Second-Order Beliefs and Gender

Andrew Dustan[1], Kristine Koutout[1] and Greg Leo[1]

[1]Department of Economics, Vanderbilt University

August 12, 2021

Abstract

Beliefs about beliefs—second-order beliefs—about the differences between populations are important to understanding differences in outcomes between those populations. To study their potential impact, we develop an incentive-compatible experimental framework for eliciting beliefs (first-order) and beliefs about beliefs (second-order) about the differences in any measurable characteristics between any two populations. We implement the procedure to study beliefs about the performance of men and women on math and abstract bargaining tasks. In the math task, 78% of participants believe that most *men* believe men outscore women. In contrast, 34% believe that most *women* believe men outscore women. Despite these differences in second-order beliefs, we observe no such difference in first-order beliefs. The pattern of results is similar in the bargaining task. These results have important labor market implications for the persistence of gender gaps.

Keywords: Higher-Order Beliefs, Gender, Experimental Methods

> "Leaders in the field— men and sometimes women— simply don't believe that women are as good at doing science."
>
> Alison Coil
> From a 2017 article
> in Wired Magazine

[1]Contact information: Leo [Corresponding], g.leo@vanderbilt.edu, Vanderbilt University Department of Economics 5201 West End Avenue Nashville, TN 37235. Dustan, andrew.dustan@vanderbilt.edu. Koutout, kristine.f.koutout@vanderbilt.edu. We thank Andrea Moro, Federico Gutierrez, Ernesto Reuben, Nikos Nikoforakis, Enrique Fatas, Lise Vesterlund, and seminar attendees at BEEMA V, SEA 2019, ESA-North America 2019, The University of Toledo, and NYU Abu Dhabi for helpful feedback. Matthew Chambers provided outstanding programming and technical support. This research was approved by the Institutional Review Board at Vanderbilt University on November 6, 2017, with approval number 171857.

1 Introduction

Do women *believe* that leaders in science, technology, engineering, and math (STEM) fields believe that women are bad at doing science? Such beliefs about beliefs—*second-order beliefs*—could drive women to sort out of STEM fields, leading to the observed gender gap in employment (Beede et al., 2011). Importantly, this belief-driven sorting could occur regardless of leaders' true beliefs about women's scientific abilities. When historically persistent beliefs about the differences between men and women—*first-order beliefs*—cause disparities, they may generate second-order beliefs that perpetuate those disparities even once first-order beliefs change.

In addition to the STEM/non-STEM employment gender gap, second-order beliefs could contribute to empirically documented differences in men's and women's outcomes in education (Lundberg, 2017) and wages (Blau and Kahn, 2017), among other outcomes of interest. Moreover, second-order beliefs about the differences between populations characterized by other dimensions, such as race/ethnicity, religious affiliation, or sexual orientation, may be important to understanding differences in outcomes between these groups. Despite their potential importance, second-order beliefs about population differences have rarely been studied and never, to our knowledge, directly measured.

In this paper, we introduce second-order beliefs about the differences between two populations as an important factor to consider in the study of unequal outcomes and provide researchers with an incentive-compatible experimental framework to measure them. We implement this procedure in a lab experiment to elicit beliefs about characteristics that have received particular attention for their potential to partially explain gender gaps: ability and negotiation behaviors (Bertrand, 2011; Croson & Gneezy, 2009). To operationalize the study of these general categories, we have chosen tasks that are commonly used in the experimental literature as abstracted versions of these two domains. Specifically, we elicit beliefs about men's and women's performance on a timed math task (Niederle & Vesterlund, 2007; Reuben et al., 2014) and choices in the ultimatum game (Eckel & Grossman, 2001; Solnick, 2001).

We find an interesting contrast between first- and second-order beliefs. There is no evidence that men's and women's first-order beliefs differ;[1] however, both men and women believe that such differences exist. In the math task, 78% of participants believe that most *men* believe that men outscore women. In contrast, only 34% believe that most *women* believe that men outscore women. Moreover, we find no

[1] This result is consistent with other studies that find no gender differences in beliefs about men and women such as Babcock et al. (2017), Bordalo et al. (2019), Moss-Racusin et al. (2012), and Reuben et al. (2014).

evidence of significant differences between men and women in these second-order beliefs. Similarly in the bargaining task, we find that people believe that men and women hold different first-order beliefs even though we observe no such differences in the data.

In summary, even when men and women have similar first-order beliefs, second-order beliefs about men and women can vary substantially. These statistically and economically significant differences in beliefs about men's and women's beliefs may imply different incentives to acquire skills or to engage in the labor market. Our results suggest that second-order beliefs are an important, yet relatively unexplored, mechanism that could perpetuate gender gaps regardless of differences in skills or first-order beliefs.

We make two contributions in this paper. First, we develop a generalized incentive-compatible experimental framework to serve as a template for eliciting first- and second-order beliefs about the differences between two populations regarding a measurable characteristic. No other paper to our knowledge has presented a methodology for measuring second-order beliefs about population differences. We discuss our design decisions extensively in Section 2 to facilitate the practitioner's careful choice of 1) the property of a participant's beliefs to target, 2) the function of two population-specific distributions that implies this property, and 3) the experimental protocol to most effectively elicit this function.

In brief, the belief elicitation procedure works as follows. First, we elicit first-order beliefs. In the math task, for example, participants are asked to reveal their belief about who correctly answered more math summations in a timed task—a randomly chosen man or a randomly chosen woman (and by how many summations). Participants' stated beliefs are then compared to a random draw from a sample of people who completed the math task. We use the Binarized Scoring Rule (BSR) (Hossain & Okui, 2013) to incentivize the truthful revelation of beliefs.

After the first-order belief elicitation, we ask participants to reveal what they believe a random man and a random woman chose *when asked the same question they just answered*. Participants are again rewarded based on how close their stated belief is to a realized outcome drawn from a sample of first-order beliefs using the BSR. In this intuitive way, participants reveal their second-order beliefs.

Our second contribution is providing the first empirical evidence that second-order beliefs could lead to unequal outcomes. Existing studies of behavioral responses to potential discrimination cannot distinguish whether second-order beliefs or beliefs about others' *preferences* drive those responses. Similar to the distinction between classical statistical (Arrow, 1973; Phelps, 1972) and taste-based (Becker, 1957) discrimination models, beliefs about beliefs about measurable characteristics (such as productivity) have different policy implications than beliefs about preferences. Our results make the case for the explicit study of the role of beliefs about beliefs about population differences in generating unequal outcomes.

We next discuss the relationship of our work to the literature. In Section 2, we detail the experimental framework and rationale for each of the design decisions, as well as reasons why practitioners might make different decisions based on their specific research question. In Section 3, we present the results of our implementation of the experimental framework to study beliefs (and beliefs about beliefs) about the differences between men's and women's performance on a math task and choices in the ultimatum game. We conclude in Section 4 with a discussion of the implications of our results and directions for future research.

1.1 Related Literature

Higher-order beliefs about the strategic sophistication of opponents have received substantial attention in the experimental literature, especially with regard to the "level-k" model (see Crawford, Costa-Gomez & Iriberri, 2013, for a survey). The level-k model predicts game behavior based on a player's level of rationality. A level-2 player, for instance, is rational and believes that other players believe they are rational—a second-order belief. Kneeland's (2015) innovative study of strategic sophistication uses a player's chosen strategies in a series of "ring games" to measure lab participants' levels of rationality. She finds that 71% of participants make choices that rely on second- or higher-order beliefs.

More specifically, second-order beliefs with respect to individual actions have been studied in the experimental literature. The most closely related paper to ours in the literature on beliefs and strategic decision-making is Manski & Neri (2013), in which the authors elicit first- and second-order beliefs about actions in a 2×2 game to study consistency between actions and beliefs. The literature on guilt aversion also elicits second-order beliefs and correlates them to own actions (Bacharach et al., 2007; Bellemare et al., 2011; Charness & Dufwenberg, 2006; Dufwenberg & Gneezy, 2000; Guerra & Zizzo, 2004). These studies elicit reflective beliefs of the form: person A's beliefs about person B's beliefs about what person A will do. In contrast, we elicit beliefs of the form: person A's beliefs about person B's beliefs about differences between two populations in some measurable characteristic.

We are the first, to our knowledge, to propose second-order beliefs as a potential mechanism driving gender differences in outcomes. The closest literature to the study of these beliefs is the theoretical work on self-fulfilling prophecies in statistical discrimination models (see Fang & Moro, 2011 for a review). In these models, minority workers choose to invest less in human capital as a rational response to employers' beliefs that minority workers are less likely to invest. Workers' response rely on their beliefs about employers' beliefs about the differences between minority and majority workers; however, empirical work on statistical discrimination has focused exclusively on employers' beliefs (e.g. Dianat, Echenique, & Yariv, 2018; Altonji & Pierret, 2001; Ewens, Tomlin, & Wang, 2014), not workers' second-order beliefs. This attention on first-order beliefs could be due to the assumption that beliefs are accurate, so changing first-order beliefs will lead to changes in second-order beliefs. Our results on the differences in first- and second-order beliefs, as well as recent work demonstrating inaccurate statistical discrimination (Bohren, Haggag, Imas, & Pope, 2019), suggest this assumption is unlikely to hold in real markets.

Several papers study mechanisms that could be consistent with second-order beliefs. For example, Alston (2019) shows that women in a lab experiment anticipate discrimination on a sports trivia task and are willing to pay to hide their gender from prospective "employers." Charness et al. (2020) similarly show in a lab experiment that men are twice as likely as women to choose to reveal their gender in a job market for a stereotypically male task. Outside the lab, Glover, Pallais, & Pariente (2017) find that minority workers in French grocery stores perform better under less-biased supervisors, where bias is measured using an implicit association test. These papers observe behavior that could be consistent with second-order beliefs, but could also be due to beliefs about employer/supervisor preferences.

Coffman (2014) and Bordalo et al. (2019) study an idea closely related to second-order beliefs. In a series of lab experiments, Bordalo et al. test for the effects of "self-stereotyping" on confidence and behavior. Stereotypes such as "women are bad at math" are first-order beliefs about a measurable characteristic. In order to self-stereotype, a person must have beliefs about what those stereotypes are; therefore, when stereotypes can be classified as first-order beliefs, the person uses their *second-order beliefs* to self-stereotype.

Exploring another type of beliefs-based mechanism, Babcock et al. (2017) consider how the distribution of low-promotability tasks may impede women's career progression. They find that beliefs about willingness to accept these low promotability tasks are a primary driver of their inequitable distribution in the lab. In a related thread of literature that studies beliefs about social norms, Bursztyn, Gonzalez, & Yanagizawa-Drott (2018) measure and treat men's beliefs about other men's opinions about women working outside the home in Saudi Arabia.

While we are the first to study second-order beliefs about population differences, *first*-order beliefs about population differences have been studied in the lab using various elicitation procedures. Some of these procedures are indirect, where beliefs are inferred from actions. For instance, in Aguiar et al. (2009) participants choose whether they prefer to have a dictator allocation from a man or woman. Similarly, Castillo & Petrie (2010) and Fershtman & Gneezy (2001) infer beliefs about different races or ethnicities from contributions in a public goods game and choices in a trust game, respectively. Beliefs have also been elicited directly. Albrecht et al. (2013) use a price list to elicit beliefs about gender differences in a spatial reasoning task. Reuben et al. (2014) directly elicit expectations about men's and women's performance on a timed math task. Similarly, Schniter & Shields (2014) directly elicit expectations about the choices of young and old people in a trust game.[2]

2 Experimental Framework

In this section, we establish a framework for eliciting first- and second-order beliefs about the differences between two populations. The first step is precisely defining the beliefs of interest. That requires determining what properties of a participant's beliefs answer the research question.

The first-order belief we want to learn about in this experiment is whether a participant believes that, given a random draw from two populations, the characteristic of interest is most likely to be larger for the person drawn from population one or two. The second-order belief we want to learn about is whether a participant believes that random draws from population one/two are most likely to have first-order beliefs that favor one population or the other.

These beliefs may be relevant in a variety of scenarios. With respect to first-order beliefs, consider a professor who must choose between two students to advise—one is a man and the other a woman. The professor might care about who is most likely to be more productive. Then, for second-order beliefs, the woman student might care about whether it is mostly likely that the professor believes it is most likely that the man or the woman is more productive.

Different decision-making models can motivate the targeting of other properties of first-order beliefs. For instance, the relevant property for a profit-maximizing employer would be the difference in *expected* productivity between a man and woman, rather than whether a man or woman is *most likely* to be more productive. While we have chosen a simple, binary property of first-order beliefs to study, our procedure may be adapted to study beliefs about the mean difference between two populations, various quantiles of that difference, or other relevant properties.

The second-order belief of interest also depends on the decision-making model. There are two choices that determine the nature of the elicited second-order belief. The choice of the property of the participant's first-order belief determines the relevant second-order belief distribution. Then, just as in first-order beliefs, the ex-

[2]The goal of this paper is not to compare our method of eliciting first-order beliefs to other methods. Rather, we focus on the novelty of eliciting second-order beliefs about population differences.

perimenter needs to choose the property of interest for the participant's second-order belief distribution. From the example above, choosing to target first-order beliefs about the difference in expected productivity between a man and woman worker, instead of whether the man or woman is most likely to be more productive, changes the relevant second-order belief distribution.

2.1 Operationalizing the Property of Interest

To operationalize our property of interest, whether a participant believes that the characteristic of interest is most likely to be larger for the person drawn from population one or two, let X_1 be the random variable measuring the characteristic of interest in population one and X_2 be the same in population two. Then, we want to learn if the participant believes that these distributions have the property $P(X_1 > X_2) \geq \frac{1}{2}$ or that $P(X_1 < X_2) \geq \frac{1}{2}$. Either condition implies that from a randomly selected pair, the most likely outcome is the person from group one (or respectively two) has a higher value in the measure of interest.

For second-order beliefs, we want to learn whether a participant believes that, when we take a random draw from the first-order belief distribution, the belief is most likely to favor population one or two. So, we want to learn if a participant believes that a random draw from population one (or two) is most likely to *believe* that $P(X_1 > X_2) \geq \frac{1}{2}$ or $P(X_1 < X_2) \geq \frac{1}{2}$.

The next step is to determine what function of a participant's subjective belief distributions reflects the property of interest. In the following subsections, we explain why we choose to elicit the median of the distribution of differences between population-specific distributions. We also discuss the alternatives in some detail so the practitioner can make an informed decision based on their own property of interest.

2.2 The Median Difference

The property we describe, $P(X_1 > X_2) \geq \frac{1}{2}$ or $P(X_1 < X_2) \geq \frac{1}{2}$, is implied by a statement about the median of the *distribution of differences* between the populations, $X_1 - X_2$. If there exists a median strictly greater than zero:

$$P\left(X_1 - X_2 \geq Median(X_1 - X_2)\right) \geq \frac{1}{2} \Rightarrow P(X_1 - X_2 > 0) \geq \frac{1}{2} \Leftrightarrow P(X_1 > X_2) \geq \frac{1}{2}.$$

By the same argument, the existence of a median strictly below zero implies $P(X_1 < X_2) \geq \frac{1}{2}$. By eliciting the median of $X_1 - X_2$, we elicit the participant's first-order belief regarding which population (if any) is most likely to have a higher value in the measure of interest.

Now let Z_1 be the random variable measuring first-order beliefs in population one and Z_2 be the same in population two. Draws from Z_1 or Z_2 are draws of beliefs about the median of $X_1 - X_2$. A belief that $Median(Z_1) > 0$ implies a participant believes the probability a person from population one believes that $Median(X_1 - X_2) > 0$ is at least $\frac{1}{2}$. Thus, the belief that $Median(Z_1) > 0$ means the participant believes there is at least a $\frac{1}{2}$ probability that a randomly chosen person from population one believes that $P(X_1 > X_2) \geq \frac{1}{2}$.

While we choose to elicit similar properties of first- and second-order beliefs, it is possible (and reasonable) to choose different properties. For example, the practitioner interested in learning about managers' first-order beliefs regarding the expected differences in productivity may still want to learn about second-order beliefs about the most likely first-order belief. In this case, it would be appropriate to elicit the *mean* of the managers' first order beliefs, but the *median* of second-order beliefs.

2.3 Alternative Approaches

We choose to elicit medians because they offer precise information about the property we are interested in—whether $P(X_1 > X_2) \geq \frac{1}{2}$ or $P(X_1 < X_2) \geq \frac{1}{2}$. We next consider alternative functions of the participant's subjective belief distributions that could also elicit this information and discuss why we did not choose them for this experiment. Practitioners with different properties of interest may find these alternative functions more appropriate.

2.3.1 Eliciting Probabilities

One alternative approach would be to directly elicit the probabilities of interest: $P(X_1 > X_2)$ and $P(X_2 > X_1)$. These probabilities are means of binary distributions equal to 1 when the event occurs and equal to 0 otherwise, where the events are $x_1 > x_2$ or $x_2 > x_1$. As will be discussed in more detail in the next subsection on the payment structure, the BSR can elicit a mean as well as a median by using the appropriate loss function, ensuring that we could robustly elicit these probabilities. In fact, eliciting probabilities provides cardinal information about the participants' beliefs that is unobserved in our procedure. The cost of this additional information is an additional task for each belief elicited.

We choose not to take this approach because it requires two belief elicitations for each comparison of interest to determine which event is more likely. To determine whether $P(X_1 > X_2) \geq \frac{1}{2}$ or $P(X_1 < X_2) \geq \frac{1}{2}$ using the elicitation of probabilities would require that we elicit both $P(X_1 > X_2)$ and $P(X_1 < X_2)$. Since the outcome $x_1 = x_2$ is possible, the complement of $P(X_1 > X_2)$ is $P(X_1 \leq X_2)$, *not* $P(X_1 <$

X_2). While the cardinal information may be interesting, we argue that the precise probabilities of each event are not important enough in this experiment to justify the additional cognitive and time costs to participants from doubling the number of elicitations. Furthermore, since we use a random task payment procedure, doubling the number of tasks would also dilute the incentives.

2.3.2 Eliciting Modes of a Ternary Distribution

Another approach to determining which of a set of mutually independent outcomes is most likely is simply to ask participants which event they would like to condition their payment on. That is, ask participants to choose which outcome they think is most likely: $x_1 > x_2$, $x_1 < x_2$, or $x_1 = x_2$. This procedure is proper for eliciting the mode of a ternary distribution.

While the incentives of this procedure are clear and simple, participants with symmetric beliefs may nonetheless be incentivized to choose $x_1 > x_2$ or $x_1 < x_2$ instead of $x_1 = x_2$. Consider a continuous distribution that is identical for X_1 and X_2. Even though $X_1 = X_2$, it is sub-optimal to bet on the outcome $x_1 = x_2$ since $P(x_1 = x_2) = 0$. This also applies when X_1 and X_2 are discrete but the probability of equality is sufficiently low.

Under this payment structure, participants in our experiment who believe that men and women perform equally well on the math task would be incentivized to choose one of the non-gender-neutral outcomes simply because there are many more ways for two people to have a different math score than there are for two people to have the same math score. Therefore, we would not be able to distinguish gender-neutral participants.

In contrast, using the median procedure, a participant with symmetric beliefs is incentivized to select zero as their median belief regardless of their belief about the probability that the two randomly chosen subjects score identically. Participants with symmetric beliefs and participants whose beliefs are substantially asymmetric can always be differentiated.

2.3.3 Eliciting Population Medians

We elicit the median of a distribution of differences. An alternative approach would be to elicit the medians of each distribution separately and take the difference. In other words, there are two possibly relevant quantities involving medians: the median of the differences and the difference in the medians.

Eliciting the medians of X_1 and X_2 does not provide us the relevant information to assess our property of interest: whether $P(X_1 > X_2) \geq \frac{1}{2}$ or $P(X_1 < X_2) \geq \frac{1}{2}$. Specifically, $Median(X_1) > Median(X_2)$ does not imply that $Median(X_1 - X_2) > 0$.

Consider the data in Table 1: $Median(X_1) > Median(X_2)$ since $Median(X_1) = 3$ and $Median(X_2) = 2$; however, $Median(X_1 - X_2) = -1$ implying that $P(X_2 > X_1) > \frac{1}{2}$.

2.4 Incentive Structure

When eliciting beliefs, the first priority in experiment design is incentivizing truthful revelation. We begin with a payment structure that is incentive-compatible for all expected utility maximizers and some non-expected utility maximizers. The Binarized Scoring Rule (BSR), generalized by Hossain & Okui (2013), works by taking any proper scoring rule (i.e. a payment rule that reaches its maximum under truthfulness) and binarizing it, so that participants are maximizing the probability of winning the "large" prize rather than maximizing the size of the prize. This change in objective makes the payment rule incentive-compatible for all risk preferences. Using "probability currency" to induce risk-neutral behavior has a long tradition in experimental economics (Smith, 1961; Roth & Malouf, 1979), and similar binary procedures for belief elicitation are discussed by Karni (2009), Schlag & van der Weele (2013), and Qu (2012). Schlag & van der Weele (2013) specifically discuss binary lotteries for eliciting medians.

The probabilistic structure of the BSR outperforms other payment rules such as the popular Quadratic Scoring Rule (QSR) introduced by Brier (1950) (Hossain & Okui, 2013). The QSR incentivizes participants by varying the amount of money earned, rather than the probability of earning some fixed amount of money. That is, the closer a participant's predicted value is to the random realization, the more money they earn. This rule works for risk-neutral participants, but risk-averse participants would be incentivized to "hedge" their guess. Hossain & Okui show that participants in a lab experiment report more accurate beliefs under the BSR compared to the QSR when reporting probabilities, but the rules perform equally well in eliciting means, as theory predicts. In general, incentivized belief elicitation outperforms non-incentivized elicitation (Trautmann & van de Kuilen, 2014), particularly when there is a social cost to revealing beliefs as is the case with gendered beliefs (Babin, 2019).[3]

The BSR proceeds as follows: participants in the experiment win either prize A or prize B, with the value of A exceeding the value of B: $U(A) > U(B)$. We are interested in the random variable X. Participants report $\theta \in \Theta$ where θ is the participant's predicted value of some function of X. A loss function $l(x, \theta)$ returns

[3] Danz et al. (2020) have recently demonstrated that certain implementations of the binarized version of the quadratic scoring rule may be subject to a pull-to-center effect when eliciting a subjective probability (participants' stated beliefs are biased towards 50%); however, we do not elicit a probability, our implementation is substantially different, and such an effect would only make our results more conservative.

the prediction error from a random realization of X and the participant's predicted value θ. The experimenter compares the prediction error to a random draw K from a uniform distribution $U(0, \overline{K})$. If the prediction error is less than K, the participant wins prize A. Otherwise, the participant wins the lesser prize B. The form of the loss function determines which function of X participants should report. For example, the BSR would elicit the mean by binarizing the QSR loss function $(x-\theta)^2$. Other payment rules elicit the mode or quantiles, for example. The BSR procedure can be reduced to calculating the probability of winning the large prize A:

$$P(A) = 1 - \frac{l(x, \theta)}{\overline{K}}$$

As discussed in the previous subsection, we are interested in the median of participants' subjective distributions. The loss function for the median is $|x - \theta|$, so in our experiment

$$P(A) = 1 - \frac{|x - \theta|}{\overline{K}} \tag{1}$$

In this case, x is defined as a draw from the distribution of $X_1 - X_2$ for the first-order belief elicitation. For the second-order belief elicitation, x is defined as a draw from the distribution of Z_1 or Z_2.

A sufficient assumption on the utility function for the incentive-compatibility of the BSR to hold is *monotonicity with respect to stochastic dominance* (Hossain & Okui, 2013), originally defined by Machina & Schmeidler (1992). Although the monotonicity assumption is not satisfied by the expected utility functions in prospect theory (Kahneman & Tversky, 1979 and 1983), *Theorem 4* of Hossain & Okui (2013) extends the incentive-compatibility of the BSR to account for this type of preferences. The incentive-compatibility of the BSR holds in this case when the participant treats the large prize as a gain and the small prize as a loss. We have followed the advice of Hossain & Okui in setting the small prize to zero.[4]

While our procedure is incentive-compatible for many decision theories discussed in the literature, there is a possibility that the formal incentive compatibility does not extend to *some* decision process used by one of our participants. This has the potential to impair our interpretation of the elicited value as the median; however, note that $P(X > 0) \geq \frac{1}{2}$ is also implied by *any* quantile below 50% being larger than zero. In other words, $P(X > 0) \geq \frac{1}{2}$ implies that $P(X > 0) \geq \frac{1}{2} - \epsilon$ for all

[4] When losing the lottery, the subjects still leave with their show-up fee of \$5; however, we believe the subjects treat this as endowed wealth at the time of assessing the lotteries. The instructions re-enforce this by emphasizing that a loss in the lottery leads to zero gain.

$\epsilon \in [0, \frac{1}{2}]$. Thus, for some hypothesized decision theory to impair our interpretation of the ordinal information we collect, it would have to lead participants to report a quantile of their subjective belief *above* 50%.

2.5 Generating Samples for Incentives

In order to pay participants using the BSR, we need a sample from which to draw realizations. We pay participants for their first-order belief elicitation by sampling the measure of interest from populations one and two, X_1 and X_2. Then, we pay participants for their second-order belief elicitation by sampling from the first-order beliefs of populations one and two, Z_1 and Z_2. Therefore, we need two samples: one measuring the characteristic of interest and the other measuring first-order beliefs.[5]

The sample measuring the characteristic of interest can be generated as part of the experiment or taken from an existing data source (e.g. past experiments or administrative data). For example, the publicly available population distributions of SAT scores by gender can be sampled to incentivize elicitation of beliefs about the differences in men's and women's SAT performances. If the experimenter generates the data themselves, a single participant can be treated as a random draw from the population. Large samples are not needed— the measurement of one person from each population is sufficient.[6]

The characteristics of interest in this experiment are choices in an abstract bargaining task and scores on a timed math task. We use the Ultimatum Game as the bargaining task (see Eckel, Oliveira, & Grossman, 2008). In the Ultimatum Game, called "Task 1" in the experiment, Player 1 is endowed with \$10 and must decide how much to offer Player 2. Player 2 decides whether to accept Player 1's offer, or to reject, in which case both participants receive nothing. We use the strategy method to elicit participants' choices as both Player 1 and Player 2. Our measure of interest is Player 2's minimum acceptable offer (MAO), the smallest amount Player 1 could propose such that the participant would accept. Online Appendix A shows the instructions for the strategy-style Ultimatum Game.

Any differences between men's and women's MAOs (their *willingness to accept*) can be interpreted in multiple ways. First, since any amount above \$0 generates a higher payoff than rejecting, a participant interested only in maximizing earnings accepts any offer above \$0. A higher MAO indicates that the participant is motivated by more than earnings and may be interested in fairness, inequality aversion, competitiveness, etc. Since the Ultimatum Game has the structure of a take-it-

[5]Note that we cannot measure both in the same sample since we need the former to pay participants in the latter.

[6]While only one data point from each distribution is needed to incentivize belief elicitation, that data point must be truly random from the perspective of the subjects.

or-leave-it offer in negotiation, differences in MAO can also be interpreted in that context. For instance, women's lower average MAO in Eckel & Grossman (2001) could be due to social norms dictating that women should be more cooperative or less demanding. This interpretation is why we call the Ultimatum Game the bargaining task.

In Task 2, the math task, participants add sets of five two-digit numbers for five minutes. Participants are paid $0.50 for each correct sum. Online Appendix B shows the instructions for the math task. Previous work (Niederle & Vesterlund, 2007; Reuben et al., 2014) use timed arithmetic tasks because women and men perform equally well on them (see also Hyde et al., 1990). Despite this, people believe that men score higher than women in math tasks (Reuben et al., 2014).

Unlike the sample measuring the characteristic of interest, the sample measuring first-order beliefs should be collected using the belief elicitation procedure detailed here. The measurement of second-order beliefs relies on the recursive nature of our procedure (a belief about a belief is measured in the same terms as the original belief) to help participants understand the procedure. In other words, to intuitively define second-order beliefs, we need to be able to tell participants that other participants who we are asking about *answered the same questions they just did*.

Like the sample measuring the characteristic of interest, the sample of first-order beliefs can be as small as one person from each population. For example, in this experiment, the measurement of the characteristics of interest in one man and one woman would be sufficient to elicit first-order beliefs. Likewise, the elicitation of first-order beliefs from one man and one woman would be sufficient to incentivize the elicitation of second-order beliefs. To the participant, it does not matter if the random draw used to incentivize them is from a sample of 1 or from a sample of 1,000 because the sample itself is a random draw from the population.

2.6 Belief Elicitation

The belief elicitation procedure begins with the first-order belief elicitations about the characteristics of interest. We elicit participants' first-order beliefs by asking them to report who they believe performed "better"[7]—a randomly drawn person from population one or a randomly drawn person from population two—and by how much. For the math task in our experiment, we ask who answered more summations correctly and, for the bargaining task, who chose the higher MAO. Participants report their beliefs by moving a slider like the one presented in Figure 1. The sequence of probabilities reported in the accompanying table are determined by equation (1).

[7] We put "better" in quotation marks because in tasks like the Ultimatum Game, it is unclear whether a higher or lower MAO is better. This language is not used in the experiment.

Our implementation of the BSR communicates all relevant incentive information without teaching participants the complex payment rule. The interactive slider allows participants to observe, for every possible stated belief, their probability of winning in every realization of the random draw. Presenting this summary of the payment rule, rather than the payment rule itself, also eliminates the need for specialized mathematical knowledge, thereby increasing the range of potential applications. Simplicity in the belief elicitation procedure is particularly important in our experimental framework because we want to elicit second-order beliefs. To incentivize truthful revelation of second-order beliefs, participants must *believe* that other participants are incentivized to tell the truth about their first-order beliefs.

The slider's starting position is always the center, reporting that the man and woman scored equally in the task. Participants move the slider to the right if they believe the randomly selected man scored higher on the math task (or chose a higher MAO) and to the left if they believe the randomly selected woman scored higher (or chose a higher MAO). When the participant moves the slider, the table updates at each point of the support to show the associated sequence of probabilities of winning the large prize based on each possible realization of the random draw. Participants are told in the instructions that the procedure is designed such that it is optimal to report their best guess about the median.

Implementing equation (1) requires a choice for $\overline{K}$. Recall that $\overline{K}$ is the maximum on the uniform distribution from which we take a draw to compare to the evaluated loss function. That means $\overline{K}$ determines the size of the support over which participants can express their beliefs. There are trade-offs in the selection of $\overline{K}$. The wider the support, the flatter the slope of the objective function, weakening the incentive to be precise; however, a narrow range for $\overline{K}$ might truncate the choices of participants with more extreme beliefs. We choose to elicit beliefs over a 21 point support for both tasks: gender neutrality at zero and ten points on either side. This support matches the natural maximum of the Ultimatum Game, in which the largest difference is between a MAO of \$10 and \$0. Since there is no natural maximum for the math task, the choice might constrain our participants, so we label the endpoints as "10+".[8]

After eliciting participants' first-order beliefs, the belief elicitation procedure continues by informing participants that people from populations one and two answered the same questions they just did. We elicit second-order beliefs by asking participants to report what they believe a randomly drawn person from population one (and two) reported when *they* answered those questions. In our experiment, we ask participants what they believe a randomly chosen man from a previous session reported and, likewise, what a randomly chosen woman reported as her first-order

[8] We do not observe responses at the endpoints in our experiment.

belief for each characteristic. That is, we elicit four second-order beliefs— one for each gender/characteristic pair. As in the first-order belief elicitation, participants report their beliefs using a slider like Figure 1.

While we collect cardinal information about participants' median beliefs, the median was chosen only because it has an ordinal interpretation about underlying probabilities. The additional cardinal information may be interesting, but the cardinal results confound two factors: the magnitude of participants' beliefs about population differences and participants' beliefs about absolute levels of characteristics in the populations.

To illustrate this point, consider a participant who reports that their median belief is that a randomly selected man answers two more summations correctly than a randomly selected woman. The interpretation of those "two more summations" differs based on whether the participant believes people answer five summations total on average or twenty summations. Moreover, it is unclear how the additional quantitative results would be more informative than ordinal results. For example, knowing that people believe that men believe men outscore women on a simple math task may inform our understanding of the employment gap in STEM fields, but knowing specifically how many more math summations they are believed to outscore women by on this one particular task would not. Thus, in the Results section we focus on the ordinal information provided by the median beliefs.

2.7 Salience of Gender

We elect to make gender salient in our procedure, rather than try to disguise our intentions. Experimenters often obfuscate the purpose of an experiment about gender to avoid confounding factors such as an experimenter demand effect or social costs associated with revealing gendered beliefs. For example, one concern with our procedure is that *most* of the possible choices involve expressing some difference between men and women. This could create a demand effect, leading participants with neutral beliefs to express differences. On the other hand, revealing beliefs that "favor" one gender over another could impose some social cost on participants. This cost would bias results towards zero. Instead of attempting to design our experiment to neutralize these biases, we rely on our relatively strong and carefully designed monetary incentives to ensure that our results indicate true patterns in participants' beliefs.

Obfuscating gender is particularly untenable in our experiment because we want to elicit second-order beliefs. To elicit true second-order beliefs in our framework, it is vital that participants clearly understand that they are revealing their beliefs about men and women *and believe that other people clearly understood* that they were revealing their beliefs about men and women. When gender is obfuscated, this requirement becomes more burdensome since participants must also believe that other participants saw and interpreted the signal of gender in the same way they did. Even supposing that participants all interpret the signal of gender identically, obfuscating gender in both the first- and second-order belief elicitations means that participants reporting their second-order belief would need to deduce both the gender of the person in the first-order belief elicitation and the implied gender difference that person is asked about. This relatively complex task would confound the results in unclear ways. This argument focuses on the difficulty of eliciting second-order beliefs when gender is obfuscated in first-order beliefs; we then find it unlikely that obfuscating gender in the elicitation of second-order beliefs would be successful after having made gender salient in the first-order belief elicitation.

These potentially confounding factors are relevant to any experiment on socially sensitive topics, but it is important that we consider the implications for our interpretation of second-order beliefs. In our procedure, we incentivize participants to report what they believe another person *reported* as their first-order belief and interpret that elicitation as the participant's second-order belief. Participants who believe there are social costs, experimenter demand effects, or any other biasing factors, should account for them when reporting their second-order belief. This argument relies on participants being rational enough to consider the incentives of other participants. We believe participants are sophisticated enough to account for the full range of incentives affecting other participants;[9] therefore, a conservative interpretation of our most compelling results would be "participants believe that men and women *reveal* different first-order beliefs" rather than "*have* different beliefs."

2.8 Implementation

We implemented this experiment at the Vanderbilt University Experimental Economics Lab (VUEEL) from November 2017 to January 2018. Participants were recruited using the ORSEE system (Greiner, 2015), with no restrictions on who could participate. No one participated in more than one session of the experiment. The belief elicitation data come from 157 participants, 80 of whom are men and 77 of whom are women. The sample is comprised almost exclusively of Vanderbilt undergraduate students. Table 2 lists the sample sizes by gender for the samples used to incentivize belief elicitations as well as the sample that generates our belief elicitation data.[10]

[9] This belief is consistent with the results of Kneeland (2015), who finds that a large majority of subjects are at least second-order rational.

[10] Recall that we need only one draw from each population for each sample, but we collect slightly more.

Participants in the belief elicitation sessions received paper copies of the instructions used to measure the characteristics of interest, but completed the experiment on laptops using the oTree software (Chen et al., 2016). All instructions were read aloud by the experimenter. After the belief elicitation, the experiment concluded with a demographic survey. Each session lasted approximately 30 minutes.[11] See Online Appendix C for screenshots of the full experiment.

Participants received $5 for participating in the experiment and could earn the "large" prize of $15 from the belief elicitations. One decision out of the six[12] was chosen at random at the end of the experiment to determine payment and participants earned $18.09 on average, including the participation fee.

3 Results

We present the experimental results for the math task, summarized in Table 3 and Figure 2, followed by the bargaining task, summarized in Table 4 and Figure 3, and an intra-participant comparison of beliefs. We do not have predefined hypotheses about these belief distributions. One way to develop such hypotheses would be to use "common knowledge" arguments; however, the beliefs underlying those common knowledge arguments are precisely what we are seeking to measure. We describe the data instead.

3.1 Math Task

Most participants believe that there is *some* difference in men's and women's performance on the math task (86%, $SE = 2.8\%$), with 55% ($SE = 4.0\%$) believing that men outscore women. Testing for a difference in proportions, we cannot reject at conventional significance levels that men and women have the same probability of believing that men outscore women (59% for men vs. 51% for women, $p = 0.308$). Similarly, using a Wilcoxon rank-sum test, we cannot reject that men's and women's first-order belief distributions are identical ($p = 0.344$).[13] We note, however, that we cannot rule out a range of differences in first-order beliefs, including both positive and negative differences. For example, the 95% confidence interval for the men-women gap in the proportion believing that men outscored women is $[-7\%, 24\%]$.

[11]The time from the actual start of the experiment to when all participants completed the six belief elicitations and demographic survey was typically 15 to 20 minutes.

[12]There were two first-order belief elicitations and four second-order belief elicitations.

[13]The Wilcoxon rank-sum test for equality of first-order belief distributions uses the ternary distributions illustrated in Figure 2. Recall that we collect cardinal information, even though our outcome of interest is ternary. The Wilcoxon rank-sum test for equality of the first-order *cardinal* distributions gives $p = 0.652$. The cardinal distributions for all elicitations are in the Appendix.

Although we lack evidence that first-order beliefs differ by gender, participants believe that such differences in first-order beliefs exist. Using the Wilcoxon signed-rank test, we reject equality of distributions of second-order beliefs about men's beliefs and women's beliefs regarding math performance ($p = 0.000$).[14] Furthermore, 78% (SE=3.3%) of participants believe that most men believe men outscore women, while only 34% (SE=3.8%) of participants believe this about women's first-order beliefs, and a test of difference in proportions rejects that they are equal ($p = 0.000$). As with the first-order beliefs, we cannot reject that men's and women's second-order belief distributions are identical, with respect to either men's ($p = 0.257$) or women's ($p = 0.137$) first-order beliefs.[15]

While this experiment lacks the statistical power to make definitive statements about whether participants' second-order beliefs are correctly calibrated, some conclusions are possible. The 95% confidence set for the median of men's first-order ternary belief distribution includes both "no difference between man and woman" and "man outscores woman," but excludes "woman outscores man." The same is true for women's first-order beliefs. Thus, only a reported second-order belief (about either a man's or a woman's reported first-order belief) of "woman outscores man" can be classified as miscalibrated. Participants are much more likely to report this miscalibrated second-order belief about women (38%, SE=3.9%) than for men (8%, SE=2.2%), a difference that is statistically significant at the 1% level. To the extent that we can detect miscalibrated second-order beliefs in the data, it seems that the difference in gender-specific second-order beliefs is driven by participants wrongly believing that most women's first-order beliefs "favor" women.

To summarize, we are unable to reject equality in first- and second-order belief distributions between men and women, but have strong evidence that participants *believe* men and women hold different first-order beliefs. In particular, most participants believe that most men believe men outscore women, while they do not believe this about women.

3.2 Bargaining Task

Most participants believe that men choose a higher MAO than women (71%, SE=3.6%). Similar to the math task, we cannot reject that the proportions of men and women

[14]The Wilcoxon signed-rank test is used to account for intra-participant dependence. The Wilcoxon signed-rank test also rejects equality of cardinal second-order belief distributions ($p = 0.000$).

[15]We again use the Wilcoxon rank-sum test for equality of distributions, comparing the gender-specific ternary distributions. The null hypothesis of no differences in the gender-specific cardinal second-order belief distributions cannot be rejected for beliefs about men ($p = 0.180$) or about women ($p = 0.173$).

believing that men report a higher MAO are equal ($p = 0.212$).[16] Nor can we reject that the distributions of men's and women's first-order beliefs are the same ($p = 0.191$).[17] Again, we cannot rule out positive or negative differences in these proportions: the 95% confidence interval for the men-women gap in the proportion believing that men had a higher MAO is $[-23\%, 5\%]$. Interestingly, the point estimates suggest that men are 9.1 percentage points (SE=7.2%) *less likely* than women to hold the "stereotypical" belief that men have a higher MAO than women, although this difference is not statistically significant at conventional levels.

Participants again believe that men and women hold different first-order beliefs. We reject equality of the distributions of second-order beliefs about men's and women's first-order beliefs about which gender proposes a higher MAO ($p = 0.027$).[18] Interestingly, 68% (SE=3.7%) of participants believe that most women believe men choose a higher MAO, which is marginally higher than the 58% (SE=3.9%) of participants believing this about men's first-order beliefs ($p = 0.072$). Again, we cannot reject that men's and women's second-order beliefs are the same about men ($p = 0.475$) or about women ($p = 0.609$).[19]

The 95% confidence sets for the medians of men's and women's first-order ternary belief distributions include only "man has higher MAO than woman," meaning that all other second-order beliefs are miscalibrated. Second-order beliefs about both genders are often miscalibrated: 42% (SE=4.0%) of second-order beliefs about men are miscalibrated, as are 32% (SE=3.7%) about women. This difference in the rate of miscalibration between genders is marginally significant ($p = 0.080$), indicating that second-order beliefs about men are less accurate than those about women in this task.

Similar to the math task, we do not have consistent evidence of gender differences in either first- or second-order beliefs about the bargaining task. Yet participants believe that men and women differ in their first-order beliefs, being more likely to believe that women believe men choose a higher MAO than they are to believe this about men.

[16]The tests used for the bargaining task analysis are the same as those used for the math task: tests for differences in proportions when comparing proportions across genders, Wilcoxon rank-sum tests when testing for differences in ternary distributions between genders, and Wilcoxon signed-rank tests when testing for within-participant differences in second-order beliefs with respect to different genders.

[17]The test for differences in the cardinal distributions does find evidence of a difference ($p = 0.020$), but given the concerns with interpreting the cardinal measures and the lack of evidence for differences between the ternary distributions, we do not interpret this finding further.

[18]We also reject equality of the cardinal second-order belief distributions ($p = 0.001$).

[19]We also fail to reject equality of the cardinal second-order belief distributions about men ($p = 0.425$) or about women ($p = 0.925$).

3.3 Intra-participant Beliefs

Here, we describe the extent to which participants' second-order beliefs mirror their first-order beliefs.[20] This analysis is useful in understanding whether people form second-order beliefs about others in the same population solely by considering their own beliefs. To do this, we compare a participant's reported first-order belief to their second-order belief about a person of the same gender. Table 5 shows that, while the majority of participants believe that other participants of their same gender believe the same as themselves (57%, $SE = 4.0\%$ for the math task and 68%, $SE = 3.7\%$ for bargaining), these proportions are far from 1 and are quite similar for men and women.

Table 6 further explores the correspondence between first- and second-order beliefs, now conditioning on the first-order belief. Participants who believe that men perform better in the math task are more likely to believe that others of the same gender share that belief (70%, $SE = 5.0\%$) than those who believe the genders perform the same (41%, $SE = 10.7\%$) and those who believe that women performed better (also 41%, $SE = 7.1\%$), and these differences in proportions are statistically significant at conventional levels ($p = 0.001$ and $p = 0.012$, respectively). Similarly for the bargaining task, participants believing that men choose a higher MAO were more likely to believe that others of the same gender shared that belief (78%, $SE = 4.0\%$) than those believing that genders performed the same (50%, $SE = 9.0\%$, $p = 0.003$ for difference in proportions) or that women choose a higher MAO (36%, $SE = 13.3\%$, $p = 0.001$ for difference in proportions).[21]

4 Discussion

We establish an experimental framework for measuring both first- and second-order beliefs about the difference in some measurable characteristic between two populations. We implement the procedure in the lab to measure beliefs about the differences between men and women in their performance on a math task and choices in an abstract bargaining task. Our results are interesting, but intuitive. While men and women exhibit no statistically distinguishable differences in their first-order beliefs, people *believe* that such differences exist.

[20] We have not provided results on whether second-order beliefs are accurate. If second-order beliefs were accurate, then the distribution of second-order beliefs would converge to a point-mass at the true median of first-order belief medians. Thus, the accuracy of the distribution of second-order beliefs requires a degenerate distribution of second-order beliefs. In this sense, the "accuracy" of second-order beliefs does not involve second-order and first-order beliefs being similar. Instead, we study the similarity in these distributions through the lens of intra-participant consistency.

[21] We disaggregate the results by gender in the Appendix and note that there is a significant difference between men's and women's intra-participant beliefs in the math task; however, we do not interpret these results as the cell sizes are very small.

The potential implications of such discordant beliefs in real-world markets are far-reaching. Consider a woman who believes that male managers believe men to be more productive than women in STEM fields. She may pay some economic cost to be matched with a female manager rather than a male manager, even though there may be, in fact, no difference in male and female managers' beliefs. These second-order beliefs could contribute to observed gender differences in outcomes like the employment gap in STEM, regardless of differences in first-order beliefs or skills. Beyond the labor market, these second-order beliefs may have important implications in human capital, healthcare, marriage, and fertility decisions.

Mechanisms have been proposed to explain gender differences in market outcomes that may be, in part, driven by second-order beliefs, further underlining their importance. For example, statistical discrimination models (see Fang & Moro, 2011 for a review) require that minority workers believe that employers believe they have lower human capital—a second-order belief—to establish the self-fulling prophecy. Dianat et al. (2019) recognize the necessity of workers' "second-order rationality" in their lab experiment artificially creating statistical discrimination. Glover et al. (2017) find evidence of a self-fulfilling prophecy in French grocery stores. Minority workers exert more effort than majority workers under unbiased managers, but perform worse under biased managers. The workers' second-order beliefs about managers' beliefs are essential to explaining this behavior. Our experimental framework can be used to test the underlying assumptions on beliefs in models and experiments such as these.

A number of avenues are open for future work. First, and foremost, is showing whether second-order beliefs affect market behavior. Second is understanding how second-order beliefs are formed. One promising theory applies the stereotype model in Bordalo et al. (2019). In this model, second-order beliefs are an exaggeration of first-order beliefs.

While we have focused on gender in this paper, the procedure is sufficiently general to study differences about other types of populations. The experimental framework can be used to elicit beliefs about differences by races/ethnicities, religious beliefs, sexual orientation, STEM/non-STEM workers, and political affiliation. Only small samples from the populations of interest are required to incentivize first- and second-order belief elicitation, enabling the study of beliefs about much smaller and difficult to recruit populations than was previously practical. Second-order beliefs likely play a role in how all of these populations interact with each other, so our experimental framework provides a general tool that can be adapted to study beliefs in most contexts.

References

Aguiar, F., Brañas-Garza, P., Cobo-Reyes, R., Jimenez, N., & Miller, L. M. (2009). Are women expected to be more generous?. Experimental Economics, 12(1), 93-98.

Albrecht, K., Von Essen, E., Parys, J., & Szech, N. (2013). Updating, self-confidence, and discrimination. European Economic Review, 60, 144-169.

Alston, M. (2019). The (Perceived) Cost of Being Female: An Experimental Investigation of Strategic Responses to Discrimination. Working paper.

Altonji, J. G., & Pierret, C. R. (2001). Employer learning and statistical discrimination. The quarterly journal of economics, 116(1), 313-350.

Andersen, S., Fountain, J., Harrison, G. W., & Rutström, E. E. (2014). Estimating subjective probabilities. Journal of Risk and Uncertainty, 48(3), 207-229.
Arrow, K. (1973). The theory of discrimination. Discrimination in labor markets, 3(10), 3-33.

Babcock, L., Recalde, M. P., Vesterlund, L., & Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. American Economic Review, 107(3), 714-47.

Babin, J. (2019). Detecting Group Gender Stereotypes: Opinion-mining vs. Incentivized Coordination Games. Journal of Economic Insight, 45 (1), 21-42.

Bacharach, M., Guerra, G., & Zizzo, D. J. (2007). The self-fulfilling property of trust: An experimental study. Theory and Decision, 63(4), 349-388.

Bernheim, B. D. (1984). Rationalizable strategic behavior. Econometrica: Journal of the Econometric Society, 1007-1028.

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., & Doms, M. E. (2011). Women in STEM: A gender gap to innovation. Economics and Statistics Administration Issue Brief, (04-11).

Bellemare, C., Sebald, A., & Strobel, M. (2011). Measuring the willingness to pay to avoid guilt: estimation using equilibrium and stated belief models. Journal of Applied Econometrics, 26(3), 437-453

Bertrand, M. (2011). New perspectives on gender. Vol. 4B Handbook of Labor Economics.

Blau, F. D., Currie, J. M., Croson, R. T., & Ginther, D. K. (2010). Can mentoring help female assistant professors? Interim results from a randomized trial. American Economic Review, 100(2), 348-52.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. Journal of Economic Literature, 55(3), 789-865.

Bohren, J. A., Haggag, K., Imas, A., & Pope, D. G. (2019). Inaccurate statistical discrimination (No. w25935). National Bureau of Economic Research.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2019). Beliefs about gender. American Economic Review, 109(3), 739-73.

Brier, G. W. (1950). The statistical theory of turbulence and the problem of diffusion in the atmosphere. Journal of Meteorology, 7(4), 283-290.

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2018). Misperceived social norms: Female labor force participation in Saudi Arabia (No. w24736). National Bureau of Economic Research.

Castillo, M., & Petrie, R. (2010). Discrimination in the lab: Does information trump appearance?. Games and Economic Behavior, 68(1), 50-59.

Camerer, C. F., & Ho, T. H. (1994). Violations of the betweenness axiom and nonlinearity in probability. Journal of risk and uncertainty, 8(2), 167-196.

Charness, G., Cobo-Reyes, R., Meraglia, S., & Sánchez, Á. (2020). nAnticipated Discrimination, Choices, and Performance: Experimental Evidence. European Economic Review, 103473.

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. Econometrica, 74(6), 1579-1601.

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

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. The Quarterly Journal of Economics, 129(4), 1625-1660.

Coil, A. (2017). "Why Men Don't Believe the Data on Gender Bias in Science." Wired, August 25, 2017.

Crawford, V. P., Costa-Gomes, M. A., & Iriberri, N. (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. Journal of Economic Literature, 51(1), 5-62.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. Journal of Economic literature, 47(2), 448-74.

Danz, D., Vesterlund, L., & Wilson, A. J. (2020). Belief elicitation: Limiting truth telling with information on incentives (No. w27327). National Bureau of Economic

Research.

Dianat, A., Echenique, F., & Yariv, L. (2018). Statistical discrimination and affirmative action in the lab.

Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. Games and economic Behavior, 30(2), 163-182.

Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. The economic journal, 108(448), 726-735.

Eckel, C. C., & Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games. Economic Inquiry, 39(2), 171-188.

Eckel, C., De Oliveira, A. C., & Grossman, P. J. (2008). Gender and negotiation in the small: are women (perceived to be) more cooperative than men?. Negotiation Journal, 24(4), 429-445.

Ewens, M., Tomlin, B., & Wang, L. C. (2014). Statistical discrimination or prejudice? A large sample field experiment. Review of Economics and Statistics, 96(1), 119-134.

Fang, H., & Moro, A. (2011). Theories of statistical discrimination and affirmative action: A survey. In Handbook of social economics (Vol. 1, pp. 133-200). North-Holland.

Fehr-Duda, H., De Gennaro, M., & Schubert, R. (2006). Gender, financial risk, and probability weights. Theory and decision, 60(2-3), 283-313.

Fershtman, C., & Gneezy, U. (2001). Discrimination in a segmented society: An experimental approach. The Quarterly Journal of Economics, 116(1), 351-377.

Friedenberg, A. (2019). Bargaining Under Strategic Uncertainty: The Role of Second-Order Optimism. Econometrica, 87(6), 1835-1865.

Glover, D., Pallais, A., & Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores. The Quarterly Journal of Economics, 132(3), 1219-1260.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. Journal of the Economic Science Association, 1(1), 114-125.

Guerra, G., & Zizzo, D. J. (2004). Trust responsiveness and beliefs. Journal of Economic Behavior Organization, 55(1), 25-30.

Harrison, G. W., Martínez-Correa, J., & Swarthout, J. T. (2014). Eliciting subjective probabilities with binary lotteries. Journal of Economic Behavior & Organization, 101, 128-140.

Harsanyi, J. C. (1967). Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model. Management Science, 14(3), 159-182.

Hossain, T., & Okui, R. (2013). The binarized scoring rule. Review of Economic Studies, 80(3), 984-1001.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. Psychological bulletin, 107(2), 139.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. Econometrica, 47(2), 263-292.

Kahneman, D., & Tversky, A. (1983). Choice, values, and frames. American Psychological Association, 39(4), 341-350.

Karni, E. (2009). A mechanism for eliciting probabilities. Econometrica, 77(2), 603-606.

Kneeland, T. (2015). Identifying Higher-Order Rationality. Econometrica, 83(5), 2065-2079.

Leo, G. (2020). Belief Elicitation through Rank-Ordering. Working paper.

Lundberg, S. J. (2017). Father absence and the educational gender gap. Working Paper.

Machina, M. J., & Schmeidler, D. (1992). A more robust definition of subjective probability. Econometrica: Journal of the Econometric Society, 745-780.

Manski, C. F., & Neri, C. (2013). First-and second-order subjective expectations in strategic decision-making: Experimental evidence. Games and Economic Behavior, 81, 232-254.

Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2014). Managing self-confidence. NBER Working paper.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. Proceedings of the national academy of sciences, 109(41), 16474-16479.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much?. The Quarterly Journal of Economics, 122(3), 1067-1101.

Pearce, D. G. (1984). Rationalizable strategic behavior and the problem of perfection. Econometrica: Journal of the Econometric Society, 1029-1050.

Perea, A. (2012). Epistemic game theory: reasoning and choice. Cambridge University Press.

Phelps, E. S. (1972). The statistical theory of racism and sexism. The american economic review, 659-661.

Porter, C., & Serra, D. (2019). Gender differences in the choice of major: The importance of female role models. American Economic Journal: Applied Economics.

Qu, X. (2012). A mechanism for eliciting a probability distribution. Economics Letters, 115(3), 399-400.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. Proceedings of the National Academy of Sciences, 111(12), 4403-4408.

Riley, E. (2017). Role models in movies: The impact of Queen of Katwe on students' educational attainment (No. 2017-13). Centre for the Study of African Economies, University of Oxford.

Roth, A. E., & Malouf, M. W. (1979). Game-theoretic models and the role of information in bargaining. Psychological review, 86(6), 574.

Schniter, E., & Shields, T. W. (2014). Ageism, honesty, and trust. Journal of Behavioral and Experimental Economics, 51: 19-29.

Schlag, K. H., & van der Weele, J. J. (2013). Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality. Theoretical Economics Letters, 3(1), 38-42.

Schlag, K. H., Tremewan, J., & Van der Weele, J. J. (2015). A penny for your thoughts: A survey of methods for eliciting beliefs. Experimental Economics, 18(3), 457-490.

Smith, C. A. (1961). Consistency in statistical inference and decision. Journal of the Royal Statistical Society: Series B (Methodological), 23(1), 1-25.

Solnick, S. J. (2001). Gender differences in the ultimatum game. Economic Inquiry, 39(2), 189-200.

Trautmann, S. T., & van de Kuilen, G. (2014). Belief elicitation: A horse race among truth serums. The Economic Journal, 125(589), 2116-2135.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and uncertainty, 5(4), 297-323.

Wilson, A., & Vespa, E. (2018). Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language. Working paper.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. Management science, 42(12), 1676-1690.

Tables

Table 1: Example distributions illustrating that $Median(X_1) > Median(X_2)$ does not imply $P(X_1 > X_2) \geq \frac{1}{2}$

	Value		
X_1	0	3	4
X_2	1	2	5

Table 2: Sample sizes for incentive and belief elicitation samples

	Task	First-order beliefs only	Belief elicitation
Female	12	4	77
Male	10	4	80

Notes: Column headers denote the samples. We measure the characteristics of interest in the “task” sample (to be used in incentivizing first-order belief elicitations). The “first-order beliefs only” sample is used to incentivize second-order belief elicitations for participants in the full “belief elicitation” sessions, which provide the data analyzed in the experiment.

Table 3: Belief elicitation results for math task

	All	Men	Women	Difference
First-Order Beliefs				
W>M	0.312	0.287	0.338	-0.050
	(0.037)	(0.051)	(0.054)	(0.074)
W=M	0.140	0.125	0.156	-0.031
	(0.028)	(0.037)	(0.042)	(0.055)
W<M	0.548	0.588	0.506	0.081
	(0.040)	(0.055)	(0.057)	(0.079)
Second-Order Beliefs, about Men				
W>M	0.083	0.087	0.078	0.010
	(0.022)	(0.032)	(0.031)	(0.044)
W=M	0.140	0.175	0.104	0.071
	(0.028)	(0.043)	(0.035)	(0.055)
W<M	0.777	0.738	0.818	-0.081
	(0.033)	(0.050)	(0.044)	(0.066)
Second-Order Beliefs, about Women				
W>M	0.376	0.412	0.338	0.075
	(0.039)	(0.055)	(0.054)	(0.077)
W=M	0.287	0.313	0.260	0.053
	(0.036)	(0.052)	(0.050)	(0.072)
W<M	0.338	0.275	0.403	-0.128
	(0.038)	(0.050)	(0.056)	(0.075)
Observations	157	80	77	157

Notes: Columns (1) to (3) reference subsamples. Column (4) reports the differences between the men and women subsamples. Standard errors are reported in parentheses underneath the proportions. The rows "W>M", "W=M", and "W<M" report the proportion of participants in the math task who believe that the woman scores higher, the woman scores the same, the woman scores lower compared to the man.

Table 4: Belief elicitation results for bargaining task

	All	Men	Women	Difference
First-Order Beliefs				
W>M	0.089	0.113	0.065	0.048
	(0.023)	(0.036)	(0.028)	(0.045)
W=M	0.204	0.225	0.182	0.043
	(0.032)	(0.047)	(0.044)	(0.064)
W<M	0.707	0.662	0.753	-0.091
	(0.036)	(0.053)	(0.049)	(0.072)
Second-Order Beliefs, about Men				
W>M	0.166	0.188	0.143	0.045
	(0.030)	(0.044)	(0.040)	(0.059)
W=M	0.255	0.188	0.325	-0.137
	(0.035)	(0.044)	(0.054)	(0.069)
W<M	0.580	0.625	0.532	0.093
	(0.040)	(0.054)	(0.057)	(0.079)
Second-Order Beliefs, about Women				
W>M	0.089	0.113	0.065	0.048
	(0.023)	(0.036)	(0.028)	(0.045)
W=M	0.236	0.225	0.247	-0.022
	(0.034)	(0.047)	(0.049)	(0.068)
W<M	0.675	0.662	0.688	-0.026
	(0.037)	(0.053)	(0.053)	(0.075)
Observations	157	80	77	157

Notes: Columns (1) to (3) reference subsamples. Column (4) reports the differences between the men and women subsamples. Standard errors are reported in parentheses underneath the proportions. The rows “W>M”, “W=M”, and “W<M” report the proportion of participants in the bargaining task who believe that the woman chooses higher MAO, the woman chooses the same, the woman chooses lower MAO compared to the man.

Table 5: Proportion of participants reporting same-gender second-order beliefs matching their own first-order beliefs

	All	Men	Women
Math	0.567	0.613	0.519
	(0.040)	(0.054)	(0.057)
Bargaining	0.682	0.675	0.688
	(0.037)	(0.052)	(0.053)
Observations	157	80	77

Notes: Comparison is with respect to ternary belief distributions. Columns refer to subsamples. Standard errors are reported in parentheses underneath the proportions.

Table 6: Proportion of participants reporting same-gender second-order beliefs matching their own first-order beliefs, by first-order beliefs.

	W>M	W=M	W<M
Math	0.410	0.410	0.700
	(0.070)	(0.105)	(0.050)
Observations	49	22	86
Bargaining	0.357	0.500	0.775
	(0.128)	(0.088)	(0.039)
Observations	14	32	111

Note: Columns specify participant's first-order belief: woman higher than man, gender-neutral, and man higher than woman. Comparison is with respect to ternary belief distributions. Standard errors are reported in parentheses underneath the proportions.

Figures

Figure 1: Example of slider interface used for belief elicitation

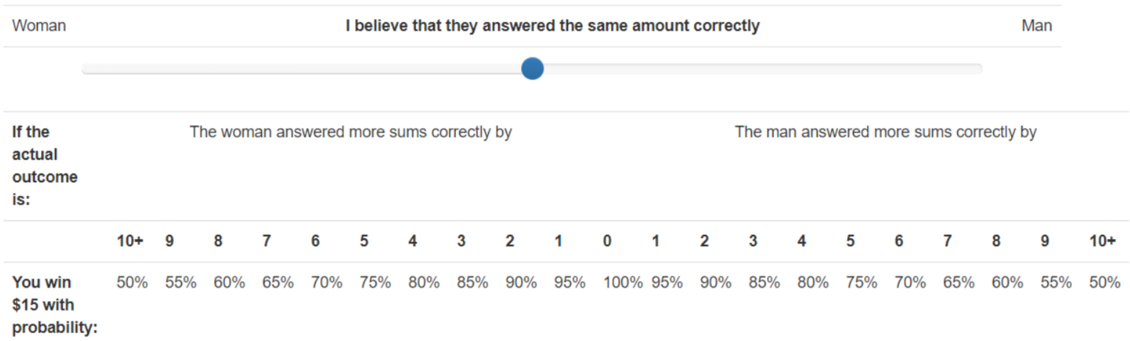
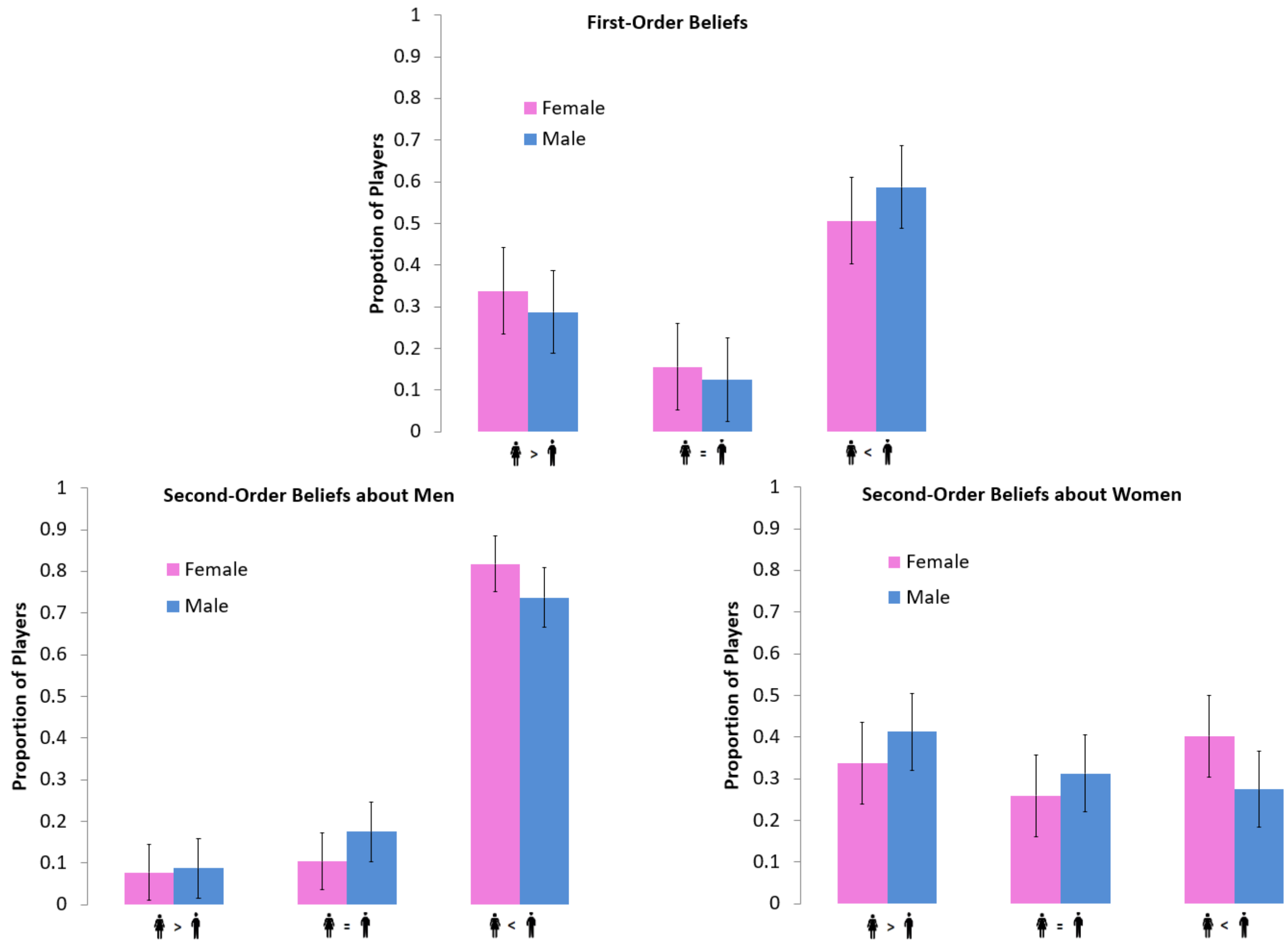
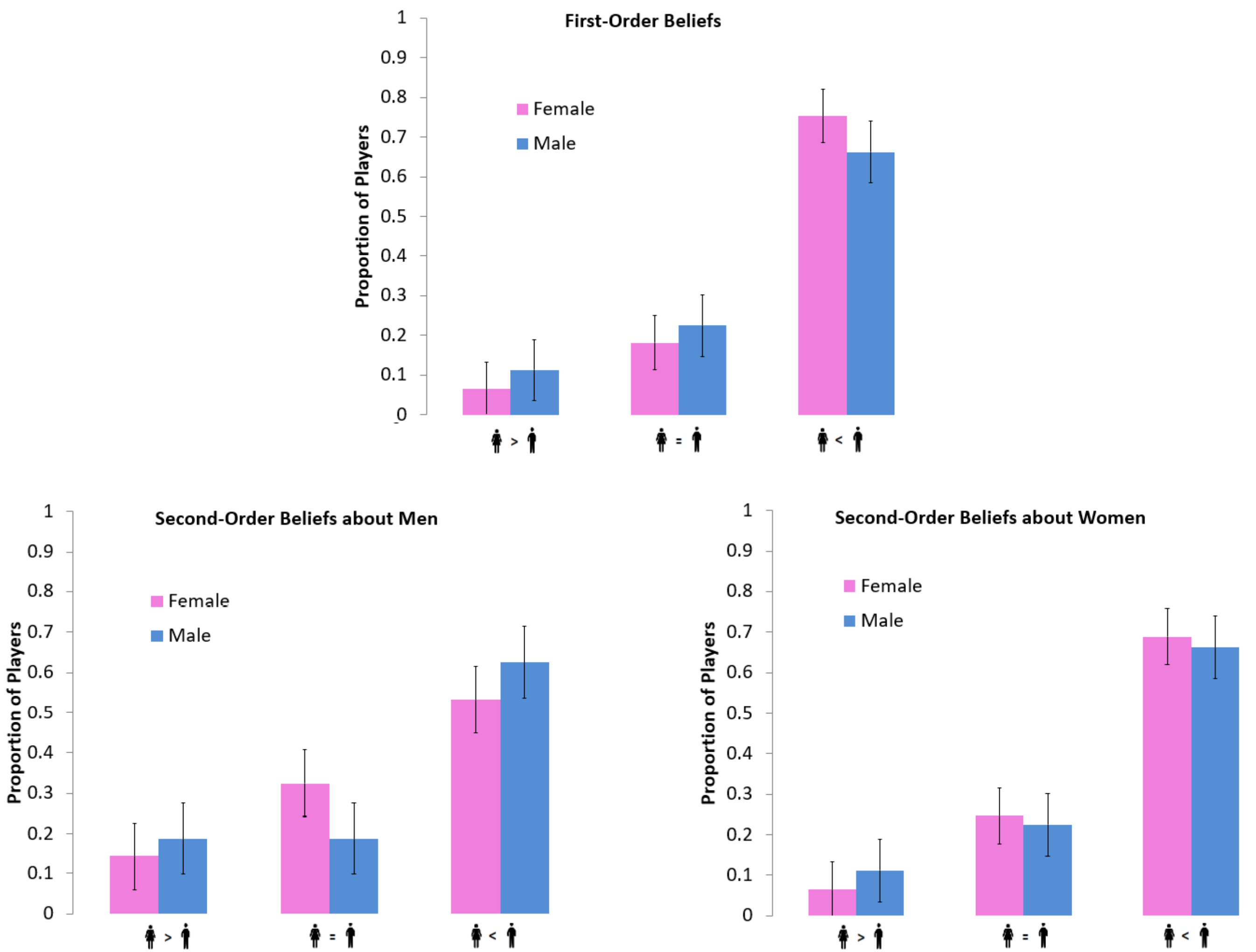


Figure 2: Belief elicitations about the math task



Notes: From left to right on each graph, the bars represent the proportion of participants in the math task who believe that the woman scores higher, the woman scores the same, the woman scores lower compared to the man.

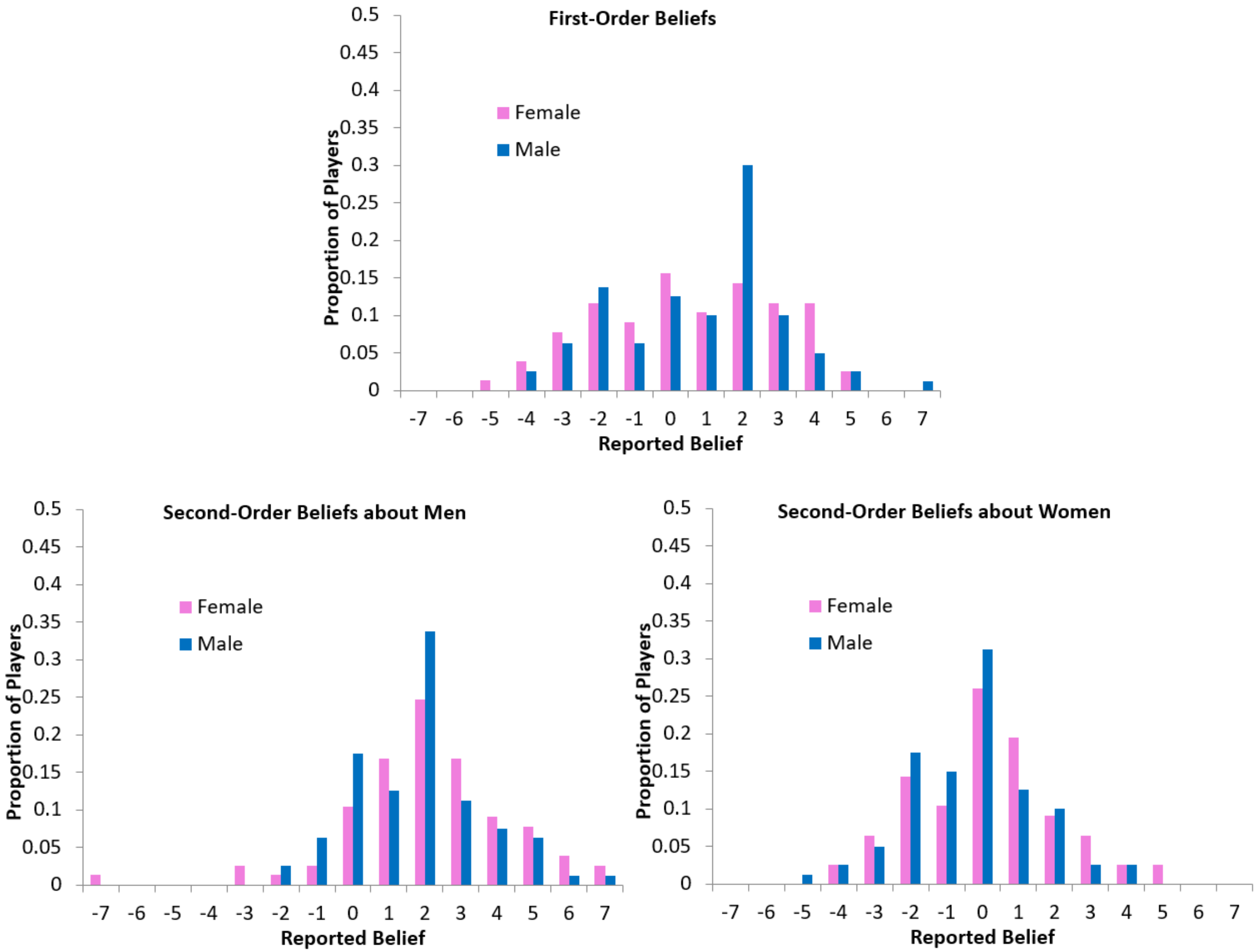
Figure 3: Belief elicitations about the bargaining task



Notes: From left to right on each graph, the bars represent the proportion of participants in the bargaining task who believe that the woman chooses higher MAO, the woman chooses the same, the woman chooses lower MAO compared to the man.

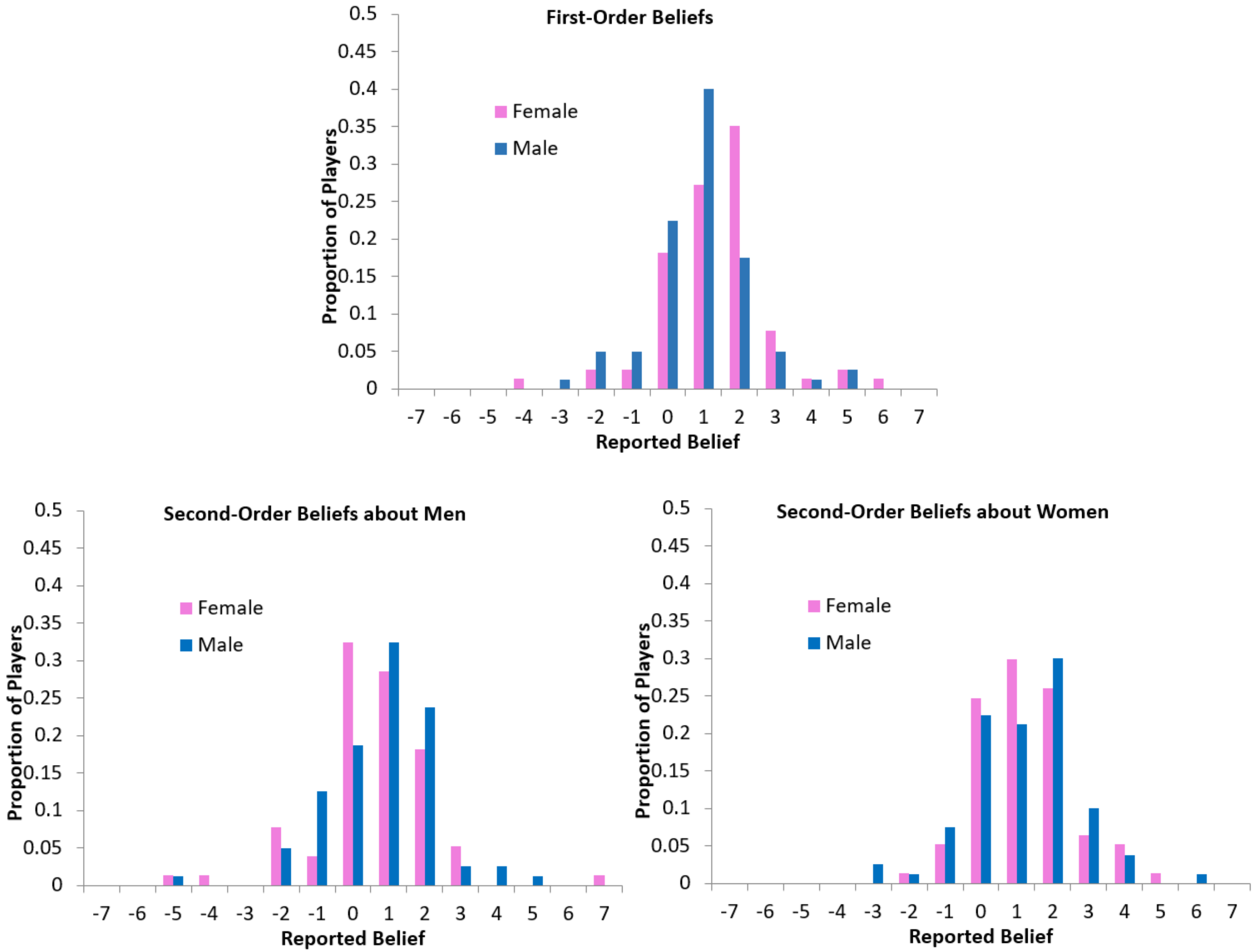
Appendix: Supplemental Figures and Table

Figure A1: Belief Elicitation about the Math Task



Notes: The bars represent the proportion of players who report each median difference between a man and a woman on the math task. A negative difference means that the woman answered more math summations correctly, while a positive difference means that the man answered more correctly.

Figure A2: Belief Elicitations about the Bargaining Task



Notes: The bars represent the proportion of players who report each median difference between a man and a woman on the bargaining task. A negative difference means that the woman chose a higher MAO, while a positive difference means that the man chose a higher MAO.

Table A1: Proportion of participants reporting same-gender second-order beliefs matching their own first-order beliefs, by first-order belief

	All	Men	Women	Difference
Math Task				
W>M	0.408	0.261	0.538	-0.278
	(0.071)	(0.094)	(0.100)	(0.134)
	[49]	[23]	[26]	
W=M	0.409	0.400	0.417	-0.017
	(0.107)	(0.163)	(0.149)	(0.210)
	[22]	[10]	[12]	
W<M	0.698	0.830	0.538	0.291
	(0.050)	(0.055)	(0.081)	(0.097)
	[86]	[47]	[39]	
Ultimatum Task				
W>M	0.357	0.556	0.000	0.556
	(0.133)	(0.176)	(0.000)	(0.166)
	[14]	[9]	[5]	
W=M	0.500	0.500	0.500	0.000
	(0.090)	(0.121)	(0.139)	(0.178)
	[32]	[18]	[14]	
W<M	0.775	0.755	0.793	-0.038
	(0.040)	(0.060)	(0.054)	(0.080)
	[111]	[53]	[58]	

Note: Columns (1) to (3) reference subsamples. Column (4) reports the differences between the men and women subsamples. Standard errors are reported in parentheses underneath the proportions. The number of participants in each cell are reported in brackets underneath the standard errors.

Online Appendix A: Task 1 Instructions

Participation ID ________

Task 1

In this task, you will be paired with a random partner. Your earnings will depend on the choice you make and the choice your partner makes. One of you will be assigned to be "Person 1" and the other to be "Person 2". The partner assigned to be Person 1 will propose how to split a total of $10 between the two partners. In other words, Person 1 proposes how much of the $10 to give to Person 2 and how much to keep for him or herself.

Person 2 then decides whether to accept or reject the split proposed by Person 1. If Person 2 accepts the proposal, the money is divided between Person 1 and Person 2 as proposed. If Person 2 rejects the proposal, both partners earn $0.

You must decide on the actions you will take in this game before knowing whether you will be Person 1 or Person 2. At the end of the experiment, we will pair you randomly with a partner and make choices on your behalf based on what you submit below. You will not know who your partner is and your partner will not know who you are. While your choices in this task will be used to determine your earnings, your choices will not be revealed during or after the experiment.

If you are **Person 1,** how much of the $10 would you like to propose to give to Person 2 (circle one)?

I propose to give Person 2:

$0 $1 $2 $3 $4 $5 $6 $7 $8 $9 $10

If you are **Person 2,** what is the smallest amount that Person 1 could propose to give you that you would accept (circle one)? If you are in the role of Person 2 and Person 1 offers you any amount equal to or larger than the number you circle below, you will automatically accept the split. If Person 1 offers you any amount less than the number you circle below, you will automatically reject the split and you will both earn $0.

The smallest amount that I would accept from Person 1 is:

$0 $1 $2 $3 $4 $5 $6 $7 $8 $9 $10

Online Appendix B: Task 2 Instructions

Participation ID ________

Task 2

During this task you earn money by correctly summing 2-digit numbers. You will be shown several sets of five two-digit numbers. Each set will be arranged in a row. For example, you could see:

60	71	41	75	81	

For each set, you will write your answer in the empty box on the right. In the above example, the correct answer is 60 + 71 + 41 + 75 + 81 = 328. You would write 328 in the empty box.

For each correct answer, you will earn $0.50. You will not be penalized for incorrect answers. You have 5 minutes to solve as many of the summations as you can. You will be told when time is up, but no time warnings will be issued.

When the experimenter instructs you to do so, please turn to the next page and begin.

Online Appendix C: Experiment Screenshots

Instructions

Today, you will be asked to make educated guesses about how people performed on two tasks in an experiment conducted recently here at the Vanderbilt University Experimental Economics Lab.

In the previous experiment, participants completed two tasks. All participants completed Task 1 first and could take as much time as they wanted to make two decisions. You will be asked about **one** of these decisions. Task 2 was a timed math exercise. Participants were paid for both tasks plus a $5 show-up fee.

We will now hand out a copy of the instructions used in this previous experiment.

Your Payment and Anonymity

In this experiment, your payment will be based on a lottery in which you receive either $15 or $0. The likelihood that you receive the larger amount of $15 is determined by how accurate your educated guess is compared to the actual outcome. (If you are interested, the lottery system is carefully designed so that it is mathematically optimal to submit your best guess about the median outcome.) So, **it is in your best interest to submit your true best guess.**

The decisions you make in this experiment are completely anonymous. Your identity will not be linked in any way to your decisions in this experiment. Your decisions are linked to your participant ID for payment purposes, but there is no record that matches your participant ID to your name.

An Example

In this example, you are asked to make an educated guess about which geographic location is closer to Vanderbilt University. You will not be paid for this example; it is only to ensure that you understand how to make your guess.

You must guess **which geographic location is closer** to our location at Vanderbilt University and **how much closer** it is.

Which is closer to our location at Vanderbilt University: the Titans Stadium or the Mall at Green Hills?

Titans Stadium	**I believe that they are the same distance**	Mall at Green Hills

Suppose you believe the Titans Stadium is 1.5 miles closer to our location at Vanderbilt University than the Mall at Green Hills. You would move the slider in to the section that says "Titans Stadium" until it says "1.5."

Titans Stadium	**I believe that the Titans Stadium is closer by 1.5 miles**	Mall at Green Hills

A chart shows your probability of winning $15 based on what the actual distance is.

If the actual distance is:	The Titans Stadium is closer by							The Mall at Green Hills is closer by					
	3+	**2.5**	**2**	**1.5**	**1**	**0.5**	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3+**
You win $15 with probability:	75%	83%	92%	100%	92%	83%	75%	67%	58%	50%	42%	33%	25%

For example, if your guess is accurate and the Titans Stadium is 1.5 miles closer than the Mall at Green Hills, you win $15 for sure (100%). On the other hand, if the Titans Stadium is actually 0.5 miles closer, you have a 83% chance of winning $15. If the Mall at Green Hills is closer than the Titans Stadium by 1.5 miles, your chance of winning $15 falls to 50%.

As you move the slider, the chart will update to show the probabilities of winning $15 at each possible value of the actual distance. So, if you decided the Titans Stadium was actually 2 miles closer than the Mall at Green Hills, the chart would change when you moved the slider.

Titans Stadium	**I believe that the Titans Stadium is closer by 2 miles**	Mall at Green Hills

If the actual distance is:	The Titans Stadium is closer by							The Mall at Green Hills is closer by					
	3+	**2.5**	**2**	**1.5**	**1**	**0.5**	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3+**
You win $15 with probability:	83%	92%	100%	92%	83%	75%	67%	58%	50%	42%	33%	25%	17%

You will now have an opportunity to test the slider and make your guess. Remember, this example is just for practice and you will not be paid for the results.

An Example

In this example, you are asked to make an educated guess about which geographic location is closer to Vanderbilt University. You will not be paid for this example; it is only to ensure that you understand how to make your guess.

Which is closer to our location at Vanderbilt University: the Titans Stadium or the Mall at Green Hills?

Titans Stadium	**I believe that they are the same distance**	Mall at Green Hills

If the actual distance is:	The Titans Stadium is closer by								The Mall at Green Hills is closer by				
	3+	**2.5**	**2**	**1.5**	**1**	**0.5**	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3+**
You win $15 with probability:	50%	58%	67%	75%	83%	92%	100%	92%	83%	75%	67%	58%	50%

Next

Example Results

The Titans Stadium is actually 0.5 miles closer to our location at Vanderbilt University than the Mall at Green Hills. You would have won $15 with 92% probability.

Titans Stadium	**You guessed that they are the same distance**	Mall at Green Hills

If the actual distance were:	The Titans Stadium is closer by								The Mall at Green Hills is closer by				
	3+	**2.5**	**2**	**1.5**	**1**	**0.5**	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3+**
You win $15 with probability:	50%	58%	67%	75%	83%	92%	100%	92%	83%	75%	67%	58%	50%

Now you will make educated guesses that determine your payment in this experiment. Consider your choices carefully. One of the guesses you make will be randomly chosen by a computer to determine your payment. Each guess is equally likely to be selected but you will not know which guess is chosen for payment until the end of the experiment. It is in your best interest to treat each guess as if it is the one that determines your payment.

Task 1

A computer will randomly draw one man and one woman from the previous experiment. Consider the decision each of these individuals made in the role of Person 2 in Task 1. You must guess **which individual chose the larger amount in the role of Person 2** and **how much larger** that amount was. In other words, who chose a larger amount in response to "The smallest amount that I would accept from Person 1 is:" and how much larger was that amount?

Woman	**I believe that they chose the same amount**	Man

If the actual outcome is:	The woman chose a larger amount by											The man chose a larger amount by									
	$10	**$9**	**$8**	**$7**	**$6**	**$5**	**$4**	**$3**	**$2**	**$1**	**$0**	**$1**	**$2**	**$3**	**$4**	**$5**	**$6**	**$7**	**$8**	**$9**	**$10**
You win $15 with probability:	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%

Next

Task 2

A computer will randomly draw one man and one woman from the previous experiment. You must guess **which individual answered more of the math sums correctly** and **how many more.**

Woman	**I believe that they answered the same amount correctly**	Man

If the actual outcome is:	The woman answered more sums correctly by											The man answered more sums correctly by									
	10+	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10+**
You win $15 with probability:	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%

Next

More Instructions

In an earlier session of this experiment, participants made the same two choices you just made. They were given the same instructions and asked to make their best guess. You must now make educated guesses about what those participants chose as their guesses. Consider your choices carefully. Again, any one of your guesses could be randomly chosen to determine your payment in this experiment and each is equally likely.

Task 2, Man

A computer will randomly draw one man from a previous session of this experiment. You must guess what he reported as his guess when asked if the man or the woman **answered more of the math sums correctly** and **how many more.**

Woman	**I believe that the man guessed they answered the same amount correctly**	Man

If the man reported:	The woman answered more sums correctly by											The man answered more sums correctly by									
	10+	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10+**
You win $15 with probability:	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%

Next

Task 2, Woman

A computer will randomly draw one woman from a previous session of this experiment. You must guess what she reported as her guess when asked if the man or the woman **answered more of the math sums correctly** and **how many more.**

Woman	**I believe that the woman guessed they answered the same amount correctly**	Man

If the woman reported:	The woman answered more sums correctly by											The man answered more sums correctly by									
	10+	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10+**
You win $15 with probability:	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%

Next

Task 1, Man

A computer will randomly draw one man from a previous session of this experiment. You must guess what he reported as his guess when asked if the man or the woman **chose the larger number in the role of Person 2 in Task 1** and **how much larger.**

Woman	**I believe that the man guessed they chose the same amount**	Man

If the man reported:					The woman chose a larger amount by											The man chose a larger amount by					
	$10	**$9**	**$8**	**$7**	**$6**	**$5**	**$4**	**$3**	**$2**	**$1**	**$0**	**$1**	**$2**	**$3**	**$4**	**$5**	**$6**	**$7**	**$8**	**$9**	**$10**
You win $15 with probability:	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%

Next

Task 1, Woman

A computer will randomly draw one woman from a previous session of this experiment. You must guess what she reported as her guess when asked if the man or the woman **chose the larger number in the role of Person 2 in Task 1** and **how much larger.**

Woman	**I believe that the woman guessed they chose the same amount**	Man

If the woman reported:					The woman chose a larger amount by											The man chose a larger amount by					
	$10	**$9**	**$8**	**$7**	**$6**	**$5**	**$4**	**$3**	**$2**	**$1**	**$0**	**$1**	**$2**	**$3**	**$4**	**$5**	**$6**	**$7**	**$8**	**$9**	**$10**
You win $15 with probability:	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%

Next

Demographics

Please complete the following brief demographic survey.

Gender:

Ethnicity:

Age:

Degree Program:

Major:

GPA (leave blank if you are a freshman):

Mother's Education:

Father's Education:

First language:

Next

Results

Your payment will be based on the following choice:

A computer will randomly draw one man and one woman from the previous experiment. You must guess **which individual answered more of the math sums correctly** and **how many more.**

Woman	**You guessed that the man answered more sums correctly by 5**	Man

If the actual outcome were:	The woman answered more sums correctly by											The man answered more sums correctly by									
	10+	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10+**
You win $15 with probability:	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%	95%	90%	85%	80%	75%

You guessed that the man answered more sums correctly by 5 and the actual outcome was the woman answered more sums correctly by 5, so you have a 50% chance of winning $15.

When you press this button, a random number between 0 and 100 will be chosen. If that number is less than 50 (your percent chance of winning), you win $15. Otherwise, you win $0.

Generate random number **20**

Next