently from the stereotype threat literature, the current study features randomly-assigned inequality about which the existence of a pre-existing stereotype is unlikely.

Finally, the study is tangentially related to a handful of fascinating studies investigating the relationship between labor provision and unequal pay among primates (e.g., Brosnan and De Waal, 2003; Brosnan et al., 2005). It has been shown that primates react negatively—particularly those disadvantaged by pay inequality—when observing other primates being paid more for performing the same task. The current study differs in many ways from the primate studies. One fundamental way in which it differs is by implementing a pay structure featuring equal marginal incentives for performance. By fixing marginal performance incentives, standard economic theory predicts actual task performance should not vary across pay groups.

3 Experiment 1

3.1 Design and procedures

Experiment 1 was conducted in the XLab facilities at the University of California, Berkeley using the software z-tree (Fischbacher, 2007). Participants were recruited among students and staff of the university. Three separate treatments were conducted. Each participant took part in exactly one of the three treatments. All three treatments shared a common structure: first, participants were informed about the task they would be performing; next, they learned about the pay structure and, if applicable, were randomly assigned one of two pay levels; subsequently, participants completed ten round of an analytical ability-intensive task; finally, after all rounds of the analytical ability task were completed, participants' beliefs about how well they they performed relative to other participants were elicited in an incentive compatible manner. Importantly, participants learned about the belief-elicitation stage only after having completed the analytical ability task so that belief elicitation should not have affected behavior.

3.1.1 Pay structure: varying inequality across treatments

To investigate the impact of inequality on relative performance and related beliefs, the pay structure varied across the three treatments comprising the experiment. In the Payoff Inequality (PI) treatment, half of the participants in each session were randomly assigned to the high pay group (HP) while the other half were assigned to a group earning low pay (LP). The fact that assignment was random, and that each person had equal chances of being assigned to either group was made known to all participants. Participants assigned to HP earned \$4 for each correct answer on the analytical ability-intensive task (described below) and \$2 for each incorrect answer. Participants assigned to LP were paid \$2 for each correct answer and \$0 for each incorrect answer. An individual's pay group persisted over all ten rounds of the task. When submitting answers, participants were informed of the pay group of each preceding guesser.

The two other treatments can be viewed as control sessions, in the sense that they implement each of the two pay structures involved in PI, separately. In the Contol-High Pay (C-HP) treatment all participants earned \$4 for each correct answer on the analytical ability-intensive task and \$2 for each incorrect answer. Meanwhile, in the Control-Low Pay (C-LP) treatment participants earned \$2 for each correct answer and \$0 for each incorrect answer.

The essential feature of these two pay structures—HP and LP—is that marginal performance incentives are identical across pay groups. Each correct answer always pays two dollars more than each incorrect answer. This implies that task performance should not vary across pay groups, which will prove essential to cleanly identifying the effect of inequality on beliefs. At the same time, the two pay structures combined implement severe inequality: the most a participant assigned to the LP group could earn was the least an HP participant could earn from task performance.

3.1.2 The analytical ability task in Experiment 1

The analytical ability task used is an urn-guessing game common in the social learning literature (see, e.g., Anderson and Holt, 1997). The game involves n individuals each of whom, in random sequence order, submits one guess about which of two urns the experimenter has chosen: "Urn A" containing two red balls and one white ball or "Urn B" containing two white balls and one red ball. An individual wins $\$X$ for a correct guess and $\$Y < \$X$ for an incorrect guess. At the time of his or her guess, each individual privately observes one draw (with replacement) from the chosen urn and all previously submitted *guesses*, if any.[6] To complete the description, it is common knowledge that the experimenter chooses among the urns with equal probability, so that individuals' common prior beliefs about the likelihood of Urn A should be $\frac{1}{2}$.

This task was chosen for three main reasons. First and foremost, it provides each player with a lot of information about his or her own performance as well as pecuniary incentives to pay attention to, and form beliefs about, others' performances *without* making such comparisons the main focus of the task.[7] Secondly, the task provides an objective performance measure: number of correct guesses.[8] And thirdly, the urn-guessing task is

[6] For example, an individual guessing third would observe one private draw from the urn (with replacement) plus the *guesses*—but not the private draws—of the first- and second-guesser. The individual chosen to guess first in the sequence would observe only his or her private draw from the urn.

[7] This is the main strength of this task over other, more standard, tasks, such as IQ tests or puzzle solving, where the opportunity for learning about others' performance does not naturally arise and where introducing such learning opportunities would run the risk of demand effects—particularly after randomly disadvantaging one group.

[8] It is true that this measure certainly includes some noise due to being correct through luck as well as skill — as would many tests or tasks I could have chosen. In particular, one might be concerned that those who guess later are more likely to be part of an informational cascade and that this fact undermines the interpretation of total correct guesses as a performance measure. There are two factors which partially ameliorate this concern: i) cascades in this game are not irrational—entering into a cascade (or not) requires analytical ability; ii) sequence orders and groups were randomly re-determined across rounds so that, what-

plausibly ability-intensive: individuals must combine information from their private draw together with information derived from observing the choices of individuals guessing before them to arrive at an accurate assessment of the likelihood that experimenters chose Urn A. Of course, like many tasks that could have been chosen, performance here will depend on ability *and* effort—one seldom succeeds at anything without trying. However, a lot of effort with no analytical ability will not produce performance better than random chance in this game, while not much effort but a lot of analytical ability will produce superior performance. This is the sense in which the task is ability-intensive.

To get a feel for the type of reasoning required to perform well in this game, let us briefly consider individuals' optimal guesses.[9] To simplify matters, assume that an individual's guess equals his or her private draw whenever the individual believes both urns are equally likely.[10] The first guesser should guess Urn A (B) upon observing a red (white) draw since the posterior probability of Urn A (B) is now $\frac{2}{3}$. The second guesser, irrespective of the first-mover's guess in period 1, should also guess Urn A conditional on privately observing a red ball drawn and Urn B otherwise.[11] Thus, the first two guesses are perfectly informative of the first two draws. For guessers $n > 2$, simply note that: i) if the history of guesses in period $1, \ldots, n-1$ is such that the number of Urn A (B) guesses exceeds the number of Urn B (A) guesses by at least 2, then guesser n and all subsequent guessers optimally submit a guess of Urn A (B) regardless of their privately observed draws; ii) for all other histories, guesser n's optimal guess is Urn A (B) conditional on a private draw of red (white). Thus, guessing correctly in this game requires a bit of analytical ability, but not an unreasonably

ever effect there is from moving late in a sequence, it is unlikely that an individual is consistently assigned a late sequence order.

[9] For brevity's sake, I only sketch the logic here. For more detail, see the classical references in the huge literature on social learning using this game: Bikhchandani, Hirshleifer and Welch (1992); Anderson and Holt (1997).

[10] That is to say, they guess Urn A (B) conditional on a private draw of red (white) whenever their posterior belief—incorporating all available information including their private draw—suggests that Urn A and Urn B are equally likely. Relaxing this assumption changes how unbalanced the history of guesses needs to be in order to trigger a cascade, but does not add much intuition.

[11] To see this note that if 2's private draw is red (white) and 1's guess was Urn B (A), then 2's posterior is $\frac{1}{2}$ since it can be inferred that 1's private draw was white (red). In this case, by assumption, 2 goes with his private information. On the other hand, if 2's private draw is red (white) and 1's guess was Urn A (B), then 2 has essentially observed two red (white) balls being drawn and should guess Urn A (B). In all cases, 2 guesses Urn A (B) whenever 2's private draw is red (white).

large facility. Performance also obviously depends on a bit of luck (guessing sequence order) and effort (paying attention), but a plausible assertion is that performance is more sensitive to ability than effort.[12]

All participants in Experiment 1 played ten separate rounds of the analytical ability task. Before each round, the experimenter randomly privately chose either Urn A or Urn B. Participants were then randomly (re-)divided into groups of 8.[13] Within each group, participants were randomly assigned a sequence order $n = 1, \ldots, 8$ so that neither the specific individuals in each group, nor the sequence order a specific individual was assigned, persisted over rounds. At the end of each round, after all members of all groups had finished submitting their guesses, the chosen urn was revealed. After all ten rounds were completed, one round of the analytical ability task was randomly chosen to count towards participants' earnings.

3.1.3 Eliciting relative performance beliefs

In each of the three treatments, after all ten rounds of urn-guessing were completed each participant was asked to estimate his or her relative performance. Participants were not aware during task performance phase that there would be a subsequent beliefs elicitation phase of the experiment. In this phase, in all treatments (PI, C-HP, C-LP), participants were asked: "Compared to others, how accurate were your urn predictions?" Each participant responded by selecting one of two possible answers, appearing in random order: "top 50 percent"; or "bottom 50 percent." Each participant earned one additional dollar for a correct answer to this question, and no additional money for an incorrect response.

[12] A potential confound to this assertion is the use of a counting heuristic—simply counting the number of previous Urn A and Urn B guesses agreeing with the majority opinion—which also performs well. It is not clear how one arrives at this heuristic from the set of all possible heuristics, but it seems reasonable that some might. Partly to address this concern, additional experiments (Experiments 2 and 3, detailed in later sections) were conducted using more traditional measures of ability: questions from the logical reasoning section of the LSAT (Experiment 2) and Raven's Progressive Matrices (Experiment 3)—a widely-used measure of general intelligence.

[13] To avoid the need to have the number of participants in each session divisible by 8—a quite severe restriction which would have resulted in sending many participants home with just a show-up fee—as many 8-person groups of participants as possible were formed with the remaining participants forming a group of less than eight. Each session consequently featured at most one group with less than eight participants.

Eliciting beliefs in this way maintains incentive compatibility under weak assumptions on preferences while maintaining simplicity. Whereas many widely used methods to elicit beliefs are either quite complex, confusing, or require strong assumptions such as risk-neutrality (see, e.g., Schlag and Van der Weele, 2009), a sufficient assumption for the simple procedure used here to provide proper incentives for truthful revelation of beliefs is that preferences obey the axioms of expected utility. Intuitively, participants are faced with a choice between two assets which pay one dollar in different states of the world. Normalizing the utility of no additional earnings to $u(w+0)=0$, the expected utility of choosing "Top 50%" is simply $Prob(Performance\ above\ median) \times u(w+1)$. Similarly, the expected utility of choosing "Bottom 50%" is $Prob(Performance\ below\ median) \times u(w+1)$. Since $u(w+1)$ is a factor common factor to both of these expected utilities, the preferred choice is the asset whose payoff-relevant state the decision-maker (subjectively) considers most likely.[14] In light of the coarse hypotheses of the current inquiry, I found the potential benefits of using a mechanism capable of eliciting beliefs more precisely did not justify the added complexity or stronger assumptions required.

A more specific concern one may have about how beliefs are elicited here is the lack of an "exactly median" performance belief option. There are two reasons I omitted this option. First of all, given the other two options an expected utility maximizing individual will never strictly prefer such an option: since "exactly median" is a subset of both of the events "top 50%" and "bottom 50%," choosing either of these latter two options always yields weakly higher expected utility. An individual might in some circumstances weakly prefer reporting an exactly median belief—e.g., when he or she is *certain* his or her performance was exactly median.[15] However, given the wealth of information participants receive on their own and others' performances during the course of the experiment, this case is *a priori* unlikely. Secondly, previous research suggests such "middle" options may be focal and chosen *in*

[14] If the individual believes these two events are equally likely—e.g., because s/he is certain his/her performance is exactly median—then the two options yield the same expected utility, making the individual indifferent between the two options. This latter case should add noise, but not a bias, to the elicited beliefs.

[15] In this special case, the expected utility of the other two options would be the same as the exactly-median option.

spite of proper incentives for truthful revelation of beliefs or preferences (e.g., Harrison and Rütstrom, 2008). Balancing these two concerns suggested leaving out a middle option. I realize, however, that omitting an exactly-median option comes at the cost of additional noise: people who are certain they performed exactly at the median will be indifferent between the two available options as they are simultaneously in the both the top and bottom 50 percent of performers.[16]

Another minor point which needs addressing is hedging. There is typically a concern when paying for beliefs and performance that individuals might try to use the belief elicitation incentives to hedge against a bad outcome—i.e., one might report doing poorly to move money into this bad state, reducing the expected variation in their earnings. However, since each individual knew his or her own performance—and hence earnings—at the time they answered the performance belief question, there is no literal scope for literal hedging here.

3.2 Experiment 1 Results

All together, eight sessions were conducted and 152 individuals participated. The PI treatment constituted four of these sessions. Two sessions each of the C-LP and C-HP treatments were conducted. In total, there were 83 participants in the PI treatment, 36 in the C-LP treatment and 33 in the C-HP treatment. This information is summarized in Table 1.

As predicted, task performance did not vary substantially by pay group (Table 2). Across all treatments and pay groups, participants guessed the correct urn about two-thirds of the time. Standard parametric and non-parametric tests fail to reject the null hypothesis of equal performance.[17] If anything, the *low*-pay group performed slightly better.

[16] For those unconvinced about the prudence of omitting an exactly-median option, Experiments 2 and 3 (detailed in later sections) provide participants with such a middle option. It is included there because participants learn nothing about their relative performance, making it much more likely *a priori* that participants truly hold an exactly-median relative performance belief.

[17] For example, Wilcoxon rank-sum test across pay groups within PI yields a test statistic with $p = 0.710$; Wilcoxon rank-sum across C-HP and C-LP yields a test statistic with $p = 0.936$; a chi-squared test across pay groups within PI data yields $p = 0.400$ and $p = 0.170$ across C-HP and C-LP data. Moreover, the null hypothesis that the distributions of the number of correct guesses were the same in the PI sessions and the pooled C-HP and C-LP data cannot be rejected: Wilcoxon rank-sum test: $p = 0.595$; Chi-squared test:

Participants' beliefs were a different story, however. Consider first only raw means, without inserting any controls. PI participants assigned to the high-pay group were significantly more likely to report believing their performance ranked in the top half. Fully 85 percent of high-pay PI participants believed they were in the top half, while only 60 percent of low-pay participants put themselves in this top category ($\chi^2(1) = 6.9165, p = 0.009$).[18] While striking, from this first comparison it is not clear whether high relative pay enhanced participants' performance beliefs, low pay undermined beliefs, or both. Hypothesis 1 suggests both. For evidence on this dual prediction, I compare relative performance beliefs in the context of inequality (PI treatment) to the equivalent pay structure without inequality (C-HP or C-LP). The results support both aspects of Hyopthesis 1—i.e., that inequality undermines the confidence of the disadvantaged and boosts the confidence of the advantaged. The proportion of high-pay (low-pay) PI participants believing their performance ranked in the top half was higher (lower) in the PI treatment than in the pooled control treatments (C-HP (76%) and C-LP (78%)) where pay had no effect on relative performance beliefs.

Moving beyond raw means, estimating a formal model allows me to control for obvious other factors which could explain the belief patterns. In Table 3 I present marginal effects from an estimated probit model using as the dependent variable whether the participant reported believing his or her performance ranked above the median. In Column 1 of Table 3, I consider only observations from the PI treatment. The main explanatory variable is a dummy for whether a participant was assigned to the high-pay group (*High pay*). I also include controls for the relevant information participants had at the time their beliefs were elicited: their *actual* performance (*Proportion Correct*); others' performances they

$p = 0.564$.

[18] Recall that after the urn-guessing phase of the experiment was over each subject was asked the following question: "compared to others, how accurate were your urn predictions?" Their answers could be either "top 50 percent" or "bottom 50 percent." One minor technical issue that must be noted is that the median number of correct guesses varies depending on how widely one interprets the reference group "others." The narrowest definition I considered was defined with respect to subjects within each session. I also considered two slightly broader definitions that turned out to be less problematic as they coincided at a value of 60% of guesses correct. These broader medians incorporated either all guesses within each version, or all guesses across all versions. Within the control versions of the experiment (C-LP and C-HP) these distinctions were unnecessary as the various definitions of the median coincided. None of the patterns change qualitatively by using different definitions of the median.

had observed during the course of the experiment (*Proportion seen correct*). An additional dummy variable is inserted taking the value of one if the participant actually performed above the median (*Above median performance*) as a specification check. Finally, since performance may also be affected by the sequence order in which an individual submits his or her guess, I include a control for the average guessing sequence order an individual was assigned (*Average guessing order*). The data support Hypothesis 1: the coefficient on *High pay* is highly statistically significant and the marginal effect is large in magnitude. Evaluated at the mean of all other controls, being *relatively* well paid leads to an approximately 30 percentage point increase in the probability of believing one's performance ranked above the median. This represents about 40 percent of the sample mean—overall, 72.2% of PI participants reported an above-median performance belief.

Columns 2 and 3 of Table 3 provide additional evidence that inequality is responsible for the difference in relative performance beliefs. Column 2 reports marginal effects from an estimated probit model, restricting attention to observations *not* coming from the PI treatment. The non-significant and near-zero coefficient on *High pay* indicates that the pay levels themselves do not lead to differences in performance beliefs. Beliefs are not completely unfounded, however, as in the pooled C-HP and C-LP data actual *absolute* performance has explanatory power. Finally, Column 3 pools all data and estimates a probit model similar to the previous two columns but now inserting controls for whether the observation comes from the PI treatment (*PI treament*) and an interaction between the PI treatment dummy and pay level (*(High pay)*×*(PI)*). The negative and highly significant coefficient on *PI treatment* implies that being assigned to the low-pay group in the context of salient inequality undermines relative performance beliefs substantially (by about 15 percentage points, evaluating other controls at their means) as compared to the same pay structure without salient inequality. The positive and significant coefficient on (*(High pay)*×*(PI)*), together with the essentially zero coefficient on *High pay*, suggest that high pay substantially enhances relative performance beliefs in the context of salient inequality relative to the same pay level implemented outside of the context of salient inequality.

In summary, the results from Experiment 1 support Hypothesis 1: salient inequality affects relative performance beliefs by undermining the confidence of those disadvantaged by inequality and boosting the confidence of those on the comfortable side of inequality. However, Experiment 1 leaves unanswered one crucial question: does inequality primarily affect beliefs about ability or effort? This is an important distinction. If pay inequality operates on effort beliefs primarily, the effects of inequality may be transient and peculiar to *retrospective* performance beliefs. On the other hand, if inequality primarily affects beliefs about ability—an ostensibly stable trait—the belief patterns may be persistent enough to color subsequent decisions. Given the ability-intensive nature of the task involved, I would like to claim the latter. However, an obvious alternative explanation for the data so far is that pay inequality demoralized the lower paid group and that demoralization, or bad mood more generally, lowers effort provision. While there is no evidence that low pay actually lowered effort provision—the low-pay group performed no worse, and if anything, better—participants may have *believed* that mood affects performance which could have possibly affected beliefs. In light of this alternative explanation, Experiments 2 and 3 were designed to directly address the "effort vs. ability" question as well as provide some reassurance of the robustness of Experiment 1 results. Moreover, Experiment 3 provides a direct test of whether past inequality affects subsequent willingness to compete—i.e., Hypothesis 2. Let us first consider Experiment 2.

4 Experiment 2

To shed light on whether inequality affects relative performance beliefs through anticipated effort reduction (demoralization) or, rather, whether inequality affects beliefs about relative ability, Experiment 2 features a 2×2 design in which pay levels and the ability-intensiveness of the task involved are indepently randomly varied. If inequality primarily affects beliefs about relative effort provision, one would expect beliefs to be affected by inequality in the effort-intensive task—perhaps even moreso than in the ability-intensive task. On the other hand, if ability beliefs are primarily affected by inequality, one would expect a strong effect of inequality in the ability-intensive task but not the effort-intensive task.

To facilitate this comparison, the two tasks were chosen to be transparently primarily about effort—simple letter-counting—or about ability—logical reasoning questions taken from an LSAT. Of course, performance on any task is always a combination of effort and ability. However, the two tasks were chosen to lie on opposite ends of the spectrum in terms of the ratio of effort to ability required to perform well. As in Experiment 1, pay levels were constructed in a way that implemented severe and salient inequality while keeping marginal performance incentives identical. High-pay (HP) participants earned \$2 (\$1) for each correct (incorrect) answer. Low-pay (LP) participants earned \$1 (\$0) for each correct (incorrect) answer. Differently from Experiment 1, here participants completed their tasks in isolation—participants did not interact in any way, nor did they receive any information on others' performance.[19] After each participant completed his or her task, relative performance beliefs were elicited in an incentive-compatible manner. Participants were unaware of the belief elicitation section of the experiment while completing their task.

4.1 Design and procedures

Experiment 2 was conducted on-line using the service Mechanical Turk.[20] Participants were recruited by placing a standard request for work (a HIT in Mechanical Turk terminology). The ad specified that participants would be asked to complete a short task and would be paid a (standard on Mechanical Turk) fixed wage of \$0.05 USD for completing the task and that, in addition, 10 percent of those who completed the task would be chosen to be paid substantially more money depending on their decisions during the task. Participants were restricted to workers residing in the U.S. or in India since, at the time Experiment 2 was conducted, these were the only two locations where mechanical turk workers could choose to be paid in local currency rather than Amazon.com credits. Each participant could complete

[19] It is this last feature which facilitated conducting the experiment on-line, allowing access to a different subject pool than the typical university population used in Experiment 1.

[20] Mechanical Turk (https://www.mturk.com) is an on-line labor market provided by Amazon.com with a focus on small, yet labor-intensive, tasks such as proofreading.

the task only once.[21]

Upon choosing to participate, participants were informed that their task was to answer a set of five questions. They were informed that there would be two pay scales used to determine potential additional earnings from the experiment—a high pay scale and a low pay scale—and that each of these scales was equally likely. Those randomly assigned to the high pay scale would earn $2 USD for each correct answer and $1 USD for each incorrect answer while the low pay scale entailed $1 USD for each correct answer and $0 USD for each incorrect answer. As in Experiment 1, these pay scales keep marginal incentives fixed while implementing severe inequality. After learning which pay scale he or she had been assigned, each participant was presented with a set of five questions to answer.

Participants were randomly assigned exactly one of two possible 5-question tasks: an effort-intensive (EI) task or an ability-intensive (AI) task. Each participant knew only about the task he or she was assigned—AI or EI—and did not know that other participants were potentially assigned a different task. The EI task consisted of five separate 30-character strings comprised solely of the letters "a" and "b." For each string, separately, participants had to count how many times the letter "a" appeared. The strings were randomly pre-generated so that, while each of the five strings was different, each participant viewed exactly the same 5 strings.[22] In the AI task, participants faced a five-part question taken from the the logical reasoning section of an LSAT exam and slightly modified to prevent "google-ing" the answers.

After each participant finished answering his or her series of five questions, but before knowing their actual absolute or relative performance, participants were asked how their number of correct responses compared to others. It was explained that a correct answer in this belief-elicitation section would contribute an additional $1 USD to a participant's potential experimental earnings, while an incorrect answer would contribute nothing.[23]

[21] The Mechanical Turk system assigns unique IDs to each potential registered worker which permits such restrictions while preserving anonymity.

[22] An example string: "bbaabaaaabaabbaaababababbbabbaab." The strings were presented as images, rather than text, to prevent simply cutting-and-pasting into a letter-counting program.

[23] As is perhaps unavoidable when participants are paid for both performance and the accuracy of their beliefs about performance, one could be concerned about whether participants report their true beliefs or

Valid responses consisted of three mutually-exlusive options: i) "More than half of other participants will answer at least as many questions correctly as I did"; ii) "Less than half of other participants will answer at least as many questions correctly as I did"; iii) "Exactly half of other participants will answer at least as many questions correctly as I did." These three options correspond to "worse than median," "better than median," and "equal to median" performance beliefs, respectively, which were presented to participants in random order.[24]

The 2×2 design yields 4 separate treatments: (AI, HP); (AI, LP); (EI, HP); and (EI, LP). If the primary effect of inequality is to alter beliefs about effort provision through, e.g., demoralization or expected demoralization, then one would expect beliefs to differ between (EI, HP) and (EI, LP). However, if the effect of inequality operates primarily through ability beliefs, one would expect no variation in beliefs across (EI, HP) and (EI, LP) but significant variation in beliefs across (AI, HP) and (AI, LP).

In total, about 300 individuals took part in Experiment 2. Responses were collected over 14 days in two countries, with each day-country pair constituting a session. Participants had 30 minutes from the time they accepted the work request on Mechanical Turk to complete the task. Participation was roughly equally divided between India and the U.S. At the time of the experiment, the exchange rate was about 44 rupees to the US dollar, providing substantial variation in stake size by country.

4.2 Experiment 2 Results

As in Experiment 1, pay level did not affect actual performance in either treatment. On average, (AI, HP) participants answered 1.43 questions correctly vs. 1.69 in (AI, LP) ($\chi^2(4) : p = 0.649$) out of 5 possible. Among (EI, -) participants, individuals assigned high

whether reported beliefs reflect a hedging motive. This concern is partially allayed here by noticing that the belief question of interest is about *relative* performance while performance incentives were based on a participant's absolute performance. Therefore, the belief question provides no direct hedge against poor performance.

[24]This belief elicitation mechanism provides proper incentives for truthful belief revelation by the same arguments used to justify the mechanism in Experiment 1. Here, a middle option is provided: such a belief cannot be exluded *a priori* on reasonableness grounds since participants receive no information on others' performance. Its focal nature may still be of concern.

pay answered 4.11 questions correctly, on average, while the analagous number in (EI, LP) was 4.16 ($\chi^2(4) : p = 0.466$).[25] These patterns are summarized in Table 4.

However, there was significant variation in relative performance beliefs by pay level—but only on the ability-intensive task. To show this formally, I re-order participants' reported beliefs to form a 3-category variable increasing in relative performance beliefs. This variable takes the value 0 if the participant reported a below-median performance belief, 1 for an exactly-median performance belief and 2 if the respondent reported believing his or her performance to be above median. I use this 3-category relative performance belief measure as the dependent variable in two ordered probit model estimates, reported in the left panel of Table 5. The first column of Table 5 restricts attention to the participants assigned the ability-intensive task (AI, HP; AI, LP); the second column reports the same model restricted to participants assigned the effort-intensive task (EI, HP; EI, LP). The main explanatory variable is an indicator for pay level (*High pay*), taking the value of one if the participant was assigned to the high pay group. While participants did not know their actual performance when submitting their relative performance beliefs, they may have formed beliefs about this value so that I include this variable (*Proportion correct*) as a specification check. Finally, I control for any country-level features which may tend to affect reported beliefs in general by inserting an indicator for whether the participant resided in the US (excluded category: India).[26] Comparing results across the first two columns of Table 5 it is evident that pay level statistically significantly affected relative performance beliefs on the abilitiy-intensive task, but had no discernible effect on relative performance beliefs on the effort-intensive task.

Because coefficients from ordered probit estimates are difficult to interpret, in the right panel of Table 5 I report marginal effects from an estimated probit model using as the dependent variable an indicator for whether the participant reported an above-median performance belief—i.e., when the 3-category measure takes its highest value. From column

[25]Obviously, performance did vary across tasks. This critique is addressed with Experiment 3, below.

[26]As there is some ambiguity about the correct level at which to cluster standard errors here, estimates were also conducted clustering instead by country alone, or by date alone. Significance patterns did not change from the reported results which reflect clustering standard errors by "session."

3 of Table 5 we see that being assigned a high pay level has a substantial impact on the propensity to report an above-median performance belief when the task is ability intensive but, again, no impact when the task is effort-intensive. Participants in the (AI, HP) treatment were about 10 percentage points more likely to report an above-median performance belief than their (AI, LP) counterparts. This represents about one-quarter of the sample mean of this variable in the pooled (AI, HP), (AI, LP) data, which is comparable in magnitude to the effect of *actually* answering one more question correctly ($\approx$ 12 percentage points).[27]

Wrapping up, the patterns in Experiment 2 are consistent with the story that inequality affects relative ability beliefs rather than beliefs about relative effort provision. An important concern with this assertion is the possibility that the variation in belief patterns across effort- and ability-intensive tasks in my data is due largely, or solely, to the fact that *actual* performance was quite different across these tasks, even if constant across pay levels within tasks.[28] A related critique is that the distribution of performance is unneccessarily truncated, again reducing the scope for beliefs to vary, because there are only five questions on each task.

I address both of these concerns with Experiment 3, to which I now turn.

5 Experiment 3

Experiments 1 and 2 provide evidence that mere inequality affects relative performance beliefs by affecting ability beliefs. In Experiment 3, I provide further evidence of the effect of inequality on relative ability beliefs and, moreover, test directly the conjecture that mere inequality has a significant impact on willingness to compete on a subsequent task (Hypothesis 2).

[27]The sample mean of the above-median performance belief indicator is 0.397 — i.e., about 40% of participants reported an above-median belief on the ability-intensive task.

[28]I, of course, designed the tasks in Experiment 2 to induce similar performance. As sometimes happens, however, participants' behavior surprised me in this respect.

5.1 Design and procedures

Experiment 3 was again conducted on-line using participants recuited via Mechanical Turk. In order to attract a lot of participants, the fixed fee paid for completing the experiment was \$1—high by Mechanical Turk standards. As in the previous experiment, participants were informed that in addition to this fixed fee there was a 10% chance of being chosen to be paid their otherwise-hypothetical earnings from the experiment (for details, see the Instructions Appendix). In total, about 300 individuals participated in the experiment. Differently from Experiment 2, participation here was restricted to individuals currently residing in the U.S. and sessions were conducted on only two dates in order to keep incentives uniform and minimize the potential for word-of-mouth communication across subjects.

Experiment 3 is comprised of three phases. Before the first phase begins, participants are instructed that only one phase will be randomly chosen to count towards their potential experimental earnings.

The first phase features a 2×2 design that, apart from involving different effort- and ability-intensive tasks, is identical to Experiment 2. Participants are randomly assigned one of two potential pay structures as well as one of two possible tasks: ability- or effort-intensive. The two possible pay structures are the same as those used in Experiment 2: high-pay (\$2 for each correct answer, \$1 for each incorrect answer); or low pay (\$1 if correct, \$0 if incorrect). As in Experiment 2, participants are made aware of both possible pay structures, in order to make inequality salient, and that each pay structure is equally likely, but are not made aware of the task they are not assigned. The ability-intensive task consists of ten questions from a Raven's Progressive Matrices booklet. This test is widely recognized as a culture-free measure of general intelligence and consists of guessing which of eight pieces best completes a picture. The effort-intensive task also features ten separate, multiple choice, questions. Each question here presents participants with a different 200-character string of upper and lower case letters and asks how many times a particular (case-sensitive) letter appears in the string.[29] Both tasks are multiple choice, with eight

[29] As in Experiment 2, the string of characters is presented as a graphics file in order to make it difficult to

possible responses for each question.

In phase two of Experiment 3, relative performance beliefs are elicited using the same 3-category question as in Experiment 2. Incentives are ramped up compared to the previous experiment: a correct belief now pays \$20, while an incorrect belief still pays nothing. Higher incentives help address one concern with belief elicitation in the previous two experiments: weak truth-telling incentives. Misreporting one's beliefs is now quite costly, particularly since it may determine a participant's entire potential experimental earnings.

In the final phase—phase 3—participants are told they must now complete another task similar to the task they performed in phase one. Before beginning the task, however, participants choose their pay structure. The choice is between: i) a piece-rate pay structure which is in-between the previous two pay structures: \$1.50 (\$0) for each correct (incorrect) answer; or ii) tournament incentives.[30] Tournament incentives are implemented as follows. After completing the phase 3 task, participant i's score (s_i) will be compared with the score of one other, randomly selected, participant (s_j); If $s_i \geq s_j (s_i < s_j)$, participant i earns \$30 (\$0). The parameters of the incentive structures were chosen to ensure a substantial fraction of participants chose each option: a risk neutral individual who believes s/he will answer 8 questions correctly and that this performance will place below the median with a 60% chance is indifferent between the two options. To minimize the possibility that extraneous considerations such as other-regarding preferences affect the choice of pay structure, it is explained that one participant's choice of pay structure in no way affects any other participant's earnings.[31]

simply cut-and-paste the string into a text editor, which would otherwise be one way to cheat on the task.

[30] Neutral wording is, of course, used to describe these two options.

[31] Another potential concern is that placing the incentive choice decision after the phase with retrospective beliefs elicitation may exaggerate the effect of beliefs on competitiveness. For example, by making beliefs salient, through a preference for consistency (Eyster, 2002; Falk and Zimmerman, 2011) the act of stating an above-median (below-median) performance belief may *cause* participants to choose tournament (piece-rate) incentives. An alternative design choice would have been to move the belief elicitation phase to the very end of the experiment, but then a similar argument could be made about pay structure choice coloring performance beliefs. Consequently, I address this concern *ex-post* by looking at the data. Specificially, if the act of stating above-median (below-median) beliefs *causes* individuals to choose (shy away from) competition, we should find this effect irrespective of whether the inequality manipulation altered beliefs. I therefore restrict attention to the effort-intensive task and estimate probit models using tournament incentive choice as the binary dependent variable and stated performance beliefs as the main explanatory variable. I control for beliefs using either a set of dummies pertaining to the 3-category performance belief measure

Finally, after participants completed the experiment they were asked to provide some demographic details about themselves. I will use these self-reported demographics as controls where appropriate in the formal analysis of the results, below. The outcomes of interest in Experiment 3 are: i) retrospective relative performance beliefs related to the initial task in phase one; and ii) the choice of pay structure for the phase 3 task. Let us now turn to the results.

5.2 Results

As desired, actual performance in phase one is quite similar across tasks. Participants assigned the ability-intensive task answered 8.25 questions correctly, on average, while the analogous figure on the effort-intensive task was only slightly better at 8.74 ($\chi^2(8) : p = 0.115$).[32] As in the previous two experiments, actual task performance did not vary significantly across pay levels and, if anything, low pay induced slightly better performance.[33] Actual phase one task performance is summarized in Table 6.

To examine the effects of inequality on relative performance beliefs, I perform the same basic estimation of models of beliefs that I did for Experiment 2 using the Experiment 3 data (Table 7). These new estimates, however, additionally incorporate demographic controls to lessen the impact of omitted variables. I also add a full set of US time zone dummy variables to serve as both a rough control for within-US regional-level differences in beliefs, as well as a control for familiarity with the state lottery used to select participants to be paid their hypothetical earnings. Reassuringly, the estimates reported in Table 7 are quite

or one dummy for a stated above-median performance beliefs. For both the 3-category belief measure and the above-median belief indicator, I estimate: i) a minimal model, controlling only for beliefs; ii) an intermediate model including beliefs and all available demographics; and iii) a maximal model including beliefs and demographics as well as pay level and actual performance. I find no significant evidence in any of these estimates that stated beliefs affect the probability of choosing tournament incentives: the estimated effect of beliefs is always nearly zero and non-significant. Moreover, the coefficients obtained from estimating these models separately for males and females provide neither qualitative nor quantitative support for this channel. For example, stating an above-median performance belief has a small, non-significant, but negative impact on choosing tournament incentives for both males and females. These tables of results are omitted for brevity, but available upon request.

[32] There are only eight degrees of freedom because nobody answered fewer than two questions correctly.

[33] A chi-squared test of the difference in performance between high-pay and low-pay participants on the ability-intensive task yields a test statistic with $p = 0.300$. On the effort-intensive task, the chi-squared test statistic has $p = 0.405$.

similar to those found in the previous experiment. Again, pay level significantly affected relative performance beliefs in the ability-intensive task, but not in the effort-intensive task. Moreover, the marginal effect estimate of high pay on the probability of reporting an above-median belief is roughly equal to the estimate in the previous experiment (15% here, 10% there). The dual patterns in beliefs showing up in the data once again—a significant effect of inequality on beliefs about ability-based performance coupled with the absence of an effect on beliefs about effort-based performance—provides further evidence that beliefs about relative ability are affected by inequality. Moreoever, these belief patterns together the similarity of actual performance across tasks and pay levels demonstrate that alternative explanations—demoralization, mood or wealth effects—which depend on inequality affecting effort provision or beliefs about effort provision are unlikely to be the main driver of the results.

Before examining the tournament choice results, note that identification here relies on inequality moving beliefs enough to alter the attractiveness of the riskier, more uncertain, tournament pay option relative to the safer piece-rate pay option. Consider two individuals who have identical relative ability beliefs before being exposed to inequality, but who differ with respect to risk preferences. The more risk averse individual is more likely to prefer the safer, piece-rate, pay option before inequality is introduced and would require a larger ability-beliefs boost from advantageous inequality to switch to preferring tournament incentives. Consequently, systematic differences in risk preferences mitigate against finding an effect of inequality on tournament choice among more risk averse groups for any given magnitude of the effect of inequality on beliefs. As a common finding in previous research on risk preferences is that women tend to be more risk averse than men, when considering the effect of inequality on selection into competitive environments I split the data by gender and analyze men and women separately.[34] To compensate for the reduction in sample size for each task resulting from splitting the data in this way, I pool observations over the effort- and ability-intensive tasks and focus on the interaction between pay level and

[34] I cannot control for risk aversion, since I do not elicit a measure of this preference.

ability-intensiveness.

In Table 8 I report marginal effects from a probit model estimate using as the dependent variable an indicator for whether the participant chose tournament incentives. The first two columns restrict attention to male participants, while the third and fourth column use only female participants' data. Observations are pooled over tasks. Columns 1 and 3 present minimal specifications with no demographic controls, while columns 2 and 4 insert demographic and time zone controls.[35] Consider first male participants (columns 1-2). The estimates suggest, reassuringly, that prior inequality does not affect subsequent competitiveness in the case where beliefs were not affected: the underlying coefficient on *High pay* is non-significant. At the same time, consistent with Hypothesis 2, the interaction of previous inequality and being assigned an ability-intensive task—where high pay had a significant effect on beliefs—has a substantial effect on competitiveness. The marginal effects imply that high pay increases the probability of choosing tournament incentives over piece-rate pay by about 13 percentage points—roughly equivalent to the effect of answering 2.5 more questions correctly and almost identical in magnitude to the increase in the probability of reporting an above-median *retrospective* (15%) performance belief.

Among female participants there is no evidence of an effect of prior inequality on subsequent competitiveness (columns 3 and 4). This non-effect could obtain for several reasons. While I have no direct evidence on the true channel, let me take a moment to rule one out. An explanation which jumps to mind, particularly in the light of previous research, is that women in my experiment are just less competitive. If that were the case, however, one would expect women to be less likely to choose the tournament incentives than men even in the case where beliefs are not affected by inequality—i.e., when the task is effort-intensive. Looking into the data, the difference in this case is tiny: 54% of male participants and 52% of female participants choose tournament incentives over piece-rate pay ($\chi^2(1) : p = 0.739$). As food for thought, an alternative explanation may be that—for whatever reason—women's ability beliefs are less sensitive to environmental factors like recent unequal treatment. Among

[35] While participants were not informed of their score on the task in phase one, I include this variable here, as before, as a specification check.

other reasons, this could be because women are more inured to inequality, because they have more prior information on their own ability, or because they view outcomes as more dependent on luck than skill. To provide support for this conjecture, I estimate the effect of inequality on performance beliefs separately for male and female participants (Table 9). While the impact of inequality on participants' performance beliefs is qualitatively similar across genders, the estimates suggest the effect is quantitatively stronger among male participants.

6 Discussion

In this paper, I have presented results from three separate experiments documenting that mere inequality can exert considerable impact on relative ability beliefs (Hypothesis 1) and that these beliefs can color an individual's subsequent willingness to compete (Hypothesis 2). These patterns are consistent with the notion that Just World beliefs apply to one's own merit or ability and are precisely those necessary for initial inequality to become self-perpetuating. The fact that neither beliefs nor competitiveness are affected by unequal pay on effort-intensive tasks provides evidence against alternative explanations—mood, morale or more traditional wealth effects. Considered as a whole, the results from all three experiments reveal a novel mechanism with the potential to generate durable inequality.

This new mechanism differs fundamentally from existing mechanisms in that it requires no actual external discrimination or beliefs that external discrimination will occur. For example, two standard explanations for persistent inequality among ethnic groups in the US context are Arrow's (1972) statistical discrimination and Becker's (1971) taste for discrimination models. In the former, if employers believe one group to be less productive–say, group A—they pay members of group A less. Anticipating that their human capital investments will be less profitable, members of group A rationally obtain less human capital, making them less productive. Employers' initially discriminatory beliefs generate persistent discrimination in equilibrium. In the latter model, employers and other employees simply prefer to discriminate against members of group A. In both of these models, the chain of inequality can be broken by addressing external discrimination—either beliefs or preferences. In the current mechanism, a random shock adversely affecting one group may set off a process leading to persistent inequality even in the absence of employers' discriminatory beliefs or preferences. Individuals' are sufficiently adept at discriminating against themselves. More perniciously, while open and transparent ability-based competition for spots at prestigious colleges or firms is an oft-recommended antidote to statistical or preference-based discrimination, such competition may exacerbate inequality through the mechanism documented in this paper. Instead, an effective policy prescription here could be providing incentives and information to the demand-side (employers, colleges, etc.) to aid in the search for and recruitment of qualified applicants.

As a final note, one may wonder whether the patterns documented here in an admittedly artificial context extend to real-world decision making. While estimating the importance of such self-discrimination in economically consequential, real-life, decisions is beyond the scope of this paper, anecdotal evidence may stimulate such future research. One of the most important decisions of many individual's lives is the decision about which college to attend. The goal of a well-publicized multi-year campaign by several elite colleges, including Harvard, was to attract talented students from low-income backgrounds. In order to accomplish this, many colleges offered an essentially free education by committing to need-blind admissions and need-based grants. In an interview with National Public Radio, economist Caroline Hoxby summarized the stunning lack of success of this approach:[36]

> After Harvard offered what was, in essence, a free college education to students whose families earned under $40,000 a year, Hoxby says, 'the number of students whose families had income below that threshold changed by only about 15 students, and the class at Harvard is about 1,650 freshmen.' ... some college administers (sic) had confided to her that they had reluctantly come to the con-

[36]The transcript of the interview can be found here: http://www.npr.org/2013/01/09/168889785/elite-colleges-struggle-to-recruit-smart-low-income-kids.

> clusion that the pool of low-income students with top academic credentials was just limited, and there wasn't much they could do to change that.

Hoxby and Avery (2012) go on to debunk the myth of limited talent among low-income high school students, showing that talented low-income students are out there—they just do not apply to elite, highly selective, colleges. Instead, they generally apply to "...a few 'par' colleges, a few 'reach' colleges, and a couple of 'safety' schools...." Given the generous financial aid offered by highly selective colleges, this strategy results in the perverse outcome of low-income students paying more for a less prestigious degree. While only suggestive—many other explanations are plausible and are likely in play—since applying to and succeeding at highly selective colleges is the quintessential example of ability-based competition, the patterns documented in college application decisions are consistent with the mechanism documented here. If so, my results suggest that one admittedly controversial remedy to low-income students' apparently sub-optimal college application strategies might be to take college application decisions out of their hands.

References

[1] Afridi, F., Li, S., and Ren, Y. (2012), "Social identity and inequality: The impact of china's hukou system," *IZA Discussion Paper Number 6417.*

[2] Allport, G. W. (1954) *The Nature of Prejudice*, Addison-Wesley Publishing Co.

[3] Anderson, L. and Charles Holt (1997), "Information cascades in the laboratory," *The American Economic Review*, 87(5): 847–862.

[4] Arrow, Kenneth (1972), *Racial Discrimination in Economic Life*, chapter "Some Models of Race in the Labor Market." D.C. Heath, Lexington, MA.

[5] Becker, Gary S. (1971), *The Economics of Discrimination, 2nd ed.* University of Chicago Press, Chicago, Il.

[6] Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992), "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, 100(5):992–1026.

[7] Brosnan, Sarah F. and Frans B. M. de Waal (2003), "Monkeys reject unequal pay," *Nature*, 425: 297–299.

[8] Brosnan, Sarah F., Hillary C. Schiff and Frans B. M. de Waal (2005), "Tolerance for inequity may increase with social closeness in chimpanzees," *Proceedings of the Royal Society B-Biological Sciences*, 272: 253–258.

[9] Caprariello, P., Cuddy, A., and Fiske, S. (2009), "Social structure shapes cultural stereotypes and emotions: A causal test of the stereotype content model," *Group Processes and Intergroup Relations*, 12:147–155.

[10] Cuddy, Amy J. C., et al. (2009), "Stereotype content model across cultures: Towards universal similarities and some differences," *British Journal of Social Psychology*, 48:1–33.

[11] Eil, David and Justin M. Rao (2010), "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics*, 3: 114-38.

[12] Eyster, Erik (2002), "Rationalizing the Past: a Taste for Consistency." *mimeo, UC Berkeley*

[13] Falk, Armin and Florian Zimmerman (2011) "Preferences for Consistency," *IZA Discussion Paper*, n. 5840.

[14] Fiske, S.T., Amy J.C. Cuddy and Glick, P. (2007). "Universal dimensions of social cognition: warmth and competence," *Trends in Cognitive Sciences*, 11(2):77 – 83.

[15] Gneezy, Uri, Kenneth L. Leonard, and John A. List (2009), "Gender differences in competition: Evidence from a matrilineal and a patriarchal society," *Econometrica*, 77(5):1637–1664.

[16] Grossman, Zachary and David Owens (2012), "An unlucky feeling: Overconfidence and noisy feedback," *Journal of Economic Behavior and Organization*, 84:510–524.

[17] Harrison, G. and Rütstrom, E. E. (2008). "Risk aversion in the laboratory," In Cox, J. and Harrison, G., editors, *Risk Aversion in Experiments*, volume 12. Emerald.

[18] Hoff, Karla and Priyanka Pandey (2006), "Discrimination, social identity, and durable inequalities," *The American Economic Review*, 96:206–211.

[19] Hoxby, Caroline and Christopher Avery (2012), "The Missing 'One-Offs': The Hidden Supply of High-Achieving, Low Income Students," *NBER working paper* n. 18586.

[20] Jost, John T., Brett W. Pelham, Oliver Sheldon and Bilian Ni Sullivan (2003), "Social inequality and the reduction of ideological dissonance on behalf of the system: evidence of enhanced system justification among the disadvantaged." *European Journal of Social Psychology*, 33:13–36.

[21] Lerner, Melvin J. (1965), "Evaluation of performance as a function of performer's reward and attractiveness." *Journal of Personality and Social Psychology*, 1:355–360.

[22] Niederle, M. and Vesterlund, L. (2007), "Do women shy away from competition? do men compete too much?," *Quarterly Journal of Economics*, 122(3):1067–1101.

[23] Schlag, Karl and Joël; van der Weele (2009), "Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality," Working Paper, Universitat Pompeu Fabra.

[24] Steele, C. and J. Aronson (1995), "Stereotype threat and the intellectual test performance of the african-americans," *Journal of Personality and Social Psychology*, 69:797–811.

Tables

Table 1: Summary of Experiment 1 design

	PI treatment		Control	
	High-pay group	Low-pay group	C-HP	C-LP
Pay for correct guess	$4	$2	$4	$2
Pay for incorrect guess	$2	$0	$2	$0
Participants	41	42	33	36

Table 2: Actual Performance, Experiment 1

	PI treatment		Control	
	High-pay group	Low-pay group	C-HP	C-LP
Number of correct guesses	6.54	6.69	6.63	6.81
Std Err	(0.25)	(0.25)	(0.31)	(0.19)
Obs	41	42	33	36

Notes: [1] The table reports the average number of correct guesses across all 10 rounds of the urn-guessing task. The maximum possible value would be 10. **[2]** Standard errors appear in parentheses.

Table 3: Relative Performance Beliefs, Marginal Effects Estimates, Experiment 1

	Dependent Variable = Participant Believed Own Performance Was in Top 50%		
	(1) PI treatment only	(2) C-HP and C-LP pooled	(3) PI, C-HP and C-LP pooled
High pay	0.29***	0.04	0.01
	(0.07)	(0.08)	(0.08)
PI treatment			-0.15***
			(0.04)
(High pay) X (PI)			0.23***
			(0.07)
Proportion correct	0.85	1.18**	1.04**
	(0.54)	(0.49)	(0.43)
Proportion seen correct	-0.08	-0.57	-0.29
	(0.11)	(0.37)	(0.22)

Above median performance	0.19*	0.05	0.12
	(0.10)	(0.16)	(0.12)
Average guessing order	-0.00	0.09	0.05
	(0.07)	(0.08)	(0.06)
Pseudo-R^2	0.26	0.26	0.25
Observations	83	69	152

Notes: **[1]** Each column presents marginal effects estimates from a separate probit model. Column 1 uses only data from PI treatment sessions, column 2 uses only data from the pooled control sessions (C-HP and C-LP) while column 3 uses all data from all sessions. **[2]** The dependent variable in each column is the same: a dummy variable taking the value one whenever that the participant stated believing their performance was in the top 50% compared to other participants in the urn-guessing task, and taking the value zero otherwise. **[3]** In order to account for arbitrary within-session correlation of observations, robust standard errors are clustered by session (in parentheses). Significance levels do not change, however, if standard errors are not clustered. **[4]** *** significant at 1%, ** significant at 5%, * significant at 10%. **[5]** The independent variables are: "*High pay*" is a dummy taking the value of one if the participants' pay structure was $4 for a correct urn-guess and $2 for an incorrect guess, and 0 otherwise; "*PI treatment*" is an indicator variable taking the value one if the participant was assigned to the PI treatment, and 0 if the participant was assigned to either the C-HP or the C-LP treatment; "*Proportion correct*" is the proportion of the participant's 10 urn guesses that were correct; "*Proportion seen correct*" is the proportion of correct guesses made by others that the participant observed, averaged over all 10 rounds of the urn-guessing task; "*Above median performance*" is a dummy variable indicating whether the participants' performance in the urn-guessing task was actually (weakly) above the median—this variable was obviously unknown to the participant.; "*Average guessing order*" is the participant's average guessing sequence order over all 10 rounds of the urn-guessing task.

Table 4: Actual Performance, Experiment 2

	Ability Intensive Task		Effort Intensive Task	
	High-pay	Low-pay	High-pay	Low-pay
Number of correct answers	1.42	1.69	4.12	4.16
Std Error	(0.13)	(0.12)	(0.21)	(0.21)
Obs	91	83	77	70

Notes: [1] The table reports the average number of correct answers out of the 5 questions asked. [2] Standard errors appear in parentheses.

Table 5: Relative Performance Beliefs, Experiment 2

	(1)	(2)	(3)	(4)
	Using 3-Category Belief Measure		Using Above Median Performance Belief Dummy	
	Ability Intensive Task	Effort Intensive Task	Ability Intensive Task	Effort Intensive Task
High pay	0.26**	0.03	0.10**	0.03
	(0.12)	(0.29)	(0.05)	(0.08)
Proportion correct	1.14***	-0.36	0.62***	-0.10
	(0.40)	(0.39)	(0.18)	(0.13)
US (dummy)	-0.18	0.54***	-0.03	0.10
	(0.14)	(0.19)	(0.07)	(0.07)
Pseudo-R^2	0.03	0.03	0.06	0.03
Observations	174	147	174	147

Notes: [1] Robust standard errors, clustered by date-country pair, in parentheses. Date-country pairs serve as an analogue to sessions in this on-line environment. The number of clusters ranges from 14 to 17. **[2]** *** significant at 1%, ** significant at 5%, * significant at 10%. **[3]** Columns 1 and 2 present ordered probit estimates using as the dependent variable the 3-category relative performance belief measure described in the text, taking values 1 = believed performance was below median to 3 = believed performance was above median. **[4]** Columns 3 and 4 report marginal effects estimates from probit models using as the dependent variable a dummy taking the value of one if the participant reported believing they performed above the median (i.e., a value of 3 in the 3-category scheme). **[5]** The explanatory variables are: "*High pay*" = a dummy taking the value of one if the participants' pay structure was \$2 for each correct guess and \$1 for each incorrect guess; "*Proportion correct*" = the participant's actual proportion of correct guesses (i.e., $\frac{Number\ correct}{5}$). This variable is unknown to participant, of course; "*US*" is a dummy variable taking the value of one if the participant was located in the USA, and 0 if the participant's location was India.

Table 6: Actual Performance, Task 1, Experiment 3

	Ability Intensive		Effort Intensive	
	High-pay	Low-pay	High-pay	Low-pay
Number of correct answers	8.20	8.31	8.55	8.78
Std Error	(0.23)	(0.25)	(0.19)	(0.14)
Obs	64	68	76	79

Notes: [1] The table reports the average number of correct answers out of the 10 questions asked on the first task of Experiment 3. **[2]** Standard errors appear in parentheses.

Table 7: Relative Performance Beliefs, Experiment 3

	(1)	(2)	(3)	(4)
	3-Category Belief Measure		Above Median Performance Belief Dummy	
	Ability Intensive Task	Effort Intensive Task	Ability Intensive Task	Effort Intensive Task
High pay	0.47***	-0.08	0.15**	-0.03
	(0.17)	(0.19)	(0.07)	(0.07)
Proportion correct	0.59*	0.35	0.19	0.43*
	(0.30)	(0.44)	(0.16)	(0.26)
Income ≤ $30K	-0.37	0.16	-0.15	0.07
	(0.32)	(0.24)	(0.13)	(0.08)
$30K<Inc ≤ $70K	-0.27	0.10	-0.17	0.06
	(0.27)	(0.25)	(0.12)	(0.11)
Bachelor's degree	0.07	0.10	-0.01	0.11
	(0.15)	(0.25)	(0.07)	(0.09)
Age	0.00	-0.03	-0.01	-0.00
	(0.06)	(0.04)	(0.03)	(0.02)
Age^2	-0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Male	0.14	0.27	0.06	0.14*
	(0.19)	(0.20)	(0.11)	(0.07)
Time Zone Dummies	Y	Y	Y	Y
Pseudo-R^2	0.05	0.02	0.05	0.04
Observations	131	154	131	154

Notes: [1] Robust standard errors, clustered by date-hour pair, in parentheses. Hour is the hour of the day in which the participant completed the experiment and can take values from 0 to 24. Date-hour pairs serve as an analogue to sessions in this on-line environment. The number of clusters ranges from 15 to 16. **[2]** *** significant at 1%, ** significant at 5%, * significant at 10%. **[3]** Columns 1 and 2 present ordered probit estimates using as the dependent variable the 3-category relative performance belief measure described in the text, taking values 1 = believed performance was below median to 3 = believed performance was above median. **[4]** Columns 3 and 4 report marginal effects estimates from probit models using as the dependent variable a dummy taking the value of one if the participant reported believing they performed above the median (i.e., a value of 3 in the 3-category scheme). **[5]** The explanatory variables are: "*High pay*" = a dummy taking the value of one if the participants' pay structure was $2 for each correct guess and $1 for each incorrect guess; "*Proportion correct*" = the participant's actual proportion of correct guesses in task 1 (i.e., (i.e., $\frac{Number\ correct}{10}$). This variable is unknown to participant, of course; **[6]** Demographic controls are all self-reported and include: income category dummies based on participant's last-year's total income; "*Age*" and "*Age^2*" which are constructed from the participant's answer to the question "*in what year were you born?*"; and Gender. Coefficients and significance levels are robust to excluding demographics. **[7]** "*Mountain Time Zone*," "*Central Time Zone*" and "*Eastern Time Zone*" together constitute a full set of dummies for the US time zone in which the participant's ip address was located—the excluded category being "*Pacific Time Zone*."

Table 8: Selecting Tournament Incentives for Second Task, Experiment 3

Dependent variable = Participant Chose Tournament Pay for Second Task				
	(1)	(2)	(3)	(4)
	Male		Female	
High pay	-0.17	-0.18	0.01	-0.00
	(0.12)	(0.13)	(0.09)	(0.10)
Ability intensive	-0.13	-0.10	-0.10	-0.17
	(0.13)	(0.14)	(0.12)	(0.14)
(Ability intensive) X (High pay)	0.30***	0.28**	-0.08	-0.02
	(0.11)	(0.13)	(0.19)	(0.22)
Proportion correct in task 1	0.59**	0.51**	0.02	-0.02
	(0.26)	(0.26)	(0.23)	(0.18)
Income ≤ $30K		0.06		-0.06
		(0.16)		(0.11)
$30K<Income ≤ $70K		-0.08		-0.08
		(0.22)		(0.11)
Bachelor's degree		-0.03		0.20**
		(0.09)		(0.09)
Age		-0.01		-0.03
		(0.03)		(0.02)
Age^2		0.00		0.00
		(0.00)		(0.00)
Time Zone Dummies	Y	Y	Y	Y
Pseudo-R^2	0.09	0.11	0.02	0.05
Observations	125	124	162	161

Notes: **[1]** Each column presents marginal effects estimates from a probit model using as the dependent variable an indicator taking the value 1 whenever the participant chose tournament incentives to apply to the second task in Experiment 3, and taking the value 0 whenever the participant instead chose piece-rate pay for this task. **[2]** Columns 1 and 2 restrict attention to only male participants, while columns 3 and 4 report estimates using only observations from female participants. **[3]** Robust standard errors, clustered by date-hour pair, in parentheses. Hour is the hour of the day in which the participant completed the experiment and can take values from 0 to 24. Date-hour pairs serve as an analogue to sessions in this on-line environment. The number of clusters ranges from 15 to 16. **[4]** *** significant at 1%, ** significant at 5%, * significant at 10%. **[5]** The explanatory variables are: "*High pay*" = a dummy taking the value of one if the participants' pay structure *on task 1* was $2 for each correct guess and $1 for each incorrect guess; "*Ability intensive*" is a dummy taking the value one if the participant was assigned ability intensive tasks, and zero if the participant was assigned effort-intensive tasks; "*(Ability intensive) X (High pay)*" is an interaction term; "*Proportion correct on task 1*" = the participant's actual proportion of correct guesses on task 1 (i.e., (i.e., $\frac{Number\ correct\ on\ task\ 1}{10}$). This variable is unknown to participant, of course; **[6]** Demographic controls are all self-reported and include: income category dummies based on participant's last-year's total income; "*Age*" and "*Age^2*" which are constructed from the participant's answer to the question "*in what year were you born?*" **[7]** "*Mountain Time Zone*," "*Central Time Zone*" and "*Eastern Time Zone*" together constitute a full set of dummies for the US time zone in which the participant's ip address was located—the excluded category being "*Pacific Time Zone*."

Table 9: Beliefs by Gender, Experiment 3

Dependent variable = 3-Category Belief Measure				
	(1)	(2)	(3)	(4)
	Male		Female	
High pay	-0.21	-0.21	-0.02	0.03
	(0.31)	(0.28)	(0.23)	(0.23)
Ability intensive	-0.19	-0.19	0.17	0.26
	(0.24)	(0.26)	(0.23)	(0.31)
(Ability intensive) X (High pay)	0.73**	0.86***	0.42	0.34
	(0.30)	(0.29)	(0.27)	(0.31)
Proportion correct in task 1	0.19	0.16	0.65*	0.75**
	(0.45)	(0.46)	(0.37)	(0.30)
Additional controls				
Demographics	N	Y	N	Y
Time Zone Dummies	Y	Y	Y	Y
Pseudo-R^2	0.03	0.07	0.04	0.06
Observations	125	124	162	161

Notes: [1] Columns 1 and 2 present estimates from an ordered probit model using data only from male participants and using as the dependent variable the 3-category relative performance belief measure described in the text. Columns 3 and 4 estimate the same models, but restrict attention to data from female participants. **[2]** Robust standard errors, clustered by date-hour pair, in parentheses. Hour is the hour of the day in which the participant completed the experiment and can take values from 0 to 24. Date-hour pairs serve as an analogue to sessions in this on-line environment. The number of clusters ranges from 15 to 16. **[4]** *** significant at 1%, ** significant at 5%, * significant at 10%. **[5]** The explanatory variables are: "*High pay*" = a dummy taking the value of one if the participants' pay structure *on task 1* was \$2 for each correct guess and \$1 for each incorrect guess; "*Ability intensive*" is a dummy taking the value one if the participant was assigned ability intensive tasks, and zero if the participant was assigned effort-intensive tasks; "*(Ability intensive) X (High pay)*" is an interaction term; "*Proportion correct on task 1*" = the participant's actual proportion of correct guesses on task 1 (i.e., $\frac{Number\ correct\ on\ task\ 1}{10}$). This variable is unknown to participant, of course; **[6]** The coefficients for the demographic and time zone controls are omitted for brevity. These sets of controls are the same listed in the notes to Table 8, above.

Experiment Instructions Appendix

Experiment 1

Instructions

This is an experiment in the economics of decision making. Your earnings will be paid to you, privately, by check, at the end of the experiment.

In this experiment you will be asked to predict from which randomly chosen urn a ball was drawn. It is equally likely that urn A or urn B will be drawn. Urn A contains 2 red balls, and 1 white ball. Urn B contains 1 red ball and 2 white balls.

To help you determine which urn has been selected, each person will be allowed to see one ball, chosen at random, from the urn. The result of this draw will be your private information and should not be shared with the other participants. After each draw, we will return the ball to the container before making the next private draw. Each person will have one private draw, with the ball being replaced after each draw.

This is done on the computer. At the beginning of each period you will be randomly (re-) matched with from 1 to 7 other participants. Within each of these groupings, the experiment proceeds identically.

When it is your turn to see your draw, your computer screen will read. “Your draw is White" if the ball the computer has randomly drawn for you is white; and “Your draw is Red" if the ball the computer has randomly drawn for you is red. The order in which you see your draw is randomly determined every period.

After each person has seen his or her draw, each person will be asked to choose the letter of the urn (A or B) that he or she thinks is more likely to have been used. The person who was chosen to see his draw first will indicate his choice by either clicking on a button marked “urn A,” or on a button marked “urn B.” All his co-participants will then be able to see his choice at the bottom of their computer screen. The second person will then see her draw, and will be asked to indicate the letter of the urn that she thinks is more likely, and all of her co-participants will be able to see her choice in the last line of their windows. This process will be repeated until all remaining people have made decisions. Finally we will inform everyone of the urn that was actually used.

[PI treatment only]
The experiment will consist of many periods. Before the first period, half of the participants will be randomly assigned to a High-Paying (HP) group; and the remaining half will be assigned to a Low-Paying (LP) group. This group assignment will be the same in all periods. The urn, however, will be randomly re-chosen at the beginning of each period.

Given these group assignments, for each period your earnings are determined as follows:

- For those assigned to the HP group: if your decision matches the urn that was actually used, then you earn $4. Otherwise you earn $2.
- For those assigned to the LP group: if your decision matches the urn that was actually used, then you earn $2. Otherwise you earn $0.

[C-HP only]
The experiment will consist of many periods. The urn, however, will be randomly re-chosen at the beginning of each period.

For each period your earnings are determined as follows:

- if your decision matches the urn that was actually used, then you earn $4. Otherwise you earn $2.

[C-LP only]
The experiment will consist of many periods. The urn, however, will be randomly re-chosen at the beginning of each period.

For each period your earnings are determined as follows:

- if your decision matches the urn that was actually used, then you earn $2. Otherwise you earn $0.

Please do not talk with anyone during the experiment. We will insist that everyone remain silent until the end of the last period. If we observe you communicating with anyone else during the experiment we will ask you to leave without completing the experiment. It is very important that you do not open other windows or leave the page in front of you while the experiment is running.

[Typical urn-guessing screen, PI]

Your draw is: White

Which urn is more likely?

[] Urn A

[] Urn B

You are number 4 in a sequence of 8 participants.

<u>History of Other Participants' Predictions this Round:</u>

	First Person	Second Person	Third Person	You				
Prediction:	Urn A	Urn A	Urn A	?				
Group:	HP	LP	HP	HP				

[Typical urn-guessing screen, C-HP or C-LP]

Your draw is: White

Which urn is more likely?

[] Urn A

[] Urn B

You are number 4 in a sequence of 8 participants.

History of Other Participants' Predictions this Round:

	First Person	Second Person	Third Person	You				
Prediction:	Urn A	Urn A	Urn A	?				

[Typical urn-reveal screen]

The actual urn used was: [Urn A / Urn B]

Please click the button below to proceed to the next period.

[Beliefs elicitation screen 1]

The urn-guessing part of the experiment has concluded. Please answer the question below concerning your performance in the experiment. You will earn one additional dollar for a correct answer.

Compared to the other participants, how accurate were your urn predictions?

- o Top 50%
- o Bottom 50%

[Beliefs elicitation screen 2, (PI only)]

Which group got the highest percentage of urn predictions correct?

- o HP Group
- o LP Group

Experiment 2

[Screen 1]

There will be at least 100 people completing this survey before it expires in the next couple of days. In addition to a small fixed participation fee (currently $0.05 USD) you can potentially make a reasonable sum of money by completing this survey.

To determine your potential additional earnings from this survey, you will be assigned one of two possible pay scales. Each of the pay scales is equally likely.

The two possible pay scales are:

- High pay scale: $ 2 USD for each correct answer; $ 1 USD for each incorrect answer.

- Low pay scale: $ 1 USD for each correct answer; $ 0 USD for each incorrect answer.

10 percent of those who complete this survey will be randomly chosen by the computer to actually be paid their potential additional earnings. Every person who completes the survey has the same chance of being chosen (a 10% chance).

Those chosen will receive their additional earnings in the form of a bonus through their mechanical turk account on the day the hit expires---so, very soon.

You will find out which pay scale you have been assigned on the next page. Click continue.

[Screen 2]

[Participants saw exactly one of the two notifications below]

You have been assigned the high pay scale

You have been assigned the low pay scale

[Screen 3: Effort-intensive task; only seen if assigned to EI treatment]

Question 1: In total, how many times does the letter "a" appear in sequence of letters below?

aaababbababaaabbababbbbabaaaaa

Question 2: In total, how many times does the letter "a" appear in sequence of letters below?

babaaabaabababbabaabaabbabaabb

Question 3: In total, how many times does the letter "a" appear in sequence of letters below?

abaaababbababaabbbbaababbbabbb

Question 4: In total, how many times does the letter "a" appear in sequence of letters below?

bbaabaaaabaabbaaabababbbabbaab

Question 5: In total, how many times does the letter "a" appear in sequence of letters below?

abaababababaabbbbababaabbababbbb

[Screen 4: Ability-intensive task; only seen if assigned to AI treatment]

A school has exactly four dormitories that are to be fully occupied: Roanoke, Tuscany, Vinci and Wigglesworth. Each dormitory is divided into two parts: a north wing and a south wing. Each student is assigned to exactly one dormitory.

The rules used to assign students to dormitories are:

- Each wing contains either entirely male or entirely female students;
- Exactly 3 wings have male students;
- Roanoke North and Tuscany North have only females;
- Whenever one wing of a dormitory has male students, the other wing has female students;
- Whenever Vinci South has male students, Wigglesworth North also has male students.

Question 1: If females are assigned to Vinci South and Vinci North, then two other wings that are also assigned females could be:

( ) Roanoke North and Tuscany South
( ) Roanoke South and Wigglesworth South
( ) Roanoke South and Tuscany North
( ) Tuscany North and Wigglesworth South
( ) Tuscany South and Wigglesworth South

Question 2: Female students are never assigned to both:

( ) Roanoke South and Wigglesworth South
( ) Roanoke South and Tuscany South
( ) Roanoke South and Vinci North
( ) Tuscany South and Wigglesworth South
( ) Vinci North and Wigglesworth South

Question 3: If female students are assigned to Wigglesworth North, the female students must also be assigned to:

( ) Roanoke South
( ) Wigglesworth South
( ) Tuscany South
( ) Vinci South
( ) Vinci North

Question 4: If male students are assigned to Vinci South, which one of the following is a complete and accurate list of the wings that cannot be assigned male students?

( ) Roanoke North, Tuscany North
( ) Roanoke North, Tuscany North, Vinci North
( ) Roanoke North, Tuscany North, Wigglesworth South
( ) Roanoke North, Tuscany North, Vinci North, Wigglesworth South
( ) Roanoke North, Roanoke South, Tuscany North, Vinci North, Wigglesworth South

Question 5: If Tuscany South is assigned female students, then it could be true that female students are assigned to both:

( ) Roanoke South and Wigglesworth North
( ) Roanoke South and Wigglesworth South
( ) Vinci North and Wigglesworth North
( ) Vinci South and Wigglesworth South
( ) Vinci North and Vinci South

[Screen 5: Beliefs elicitation]

Everyone who is chosen to be paid their potential earnings will earn an additional $1.00 (USD) for each correct answer on this page. [Question order randomized]

Question: How many questions did you answer correctly in the previous section?

Question: Consider everybody that will take this survey in the next couple of days (before the HIT expires on mturk). Think about how the number of questions you answered correctly in the previous section compares with others. Which of the following is true?

( ) More than half of other participants will answer at least as many questions correctly as I did.
( ) Fewer than half of other participants will answer at least as many questions correctly as I did.
( ) Exactly half of other participants will answer at least as many questions correctly as I did.

Question: Considering everybody who completes this survey in the next couple of days, which group will have the highest percentage of correct answers?

() High pay scale group
() Low pay scale group
() Both groups will have the same percentage of correct answers

Question: Consider only those participants taking this survey assigned the high pay scale. How will the number of questions you answered correctly compare to the high pay scale group?

() More than half of participants in the high pay scale group will answer at least as many questions correctly as I did.
() Exactly half of participants in the high pay scale group will answer at least as many questions correctly as I did.
() Less than half of participants in the high pay scale group will answer at least as many questions correctly as I did.

Question: Consider only those participants taking this survey assigned the low pay scale. How will the number of questions you answered correctly compare to the low pay scale group?

() Less than half of participants in the low pay scale group will answer at least as many questions correctly as I did.
() More than half of participants in the low pay scale group will answer at least as many questions correctly as I did.
() Exactly half of participants in the low pay scale group will answer at least as many questions correctly as I did.

[Thank you page]

Thank you for completing this survey.

Please enter the following code in your mechanical turk HIT to complete your task:

[participant given a unique code here as another way to weed out robots]

Experiment 3

[Screen 1]

Thank you for agreeing to participate in our study. We appreciate your time and effort. Your earnings from this study will depend on the decisions you make during the study, so it is in your best interest to read all instructions carefully.

Please enter your mechanical turk worker id. This helps us to screen out robots.*

[Screen 2]

General Instructions

In this experiment, you will be asked to complete multiple tasks. Only one of these tasks will be chosen, at random, to determine your potential earnings from the experiment. Each task has the same chance of being chosen to count so it is in your best interest to complete each task carefully.

Each task will be labeled clearly and described as it arises. In total, this experiment will require approximately 15 minutes to complete. However, to allow some flexibility, you will have a maximum of 60 minutes to complete the experiment.

Approximately ten percent (10%) of participants will actually be paid their potential earnings. The remaining participants will be paid only the fixed fee listed on the HIT.

To be as fair and transparent as possible, the 10 percent of participants who will actually be paid their potential earnings are determined as follows:

- On the next page, you will choose a number from 0 to 9.
- We will compare your chosen number to the first number drawn in the next "California Mid-Day Daily 3" (http://www.calottery.com/play/draw-games/daily-3) conducted after you complete this experiment.
- If the number you choose matches the first number drawn in the next California Mid-Day Daily 3, you will actually be paid your potential earnings from this experiment. These earnings will be paid as a bonus to your mechanical turk worker account.
- Otherwise, you will earn only the fixed participation fee listed on the HIT.

When you have read and understood these terms, please click the button below to begin the experiment.

[Screen 3]

Please select a number from 0 to 9:

( ) 0
( ) 1
( ) 2
( ) 3
( ) 4
( ) 5
( ) 6
( ) 7
( ) 8
( ) 9

[Screen 4]

Task 1

In this task you will answer ten multiple choice questions. To determine your earnings from this task, you will be assigned one of two pay scales at random: a high pay scale or a low pay scale. Each pay scale is equally likely to be chosen so that half of the people participating in this experiment will be assigned the high pay scale and half the low pay scale.

If you are assigned the high pay scale:
- For each correct answer you will earn $2
- For each incorrect answer you will earn $1

If you are assigned the low pay scale:
- For each correct answer you will earn $1
- For each incorrect answer you will earn $0

On the next page, you will learn which pay scale you have been assigned.

[Screen 5a: shown only to those randomly assigned high pay]

Task 1 pay scale assignment

You have been assigned the high pay scale.

- For each correct answer on this task you will earn $2

- For each incorrect answer on this task you will earn $1

[Screen 5b: shown only to those randomly assigned low pay]

Task 1 pay scale assignment

You have been assigned the low pay scale.

- For each correct answer on this task you will earn $1

- For each incorrect answer on this task you will earn $0

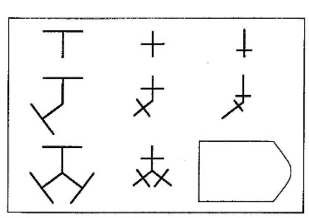
[Screens 6a-15a: ten separate effort-intensive tasks, typical task shown]

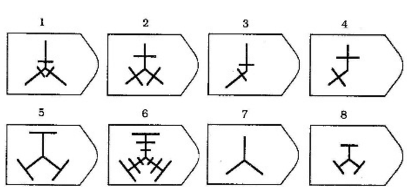
yufuWLDUthknYEgbNcYK
KcKtOoNJEMJvTlEffDmG
GZvjDwAstGKHJuhLvWLe
TZsvfTcVifbeEUQuexKa
jAekxnSfEbgnPIabzHko
kAtsvLZUecjoOvPpcayZ
eTLOQZzvdXZZJoCnOFeQ
uvlMHrPXgblZoQujraqB
sJoBWNNdUpRRmIMMxYxB
YjPkyHWZiCiqjoCmTPoK

How many times does "f" - i.e., lower case f - appear in the image above?

Your answer
( ) 1
( ) 2
( ) 3
( ) 4
( ) 5
( ) 6
( ) 7
( ) 8

[Screens 6b-15b: ten separate ability-intensive tasks, typical task shown]





Which piece number best completes the larger image?

Your answer
( ) 1
( ) 2
( ) 3
( ) 4
( ) 5

( ) 6
( ) 7
( ) 8

[Screen 16]

Task 2, Instructions

You will now begin Task 2. In this task you will answer two questions. Each question concerns how well you think you and others did on Task 1.

To determine whether your answers are correct, the "others" we will use in our calculations will be all experimental participants who are eligible to be selected to be paid their potential earnings during the same period as you: all participants who complete and submit this experiment between the last and the next mid-day California Daily 3 drawing.

One question will be chosen, at random, to determine your earnings from Task 2. Each question is equally likely to be chosen to count - we essentially flip a fair coin. Since each question can determine your entire potential earnings from this experiment, it is in your best interest to answer each question carefully.

For the question that is chosen to count:

- A correct answer will earn you $20
- An incorrect answer will earn you $0.

To begin Task 2, click the button below:

[Screen 17]

[Question order randomized.]

[Question 1]
<u>Question</u>: How will the number of questions you answered correctly on Task 1 compare to the number of Task 1 questions other participants in this experiment answer correctly?

On Task 1 ... [option order randomized]
( ) ... more than half of other participants in this experiment will answer at least as many questions correctly as I did.
( ) ... fewer than half of other participants in this experiment will answer at least as many questions correctly as I did.
( ) ... exactly half of other participants in this experiment will answer at least as many questions correctly as I did.

[Question 2]
Recall that on Task 1 there were two pay scales. Those assigned the high pay scale earned $2 for each correct answer and $1 for each incorrect answer. Those assigned the low pay scale earned $1 for each correct answer and $0 for each incorrect answer.

<u>Question</u>: Consider what percentage of answers each group, on average, will answer correctly in Task 1. Which of the two groups will have the largest percentage of correct answers?

Please note: when determining whether your answer is correct or not, we will exclude your own percentage of correct answers from our calculations. Therefore, you should consider only how other members of each group, excluding yourself, will perform.

On Task 1 ... [option order randomized]
( ) ... the low pay scale group will have the largest percentage of correct answers.
( ) ... the high pay scale group will have the largest percentage of correct answers.
( ) ... the high pay scale group and the low pay scale group will have exactly the same percentage of correct answers

[Screen 18]

Task 3, instructions

You will now begin a new task: Task 3. In Task 3 you will answer 10 questions similar to the questions you answered in Task 1.

To determine your earnings from this task, you must choose one of two options.

Option S:
- for each correct answer you will earn $ 1.50;
- for each incorrect answer you will earn $ 0.50

Option C:
- one other experimental participant will be selected at random;
- If you answer at least as many questions correctly as this other participant, you will earn $30;
- If you answer fewer questions correctly than this other participant, you will earn $0.

Note that every participant will have this choice, so that your decision will not affect anybody else's earnings besides your own.

Please select which earnings option you prefer [option order randomized]

( ) Option S
( ) Option C

[Screen 19a-28a: ten separate effort-intensive tasks, similar to screens 6a-15a]

[Refer to example above]

[Screen 19b-28b: ten separate effort-intensive tasks, similar to screens 6b-15b]

[Refer to example above]

[Screen 29]

Tell us a bit about yourself

You have now completed all tasks comprising the experiment. Please take a few moments to tell us a few things about yourself. Every response on this page is voluntary, and will not affect your earnings in any way.

In what year were you born?

In what month were you born?

Gender
( ) Male
( ) Female
( ) Other
In what country were you born?

[If US: In what US state were you born?]

Approximately what was your total income last fiscal year?
( ) $10,000 or less
( ) $10,001 - $30,000
( ) $30,001 - $50,000
( ) $50,001 - $70,000
( ) $70,001 - $90,000
( ) $90,001 - $110,000
( ) $110,001 - $130,000
( ) $130,001 - $150,000
( ) $150,001 - $170,000
( ) $170,001 - $190,000
( ) $190,001 or more

Which of the following academic degrees have you obtained (check all that apply)
[ ] High school diploma
[ ] Some college, but no degree
[ ] Associate degree
[ ] Bachelor's degree
[ ] Master's degree
[ ] Technical certificate
[ ] Professional degree
[ ] PhD

[Screen 30]

End of Experiment

You are now finished with the experiment. Thank you for participating! We appreciate your time and effort.

Please note that:
- There is an "end of experiment" code at the bottom of this page.
- Please copy this code and paste it into the appropriate box on your mechanical turk HIT.
- This makes it much easier for us to pay you quickly.
- Once you have pasted your code into the HIT, click "submit" below to send us your experiment.

For scientific reasons, because this experiment is on-going, we cannot tell you how well you performed. Everybody will be paid their fixed HIT fee as soon as possible. For the 10 percent of participants who will be chosen to be paid their potential earnings, these additional earnings will be paid as a bonus to their mechanical turk accounts shortly after the next time California Mid-Day Daily 3 drawing is conducted.

Your End of Experiment Code is: [unique code given to participants as a way to weed out robots]

Thank you for completing our experiment. Your time and effort are very important to us.