Journal of Economic Behavior and Organization 210 (2023) 441–452

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Do women receive less blame than men? Attribution of outcomes in a prosocial setting☆

Nisvan Erkal[a], Lata Gangadharan[b,*], Boon Han Koh[c]

[a] Department of Economics, University of Melbourne, VIC 3010, Australia
[b] Department of Economics, Monash University, VIC 3800, Australia
[c] School of Economics and Centre for Behavioural and Experimental Social Science, University of East Anglia, NR4 7TJ, UK

ARTICLE INFO

Article history:
Received 12 July 2022
Revised 20 March 2023
Accepted 6 April 2023

JEL Classification:
C92
D91
J16

Keywords:
Gender biases
Beliefs
Attribution biases
Leadership
Social preferences
Laboratory experiments

ABSTRACT

We examine gender biases in the attribution of leaders' outcomes to their choices versus luck. Leaders make unobservable investment choices that affect the payoffs of group members. High investment is costly to the leader but increases the probability of an outcome with a high payoff. We observe gender biases in the attribution of low outcomes. Low outcomes of male (female) leaders are attributed more to their selfish decisions (bad luck). These biases are driven by male evaluators. We find no gender differences in the attribution of high outcomes.

1. Introduction

This paper examines whether outcomes of men and women are attributed differently in an environment where outcomes are determined by a combination of luck and unobservable actions motivated by prosocial preferences. In such an environment, we ask whether the outcomes of women are more likely to be attributed to luck or altruism than those of men. The specific context we consider relates to the evaluation of individuals in positions of power. Individuals in such roles, in both the public and the private sphere, make decisions which affect the wellbeing of others. For example, political leaders make policy choices that affect the welfare of their constituents. In many situations, they face a choice between maximizing their own payoffs and maximizing the payoffs of others.

☆ We would like to thank Billur Aksoy and participants at Monash Mini-Workshop of Motivated Beliefs (2021) for their comments and feedback. We gratefully acknowledge funding from the Australian Research Council (DP210102183), the Department of Economics at the University of Melbourne, and the Department of Economics at Monash University. This study has been approved by the Business and Economics Human Ethics Advisory Group at the University of Melbourne (ID 1544873).

* Corresponding author.
E-mail addresses: n.erkal@unimelb.edu.au (N. Erkal), lata.gangadharan@monash.edu (L. Gangadharan), b.koh@uea.ac.uk (B.H. Koh).

https://doi.org/10.1016/j.jebo.2023.04.003
0167-2681/

Gender diversity in leadership and management roles is of increasing interest to both policy makers and researchers. It is important to understand whether male and female leaders are evaluated differently since the evaluations are likely to affect individuals' decision-making processes. Traditional gender roles or stereotypes would suggest that women in positions of power may be expected to act in a more altruistic manner (see, e.g., Solnick 2001, Heilman and Chen 2005, Aguiar et al. 2008, Brañas-Garza et al. 2018). As a result, gender biases in the attribution of outcomes may still emerge, but gender stereotypes may lead to evaluations that are in favor of female leaders. That is, if female leaders are expected to be more prosocial, then it may be more likely that their successes are attributed to their altruistic decisions and their failures are attributed to bad luck.

We answer our core research question using a controlled laboratory experiment which allows us to elicit individuals' beliefs that are typically unobservable in practice. In our experiment, participants are divided into groups of three and make a series of investment decisions on behalf of the group. One individual is then randomly assigned to be the leader and their investment decisions are implemented and used to determine the payoffs of all group members. Specifically, for each investment decision, the leader chooses between a low and high investment option. High investment is more costly to the leader, but it is more likely to generate a high return (i.e., a high outcome) for the group. Hence, leaders face a tradeoff between maximizing their own expected payoff and those of the group members. The other group members serve as evaluators. They do not observe the leader's investment decision, but the leader's gender is revealed to them. Based on this information, the evaluators are first asked to report initial (prior) beliefs about the investment decision made by the leader. Subsequently, they are asked to report their updated (posterior) beliefs given the outcome of the investment chosen by the leader.

Since leaders' decisions are motivated by prosocial preferences, the evaluators form beliefs about the altruism of the leaders. Our main research question is whether the evaluators' beliefs and their updating process based on the outcome are influenced by the leader's gender. Specifically, in the updating process, the leader's gender may affect whether the evaluators attribute high (low) outcomes to altruism (selfish behavior) or good luck (bad luck).

Evaluators' beliefs in our experiment can be thought of as representing citizens' or the public's assessment of leaders both in private and public sectors. Due to society's growing demand for social responsibility, those in leadership roles are under more scrutiny than ever before. As a result, leaders across the board are expected to be aware of their impact on social welfare and engage in more prosocial activities. That leaders impact social welfare is perhaps easier to grasp in the case of public managers, who serve the public by ensuring good governance, planning for cities, regulating industries, promoting public health, providing security, etc. In the private sector, the increasing interest in corporate social responsibility has led to a growing awareness of the role that CEOs can play in driving social outcomes (Bénabou and Tirole, 2010). In addition to leadership, prosocial motivation plays an important role in many other jobs in the economy (Bowles and Polanía-Reyes, 2012; Besley and Ghatak, 2018). For example, teachers and doctors inherently assume responsibility for the outcomes of others.

Our results reveal that there are gender biases in the attribution of low outcomes but not high outcomes. Specifically, low outcomes of male leaders are attributed more to their selfish decisions while those of female leaders are attributed more to bad luck. Hence, we find a gender bias in the attribution of leaders' failure in a direction that is in favor of female leaders: they receive less blame than men after delivering a low outcome. On the other hand, both male and female leaders are believed to be equally altruistic following the news of a high outcome.

We observe that the gender biases are exhibited by male evaluators and, to some extent, by evaluators who are prosocial. This suggests that evaluators, particularly male evaluators, may be giving female leaders the benefit of the doubt after observing a low outcome, thus leading to a form of sexism that is motivated by good or prosocial intentions.

Our study contributes to the growing literature on the role that beliefs play in gender discrimination. This literature has shown that gender stereotypes can affect both ego-related beliefs and beliefs held about others. Coffman (2014), Bordalo et al. (2019), and Coffman et al. (2020) focus on how gender stereotypes influence the prior and updated beliefs individuals hold about their own ability. Bohren et al. (2019), Fenske et al. (2020), Campos-Mercade and Mengel (2023), Coffman et al. (2021) and Barron et al. (2022) focus on biases in the beliefs held about others.[1] We contribute to these second set of studies by examining biases in the attributions of leaders' outcomes through the lens of evaluators' belief-updating process.

Our findings also contribute to the growing literature on gender biases in the evaluation of leaders or individuals in positions of power. In this literature, several studies have shown that female leaders are evaluated more negatively than male leaders (see, e.g., Eagly et al. 1992, Grossman et al. 2019, Chakraborty and Serra 2021, Funk et al. 2021, Abel 2022).[2] In a similar vein, researchers have also shown that followers are more likely to follow the lead of male leaders than female leaders (see, e.g., Andreoni and Petrie 2008, Gangadharan et al. 2016, Reuben and Timko 2018, Gangadharan et al. 2019,

[1] See also the literature in psychology which has examined attribution biases about others in addition to self (e.g., Miller and Ross 1975). Findings present evidence of gender differences in the attribution of outcomes of others to effort, skill, and luck (e.g., Swim and Sanna 1996).

[2] Gender biases in evaluation also exist in other contexts. For instance, Brooks et al. (2014) find that investors prefer entrepreneurial pitches presented by male entrepreneurs to those by female entrepreneurs. Sarsons (2019) documents gender biases in physicians' referrals to surgical specialists by showing that the referring physicians evaluate patient outcomes differently depending on the performing surgeon's gender. Sarsons et al. (2021) report that women receive less credit for group work when employers cannot perfectly observe their contributions.

Ayalew et al. 2021).[3] We contribute to this literature by specifically focusing on an environment where prosocial preferences (and not ability) of leaders shape their decisions and outcomes. This is an important paradigm to investigate as leadership often involves decision making for others, and prosocial motivation is regarded as an important attribute of leaders (Bénabou and Tirole, 2010). In this environment, consistent with stereotypes of women being more prosocial than men, we show that female leaders receive less blame than male leaders for their low outcomes.

Finally, the setting we consider is similar to the one modeled in Erkal et al. (2022) where decision makers' actions may be influenced by prosocial motives. However, while Erkal et al. (2022) focus on attribution biases, asymmetry in the attribution of outcomes, and the consensus effect, the current paper is specifically designed to investigate whether attribution biases vary by gender.[4] We find evidence of biases that seem to be in favor of women that are driven by men.[5]

2. Experimental design

The main task in the experiment is the leadership task, which is based on the design used in Erkal et al. (2022). Participants are divided into groups of three. Each group has one leader and two evaluators. In the first stage of the task, all participants make a series of investment decisions, and they are informed that their decisions will be implemented for their group if they are selected as the leader. Next, participants are informed of their assigned group and role, as well as the gender of their group leader. In the second stage of the task, the evaluators report their beliefs about the leader's investment decisions in the first stage. We explain each stage of the design in further detail below.[6]

2.1. Experimental details

Stage 1: Participants make investment decisions. In Stage 1, participants are asked to make investment decisions for the group, assuming that they have been assigned to be the leader. The leader of the group is provided with 300 ECU to cover the cost of the chosen investment. The leader bears the cost of the investment, but the investment return is shared equally by the leader and evaluators. There are two investment options: Investment X, which represents a high investment decision, and Investment Y, which represents a low investment decision. Both investments can either succeed (leading to a high return) or fail (leading to a low return). Investment X costs the leader 200 ECU and succeeds with a probability of 0.75. Investment Y costs the leader 50 ECU and succeeds with a probability of 0.25.

Each participant makes decisions in five different investment tasks, where the returns to the two investment options vary across the tasks (see Table 1).[7] In all five tasks, a similar trade-off exists: the expected return is always higher for each evaluator if the leader chooses Investment X, but the expected return is higher for the leader if s/he chooses Investment Y. Consequently, participants' prosocial preferences are a key determinant of their investment decisions as leaders in Stage 1. Participants are informed that the evaluators will only observe the realized outcome of the investment chosen by the leader, but not the actual investment that is chosen.

Between Stages 1 and 2: Revealing participants' roles and gender. After all participants have made their decisions, their groups and roles are revealed to them. Specifically, on their individual computer screens, each participant is informed of their group number and whether they have been assigned to be the leader or the evaluator.

Next, the leader's gender is revealed to each group. The experiment follows a between-subject treatment design where the main treatment variation is the gender of the group leader. We reveal the leader's gender to the group in a salient but non-intrusive manner, following the protocol used by Bordalo et al. (2019). For each group, we call out the last three digits of the leader's six-digit ID number, and the group leader is asked to announce "Here!" loudly and clearly. Hence, the leader's gender is revealed to the group through the leader's voice. Participants sit in individual cubicles separated by high partitions, and the leader's identity is never revealed to the group members.

Finally, for the remainder of the experiment, we use the appropriate gender-specific pronoun (e.g., "he", "him", "she", "her", etc.) on the decision screens whenever we refer to the leader of the group. This allows us to further reinforce information about the leader's gender for the rest of the experiment.[8]

Stage 2: Evaluators report beliefs about leader's decisions. For each of the five investment tasks they faced in Stage 1, evaluators are asked to report their beliefs about the likelihood that the leader has chosen Investment X (high investment).

---

[3] It is also important to understand how the biases leaders face or expect to face in the evaluation process may impact their decision-making process. For example, Chakraborty and Serra (2021) show that women may be reluctant to become leaders due to an aversion to backlash. See Eckel et al. (2021) for a recent review of the gender leadership gap.

[4] Erkal et al. (2022) also explore whether belief updating depends on the process by which the decision maker is selected. For instance, it may be the case that evaluators are more likely to blame decision makers for low outcomes if they are randomly selected and not appointed by the group.

[5] Erkal et al. (2023) consider a different decision-making environment where evaluators can choose to make discretionary payments to leaders. They investigate the role beliefs versus outcomes play in the determination of the payments made to leaders and find gender differences in the criteria used to determine the payments.

[6] Instructions can be found in Appendix A. In the experiment, we refer to the evaluators as "Members".

[7] The tasks are presented in a random order, and this randomization is done at the session level.

[8] Based on unincentivized responses in the post-experimental questionnaire, about 85% of evaluators were able to correctly recall the leader's gender, and this proportion does not differ by the leader's gender or the evaluator's gender (Fisher's exact tests: p-values = 0.720 and 1.000, respectively). We analyze the data based on evaluators' beliefs of the leader's gender.

Table 1
Investment tasks.

Task	Net Payoff (ECU) in Stage 1					
	Investment X (High Investment)			Investment Y (Low Investment)		
	Succeeds	Fails	Expected	Succeeds	Fails	Expected
Task 1						
Leader	250	100	212.5	400	250	287.5
Each Evaluator	150	0	112.5	150	0	37.5
Task 2						
Leader	300	100	250	450	250	300
Each Evaluator	200	0	150	200	0	50
Task 3						
Leader	350	100	287.5	500	250	312.5
Each Evaluator	250	0	187.5	250	0	62.5
Task 4						
Leader	350	150	300	500	300	350
Each Evaluator	250	50	200	250	50	100
Task 5						
Leader	400	150	337.5	550	300	362.5
Each Evaluator	300	50	237.5	300	50	112.5

Only the returns of both investments to the leader and each evaluator vary across the tasks. The costs of each investment (200 ECU for Investment X and 50 ECU for Investment Y) are fixed for all five tasks. Similarly, the probabilities of each investment succeeding (0.75 for Investment X and 0.25 for Investment Y) are fixed for all five tasks. In addition to the return from the investment, the net payoff to the leader also includes his/her endowment (300 ECU) and the cost of the chosen investment. To illustrate, if the leader chooses Investment X in Task 1, then the cost of 200 ECU is deducted from the leader's endowment of 300 ECU, and the investment provides a return of 150 ECU if it succeeds (75% chance) and 0 ECU if it fails (25% chance). Hence, the expected net payoff for the leader in Stage 1 if s/he chooses Investment X is given by 300 (endowment) – 200 (cost) + $(0.75 \times 150 + 0.25 \times 0)$ (expected return from Investment X) = 212.5 ECU, while the expected net payoff for each evaluator is given by $(0.75 \times 150 + 0.25 \times 0)$ (expected return from Investment X) = 112.5 ECU.

For each task, we elicit two sets of beliefs. First, we ask evaluators to report their prior beliefs of the likelihood that the leader has chosen Investment X. Second, we ask evaluators to report their posterior beliefs conditional on each possible outcome of the investment. That is, evaluators are asked to state their beliefs on the likelihood that the leader has chosen Investment X both conditional on the investment succeeding (i.e., a high outcome was attained for the group) and conditional on the investment failing (i.e., a low outcome was attained for the group).

Evaluators' reported beliefs are incentivized using the Binarized Scoring Rule (Hossain and Okui, 2013; Erkal et al., 2020). Evaluators are paid only for their prior beliefs or for their posterior beliefs. If they are paid for their posterior beliefs, then they are paid for the reported posterior belief conditional on the outcome which matches the actual realized outcome of the investment.

2.2. Experimental procedures

The experiments were conducted at the Experimental Economics Laboratory at the University of Melbourne (E^2MU) and programmed using z-Tree (Fischbacher, 2007). Equal number of male and female participants were recruited using ORSEE (Greiner, 2015) and consisted mainly of university students.

Before arriving at the session, participants were invited to complete a pre-experimental questionnaire on Qualtrics. The questionnaire elicited participants' demographic characteristics, including their gender. Each participant was then given a randomly generated six-digit ID number, where the range of the first digit was determined based on the participant's reported gender (i.e., between 1 and 4 for a male participant, and between 5 and 8 for a female participant). At the beginning of each session, participants were first asked to enter the six-digit ID number given to them. This provided us with information about each participant's gender, and we were able to achieve gender balance in the allocation of roles in the leadership task. Specifically, we had an equal number of males and females assigned as leaders, and roughly an equal number of males and females assigned as evaluators.

Participants were provided with printed instructions and asked to answer a set of control questions to reinforce their understanding of the instructions for the leadership task. After they completed the leadership task, they also participated in a simple one-shot dictator game in groups of two. Each participant was given 300 ECU and asked to decide how to divide this endowment within the pair. One decision within each pair was randomly chosen to determine the final allocation between the players. Participants then completed a post-experimental questionnaire that included questions relating to their decisions in the experiment, a cognitive reflection task (CRT), and an incentivized risk-elicitation task.

Participants were paid for either one randomly chosen investment task in the leadership task, or for their decisions in the dictator game. If they were paid for the investment task, then the leader was paid based on his/her decision in Stage 1.

To minimize hedging, each evaluator was paid either for the leader's decision in Stage 1 or for their reported beliefs in Stage 2. In addition, participants received a fixed payment of 7 AUD for completing the pre-experimental questionnaire, as well as payments for their decisions in the CRT and the risk task. Each session lasted between 90 and 120 min, and participants earned 39.51 AUD on average.

In total, we analyze data from 350 participants.[9] Our main analysis will focus on evaluators' belief-updating behavior (see the econometric framework presented in Section 3.2). Based on simulations using data from Erkal et al. (2022), our sample size allows us to detect a difference of about 0.25–0.3 in the attribution of outcomes (i.e., the estimated values of γ_H and γ_L in Eq. (3) of Section 3.2) between treatment groups.

3. Research strategy

3.1. Conceptual framework

In this section, we present a conceptual framework to examine both the leaders' decisions and the evaluators' beliefs given the outcomes of these decisions.

We assume that leaders are differentiated based on their other-regarding preferences, denoted by $\beta_i \in [0, 1]$. β_i is a private draw from a distribution $F(\beta)$, which is common knowledge and has density $f(\beta)$. Since leaders make investment decisions (by choosing either Investment X or Y) on behalf of their groups, β affects their decision. We assume that the cost of investment is given by $c \in \{c_X, c_Y\}$, where $c_X > c_Y$, and is deducted from the leader's endowment ω.

For a given outcome Q, the utility of the leader is given by

$$U = u\left(\frac{Q}{N} + \omega - c\right) + \beta \cdot \sum_j v_j\left(\frac{Q}{N}\right), \tag{1}$$

where, $u(\cdot)$ is the direct utility the leader receives from their own monetary payoff and $v_j(\cdot)$ represents the utility evaluator j receives from their own monetary payoff. Leaders are expected utility maximizers and will choose Investment X if $EU(X) \geq EU(Y)$. Given the choice of parameters in our design, leaders face a trade-off between maximizing their own expected payoff and that of the evaluators. Choosing Investment X is more costly to the leader, but it provides a higher expected return for all individuals in the group. Hence, it is straightforward to show that there is a cutoff β^* such that the leader chooses Investment X if $\beta \geq \beta^*$. That is, leaders will choose Investment X if they care sufficiently about the evaluators' payoffs.

In the experiment, we elicit evaluators' beliefs about the likelihood that the leader has chosen Investment X. This is equivalent to eliciting their beliefs about the leader's social preferences (i.e., the probability that $\beta \geq \beta^*$). Formally, let μ_j denote evaluator j's prior belief that the leader has chosen Investment X. Since the outcome of the investment Q serves as a signal about the leader's decision (and their social preferences), let $\sigma_j(X|Q)$ be evaluator j's posterior belief given the outcome Q. Theoretically, evaluators will update their beliefs in accordance with Bayes' rule, which implies that evaluator j's posterior beliefs given a high and a low outcome are $\sigma_j(X|Q_H) = \frac{\mu_j p}{\mu_j p + (1-\mu_j)(1-p)}$ and $\sigma_j(X|Q_L) = \frac{\mu_j (1-p)}{\mu_j (1-p) + (1-\mu_j) p}$, respectively.

Our main research question is to examine whether the evaluators' beliefs about the decisions of the leaders, which are motivated by prosocial preferences, depend on the leaders' gender. In our setup, both the prior and the posterior beliefs of the evaluators are likely to be shaped by the stereotypes which exist about the prosociality of women versus men (see, e.g., Solnick 2001, Heilman and Chen 2005, Aguiar et al. 2008, Brañas-Garza et al. 2018, Cason et al. 2022).[10] In their survey papers, Croson and Gneezy (2009), Niederle (2016) and Bilén et al. (2021) conclude that the relative prosociality of women is a context-dependent phenomenon. If female leaders are perceived to be more prosocial than male leaders in our environment, then evaluators may hold more favorable prior beliefs about female leaders than male leaders. In addition, the outcomes of female leaders may be evaluated more favorably as compared to their male counterparts. For example, it may be the case that high outcomes of female leaders are attributed more to prosocial decisions as compared to those of male leaders, while their low outcomes are attributed more to luck.

Moreover, biases in the attribution of outcomes may not only depend on the leader's gender, but also on the evaluator's gender. On the one hand, due to homophily, it may be that female evaluators treat female leaders more favorably. In this case, female evaluators would be more likely to evaluate the outcomes of female leaders more favorably as compared to those of male leaders.[11] On the other hand, following discriminatory social norms, women may treat other women more

[9] 354 participants took part in the experiment. However, two participants (one evaluator and one leader from two different sessions) misreported the ID number they received from the pre-experimental questionnaire. We were therefore unable to correctly match their demographic information (including gender) to the experimental data. For the evaluator, we simply dropped their data point. For the leader, because the pronouns used during the experiment reflected an incorrect gender, we dropped the data points for the entire group.

[10] Bordalo et al. (2019) and Coffman et al. (2020) show that stereotypes influence both belief formation and belief updating in the context of self-evaluation.

[11] Recent evidence by Cappelen et al. (2019) also shows that female spectators are more likely to consider males to have a productivity advantage and therefore evaluate their failure less favorably.

harshly.[12] In this case, female evaluators would instead evaluate the outcomes of female leaders less favorably as compared to those of male leaders. In our analysis, we will therefore also examine the role of evaluators' gender in driving biases in the attribution of outcomes.

3.2. Estimation approach

As highlighted in the previous section, we assume that evaluators abide by Bayes' rule when revising their beliefs after observing the leader's outcome. Note that we can also express Bayes' rule in a log-odds form as follows:

$$\text{logit}\big(\sigma_j(X|Q)\big) = \text{logit}\big(\mu_j\big) + \ I(Q = Q_H) \cdot \text{logit}(p) + \ I(Q = Q_L) \cdot \text{logit}(1-p), \tag{2}$$

where, $\text{logit} z = \log(\frac{z}{1-z})$, $I(\cdot)$ is an indicator function for the observed output or return Q from the investment. We next augment Eq. (2) in the following way:

$$\text{logit}\big(\hat{\sigma}_j(X|Q)\big) = \boldsymbol{\delta}\text{logit}\big(\hat{\mu}_j\big) + \boldsymbol{\gamma_H}\ I(Q = Q_H) \cdot \text{logit}(p) + \boldsymbol{\gamma_L}\ I(Q = Q_L) \cdot \text{logit}(1-p) + \varepsilon_j \tag{3}$$

where, ε_i captures non-systematic errors.

The specification in Eq. (3) nests the theoretical Bayesian benchmark as a special case when $\delta = \gamma_H = \gamma_L = 1$.[13] The parameters δ, γ_H, and γ_L therefore allow us to capture the emphasis that evaluators place on prior beliefs, high outcomes, and low outcomes, respectively, when updating their beliefs. Specifically, in the context of our experiment, if γ_H (γ_L) is higher, then this implies that the evaluator is more likely to attribute the leader's high (low) outcome to the leader's decision. If γ_H (γ_L) is lower, then the leader's high (low) outcome is attributed more to luck. Our goal is to examine whether there are gender differences in the attribution of outcomes (as represented by γ_H and γ_L).

4. Results

Our main research question is to understand how the high and low outcomes of male and female leaders are evaluated. To answer this question, we study whether evaluators attribute male and female leaders' outcomes to their investment decisions or luck. In all our analyses, belief is a variable that takes an integer value in [0, 100], where a higher belief implies that the evaluator thinks the leader is more likely to have chosen high investment.

4.1. Evaluators' beliefs

In general, the evaluators in our sample behave as predicted by Bayes' rule in that posterior beliefs for high outcomes are on average higher than those for low outcomes. Specifically, the mean posterior beliefs across rounds are 51.2% and 31.9% for high outcomes and low outcomes, respectively.[14]

Fig. 1 shows evaluators' posterior beliefs given a low outcome (panel a) and a high outcome (panel b) against the Bayesian posteriors beliefs as derived using evaluators' prior beliefs. In each panel, we present the graphs separately for male leaders (gray solid line) and female leaders (black dotted line). The dashed 45° line represents the situation where evaluators' posteriors fully coincide with the Bayesian benchmark. The graphs are plotted based on the estimates reported in Table B1 of the Online Appendix (see, e.g., Eil and Rao 2011 for a similar estimation approach).

The figure reveals several key insights. First, evaluators generally do not conform to the Bayesian benchmark, as evidenced by departures from the 45-degree line in both panels. Second, panel (a) reveals that while evaluators' posterior beliefs for low outcomes are lower than those predicted by Bayes' rule for both male and female leaders, their posterior beliefs for male leaders are much lower than those for female leaders (p-value = 0.067 in column 4 of Table B1). Third, panel (b) reveals that evaluators' belief-updating patterns in response to high outcomes are almost identical between male and female leaders. We investigate these findings further by estimating Eq. (3), which allows us to disentangle whether departures from the Bayesian benchmark are driven by evaluators' treatment of their prior beliefs and/or their treatment of the leader's outcomes.

Table 2 presents ordinary least squares (OLS) estimates of Eq. (3) both at the pooled level (column 1) and separately by the leader's gender (columns 2 and 3). In the last column, we report p-values from tests of differences in the estimated

[12] That is, gender discrimination may be a social norm that female evaluators internalize or conform to. See the literature on the queen bee effect for a discussion of this type of behavior (e.g., Derks et al. 2016, Arvate et al. 2018).

[13] Hence, we can use this framework to examine systematic deviations from the Bayesian benchmark. See, e.g., Grether (1980), Ambuehl and Li (2018), Buser et al. (2018), Coutts (2019), Barron (2021), Erkal et al. (2022) and Möbius et al. (2022). Benjamin (2019) provides a review.

[14] As shown in Fig. C1 of Appendix C, some evaluators in our sample update their beliefs inconsistently i.e., in the opposite direction to that predicted by Bayes' rule) or not at all (i.e., have posterior beliefs equal to prior beliefs). The inclusion of these observations in the analysis may result in biased or incorrect conclusions, particularly if these evaluators are reporting beliefs that do not genuinely reflect their true posterior beliefs. Hence, we exclude an evaluator in our main analysis if 25% or more of their posterior beliefs are in the opposite direction to that predicted by Bayes' rule (inconsistent), or if all their posterior beliefs are equal to their prior beliefs (non-updaters). This corresponds to 23.5% and 8.1% of the sample, respectively, which is largely in line with what has been previously found in the literature (Coutts, 2019; Barron, 2021; Erkal et al., 2022; Möbius et al., 2022). In Appendix C, we present as a robustness check the analysis with the full sample. Our main conclusions remain largely unchanged. Further, our results remain robust to alternative specifications where we only exclude the inconsistent updaters.

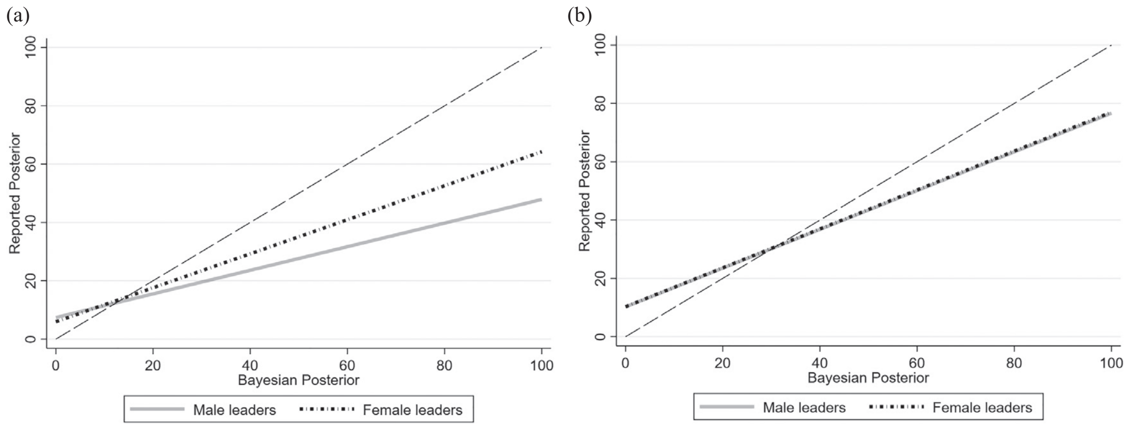

Fig. 1. Evaluators' posterior beliefs that the leader has chosen high investment against Bayesian posteriors, by the leader's outcomes. (a) Low outcomes, (b) High outcomes.

Table 2
OLS regressions of evaluators' posterior belief that the leader has chosen high investment, at pooled level and by the leader's gender.

	Dependent variable: Logit(posterior belief)			
Variables	Pooled (1)	Male Leader (2)	Female Leader (3)	(2) vs. (3) p-value [q-value]
δ: Logit(prior belief)	0.624	0.598	0.658	0.472
	(0.041)	(0.053)	(0.063)	[0.364]
γ_H: High outcome $\times$ logit(p)	0.783	0.854	0.725	0.582
	(0.117)	(0.210)	(0.104)	[0.365]
γ_L: Low outcome $\times$ logit($1-p$)	1.064	1.287	0.843	0.018**
	(0.094)	(0.150)	(0.110)	[0.042**]
Observations	1600	780	820	
R-squared	0.536	0.530	0.553	

Robust standard errors clustered at the participant level in parentheses. This analysis excludes participants classified as inconsistent or non-updaters. Since the regression specification estimates parameters of an augmented Bayes' rule, no controls can be included as the presence of any controls would invalidate the interpretation of the parameters. Moreover, since $I(\text{High Outcome}) + I(\text{Low Outcome}) = 1$, there is no constant term in the regression. q-values reported in brackets correct for multiple hypotheses testing using Anderson (2008)'s method. For tests of coefficients = 0: *** p<0.01 / q<0.01, ** p<0.05 / q<0.05, * p<0.10 / q<0.10.

coefficients by the leader's gender, as well as p-values adjusted for multiple hypotheses testing (referred to as q-values) in brackets. The q-values show that our results overall are robust to the correction of p-values for multiple hypotheses testing.

Column (1) reveals that evaluators suffer from base-rate neglect (test of $\delta = 1$: p-value < 0.001), attribute high outcomes more to luck than a Bayesian would on average (test of $\gamma_H = 1$: p-value $= 0.064$), but they are no different from a Bayesian in their attribution of low outcomes (test of $\gamma_L = 1$: p-value $= 0.495$). Consequently, there is an overall asymmetry in the attribution of outcomes, where, on average, the leader's low outcomes are attributed more to their investment decisions and high outcomes are attributed more to luck (test of $\gamma_H = \gamma_L$: p-value $= 0.083$). This result of an overall asymmetry in the attribution of high and low outcomes is consistent with Erkal et al. (2022).

We next focus on our core research question and explore whether the attribution of outcomes depends on the leader's gender. Columns (2) and (3) reveal that there are no gender differences in the evaluation of high outcomes (p-value $= 0.582$). The direction of the estimates of γ_H in both columns also suggests that the high outcomes of male and female leaders are on average attributed to luck relative to a Bayesian. While this difference is not statistically significant for male leaders, it is statistically significant for female leaders (tests of $\gamma_H = 1$ in columns 2 and 3: p-values $= 0.489$ and 0.010, respectively).

On the other hand, the estimates in columns (2) and (3) reveal that there are gender differences in the evaluation of low outcomes (p-value $= 0.018$). Moreover, evaluators attribute the low outcomes of male leaders more to their investment decisions than a Bayesian (test of $\gamma_L = 1$ in column 2: p-value $= 0.059$), but they are no different from a Bayesian in their attribution of the female leaders' low outcomes (test of $\gamma_L = 1$ in column 3: p-value $= 0.156$). Hence, even though we do not observe an overall bias in the attribution of low outcomes (column 1), the result at the pooled level

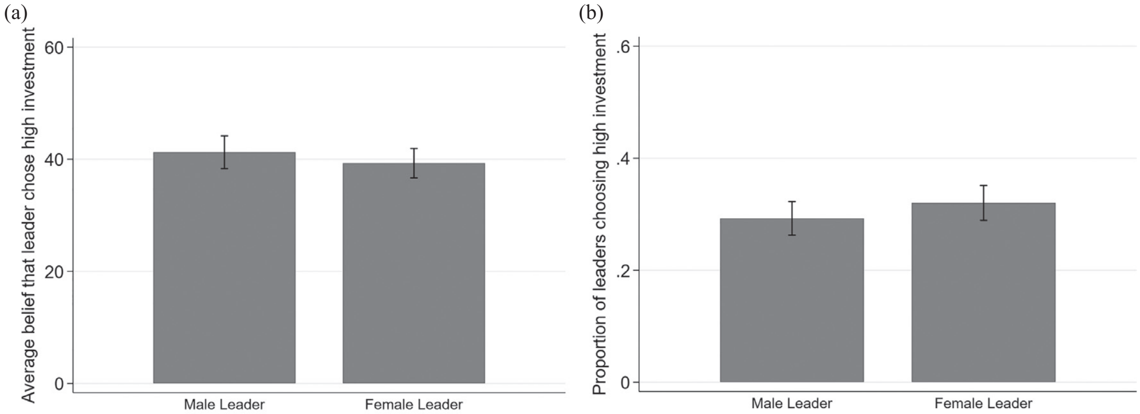

Fig. 2. Evaluators' prior belief that the leader has chosen high investment and leaders' investment decisions, by the leader's gender.
(a) Evaluator's average prior belief that leader has chosen high investment, (b) Proportion of high investment decisions by leaders.

masks a gender difference in the way low outcomes are attributed between male and female leaders. We summarize as follows.

Result 1. *Evaluators attribute the low outcomes of male leaders more to their selfish decisions than those of female leaders on average. However, evaluators are no different in their attribution of high outcomes between male and female leaders on average.*

Our results suggest that evaluators are more likely to believe that a low outcome from a male leader is the result of their selfish behavior while a low outcome from a female leader is the result of bad luck. Hence, evaluators are likely to hold male leaders more responsible than female leaders for low outcomes. This gender difference we observe in the updating behavior with respect to low outcomes is consistent with the stereotype of women being more prosocial than men. That is, the stereotype of women being more prosocial could affect the belief-updating process such that their low outcomes are attributed more to luck. However, we do not observe any statistically significant gender difference in the evaluation of high outcomes.

We next explore whether similar gender differences exist in evaluators' prior beliefs. Specifically, we ask whether it is the case that, a priori, evaluators expect women to be more prosocial. Panel (a) of Fig. 2 presents evaluators' average prior beliefs. The figure reveals that there are no statistically significant differences in evaluators' average prior beliefs toward male and female leaders (Wilcoxon rank-sum test: p-value = 0.465). Panel (b) shows the proportion of high investment decisions made by participants (in the role of leaders) in Stage 1. The figure reveals that consistent with the prior beliefs, there are no statistically significant differences between male and female leaders in their investment decisions on average (Fisher's exact test: p-value = 0.214).[15] We summarize as follows.

Result 2. *There is no evidence to suggest that gender stereotypes influence evaluators' prior beliefs. That is, there are no statistically significant differences between male and female leaders in terms of evaluators' prior beliefs. Similarly, there are no gender differences in leaders' investment decisions.*

Result 1 and Result 2 together imply that although there are no gender differences in evaluators' expectations, as reflected in their prior beliefs, about the prosociality of female and male leaders, there are gender differences in the attribution of low outcomes to luck versus selfish decisions. This suggests that even if gender stereotypes do not affect the prior beliefs, they may still emerge in the belief updating process.

4.2. Evaluators' gender and preferences

We next examine whether the attribution biases differ by the evaluator's gender. Table 3 presents OLS estimates of Eq. (3) separately for female evaluators (columns 1 and 2) and male evaluators (columns 3 and 4). The estimates reveal that both female and male evaluators are no different in their attribution of high outcomes between male and female leaders (p-values = 0.971 and 0.503, respectively).[16] On the other hand, while female evaluators are no different in their attribution of low outcomes between male and female leaders (p-value = 0.863), male evaluators are more likely to attribute the low outcomes of male leaders to their investment decisions as compared to those of female leaders (p-value = 0.003). We

[15] Our conclusions from the non-parametric tests are also supported by regression analyses, as reported in Tables B2 and B3 of Appendix B.

[16] On average, evaluators are also no different from a Bayesian in their attribution of the leader's high outcomes, with one exception being that male evaluators tend to attribute the high outcomes of female leaders more to luck relative to a Bayesian (test of $\gamma_H = 1$ in columns 1 to 4: p-values = 0.269, 0.284, 0.835, and 0.011, respectively).

Table 3
OLS regressions of evaluators' posterior belief that the leader has chosen high investment, by the leader's gender and evaluator's gender.

	Dependent variable: Logit(posterior belief)					
	Female Evaluator			Male Evaluator		
Variables	Male Leader (1)	Female Leader (2)	(1) vs. (2) p-value [q-value]	Male Leader (3)	Female Leader (4)	(3) vs. (4) p-value [q-value]
δ: Logit(prior belief)	0.681	0.621	0.652	0.533	0.693	0.097*
	(0.086)	(0.101)	[0.862]	(0.065)	(0.070)	[0.219]
γ_H: High outcome $\times$ logit(p)	0.812	0.820	0.971	0.915	0.630	0.503
	(0.168)	(0.166)	[1.000]	(0.404)	(0.138)	[0.845]
γ_L: Low outcome $\times$ logit($1-p$)	0.962	0.923	0.863	1.631	0.753	0.003***
	(0.150)	(0.177)	[1.000]	(0.262)	(0.128)	[0.016**]
Observations	410	420		370	400	
R-squared	0.619	0.560		0.479	0.551	

Robust standard errors clustered at the participant level in parentheses. This analysis excludes participants classified as inconsistent or non-updaters. Since the regression specification estimates parameters of an augmented Bayes' rule, no controls can be included as the presence of any controls would invalidate the interpretation of the parameters. Moreover, since $I(\text{High Outcome}) + I(\text{Low Outcome}) = 1$, there is no constant term in the regression.
q-values reported in brackets correct for multiple hypotheses testing using Anderson (2008)'s method. For tests of coefficients = 0: *** p<0.01 / q<0.01, ** p<0.05 / q<0.05, * p<0.10 / q<0.10.

also observe that relative to a Bayesian, male evaluators tend to attribute the low outcomes of male leaders more to their investment decisions (test of $\gamma_L = 1$ in column 3: p-value = 0.021) and the low outcomes of female leaders more to luck (test of $\gamma_L = 1$ in column 4: p-value = 0.060).[17] However, female evaluators are consistently no different from the Bayesian benchmark in their attribution of low outcomes regardless of the leader's gender (test of $\gamma_L = 1$ in columns 1 and 2: p-values = 0.803 and 0.664, respectively). Hence, the gender difference in the attribution of low outcomes we observe in Result 1 appears to be driven by male evaluators.[18]

We summarize our result as follows.

Result 3. *Male evaluators attribute the low outcomes of male leaders more to their investment decisions and those of female leaders more to luck, while female evaluators are no different in their attribution of male and female leaders' low outcomes. Both female and male evaluators are no different in their attribution of high outcomes between male and female leaders.*

Further, we examine whether the attribution biases differ by the evaluator's own social preferences. Because we elicit all participants' investment decisions as leaders in Stage 1 of the experiment, we can use these decisions as a proxy for their prosocial attitudes toward group members. Table 4 presents OLS estimates of Eq. (3) separately for evaluators who chose low investment (columns 1 and 2) and evaluators who chose high investment (columns 3 and 4).

The estimates in Table 4 reveal that the two type of evaluators are no different in their attribution of high outcomes between male and female leaders (p-values = 0.481 and 0.318, respectively).[19] Nonetheless, while evaluators who are less prosocial attribute leaders' high outcomes to luck relative to a Bayesian (p-values = 0.012 and 0.003 in columns 1 and 2, respectively), those who are more prosocial are no different from a Bayesian in their attribution of high outcomes (p-values = 0.313 and 0.871 in columns 3 and 4, respectively).[20]

Importantly, we find that the gender biases in the attribution of low outcomes appear to be exhibited by evaluators who are more prosocial. Specifically, columns (3) and (4) reveal that prosocial evaluators are more likely to attribute the low outcomes of female leaders to luck than those of male leaders (p-value = 0.002), while columns (1) and (2) reveal that evaluators who are less prosocial are no different in their attribution of low outcomes between male and female leaders (p-value = 0.481).[21] However, the difference-in-differences in the attribution of low outcomes based on both the leader's gender and the evaluator's prosociality is not statistically significant (p-value = 0.136). As a result, we conclude that there

[17] We also examine difference-in-differences in the attribution of leaders' outcomes based on both the leader's and evaluator's gender. Specifically, the difference-in-differences in the attribution of high outcomes is not statistically significant, while that for the attribution of low outcomes is statistically significant (p-values = 0.544 and 0.024, respectively).

[18] In addition to attribution biases, we can also examine how posterior beliefs differ by the evaluator's gender. Table B4 of the Online Appendix reports reduced-form regressions of evaluators' posterior beliefs against the leader's gender and outcomes. Consistent with the estimates in Table 3, columns (3) and (4) reveal that the posterior beliefs of male evaluators given a low outcome are statistically significantly lower for male leaders than for female leaders (p-values = 0.031 and 0.065, respectively). Note that since differences in average posterior beliefs do not distinguish between evaluator's emphasis on prior beliefs versus the leader's outcomes, we use Eq. (3) to examine departures from the Bayesian benchmark in detail.

[19] The difference-in-differences in the attribution of high outcomes based on the leader's gender and the evaluator's prosociality is not statistically significant either (p-value = 0.208).

[20] This difference in the attribution of high outcomes between prosocial and selfish evaluators is consistent with Erkal et al. (2022), who find that a consensus effect exists in the attribution of decision makers' outcomes.

[21] Similarly, relative to a Bayesian, more prosocial evaluators attribute low outcomes of female leaders more to luck and those of male leaders more to their actions (p-values = 0.019 and 0.043, respectively). However, less prosocial evaluators are no different from a Bayesian in their attribution of leaders' low outcomes (p-values = 0.218 and 0.887 in columns 1 and 2, respectively).

Table 4
OLS regressions of evaluators' posterior belief that the leader has chosen high investment, by the leader's gender and evaluator's investment decision.

	Dependent variable: Logit(posterior belief)					
	Evaluator Chose Low Investment			Evaluator Chose High Investment		
Variables	Male Leader (1)	Female Leader (2)	(1) vs. (2) p-value [q-value]	Male Leader (3)	Female Leader (4)	(3) vs. (4) p-value [q-value]
δ: Logit(prior belief)	0.562	0.669	0.272	0.609	0.457	0.534
	(0.076)	(0.061)	[0.319]	(0.072)	(0.232)	[0.501]
γ_H: High outcome $\times$ logit(p)	0.471	0.638	0.480	1.434	0.970	0.318
	(0.204)	(0.119)	[0.457]	(0.427)	(0.182)	[0.320]
γ_L: Low outcome $\times$ logit($1-p$)	1.290	0.980	0.256	1.407	0.496	0.002***
	(0.233)	(0.142)	[0.318]	(0.196)	(0.209)	[0.012**]
Observations	506	548		274	272	
R-squared	0.575	0.661		0.479	0.238	

Robust standard errors clustered at the participant level in parentheses. This analysis excludes participants classified as inconsistent or non-updaters. Since the regression specification estimates parameters of an augmented Bayes' rule, no controls can be included as the presence of any controls would invalidate the interpretation of the parameters. Moreover, since $I(\text{High Outcome}) + I(\text{Low Outcome}) = 1$, there is no constant term in the regression.
q-values reported in brackets correct for multiple hypotheses testing using Anderson (2008)'s method. For tests of coefficients = 0: *** p<0.01 / q<0.01, ** p<0.05 / q<0.05, * p<0.10 / q<0.10.

exists some suggestive evidence that prosocial evaluators are more likely to attribute male and female leaders' low outcomes differently.

5. Conclusion

In environments where the outcomes of leaders are determined by a combination of luck and unobservable actions, are outcomes of male and female leaders attributed differently by their evaluators? We answer this question in an environment where the actions taken by leaders affect both their own welfare and those of the group members (evaluators). Such environments are pervasive in many real-world settings both in the public and private domain. Gender biases in evaluations may emerge in these situations if, for example, women are expected to behave in a more prosocial manner.

We find an asymmetry between the evaluation of high outcomes and low outcomes. High outcomes of male and female leaders are not treated differently, suggesting that men and women are deemed to be equally altruistic (but less so than what a Bayesian would believe) after observing a high outcome. However, while the low outcomes of male leaders are attributed more to their selfish decisions, those of female leaders are attributed more to bad luck. Hence, in the case of failure, men are assigned more blame than women for being selfish. This is despite the fact that there are no gender differences in the evaluators' prior beliefs about male and female leaders' prosocial preferences. That is, evaluators start from a gender-neutral position, but they update their beliefs differently based on the gender of the leader when they observe a low outcome.

We find that this gender bias in the evaluation of low outcomes is driven by male evaluators and potentially, by evaluators who are prosocial. One interpretation of these results is that male evaluators may see the need to treat female leaders more favorably, thus giving them a greater benefit of the doubt in the face of failure. Interpreted in this way, one possible explanation for our findings is benevolent sexism (Glick and Fiske, 1996). Unlike hostile sexism, benevolent sexism tends to lead to behaviors toward women that are often characterized as prosocial.[22] However, such biases in the evaluation process may still lead to adverse outcomes for women. For example, gender biases in evaluations that favor women may hinder the development of their careers and increase the possibility of backlash against female leaders in the long run.[23] In general, gender biases are important to understand because they may lead to distortions in the incentives provided to all decision makers in positions of power (male and female) and may harm the future actions taken by them. Future research can shed light on the distortionary impact that such gendered evaluations can have on decision making by leaders and the consequent labor market outcomes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[22] Benevolent sexism stereotypes women as affectionate and delicate individuals who need to be protected (e.g., Glick and Fiske 1996, Glick et al. 2000). Researchers have argued that benevolent sexism can be as oppressive as hostile sexism (e.g., Glick and Fiske 2001). According to Glick and Fiske (1996), hostile sexism is behavior that is in line with the classic definition of prejudice by Allport (1954). Glick et al. (2000) show that while men are more likely than women to exhibit hostile sexism, both men and women are equally likely to exhibit and endorse benevolent sexism.

[23] Jampol and Zayas (2021) find that underperforming women are evaluated in a less truthful but kinder way (using "white lies") as compared with equally underperforming men.

Data availability

Data and experimental software are available at: https://github.com/boonhankoh/EGK-beliefs_gender.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2023.04.003.

References

Abel, M., 2022. Do workers discriminate against female bosses? J. Human Resour., 1120-11318R3.
Aguiar, F., Brañas-Garza, P., Cobo-Reyes, R., Jimenez, N., Miller, L.M., 2008. Are women expected to be more generous? Exp. Econ. 12 (1), 93–98.
Allport, G.W., 1954. The Nature of Prejudice. Addison-Wesley, Reading, MA.
Ambuehl, S., Li, S., 2018. Belief updating and the demand for information. Games Econ. Behav. 109, 21–39.
Anderson, M.L., 2008. Multiple inference and gender differences in the effects of early intervention: a reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. J. Am. Stat. Assoc. 103 (484), 1481–1495.
Andreoni, J., Petrie, R., 2008. Beauty, gender and stereotypes: evidence from laboratory experiments. J. Econ. Psychol. 29 (1), 73–93.
Arvate, P.R., Galilea, G.W., Todescat, I., 2018. The queen bee: a myth? The effect of top-level female leadership on subordinate females. Leadersh. Q. 29 (5), 533–548.
Ayalew, S., Manian, S., Sheth, K., 2021. Discrimination from below: experimental evidence from Ethiopia. J. Dev. Econ. 151.
Barron, K., 2021. Belief updating: does the 'good-news, bad-news' asymmetry extend to purely financial domains? Exp. Econ. 24, 31–58.
Barron, K., Ditlmann, R., Gehrig, S., Schweighofer-Kodritsch, S. (2022). Explicit and implicit belief-based gender discrimination: a hiring experiment. Working Paper.
Bénabou, R., Tirole, J., 2010. Individual and corporate social responsibility. Economica 77 (305), 1–19.
Benjamin, D.J., Bernheim, B.D., DellaVigna, S., Laibson, D., 2019. Errors in probabilistic reasoning and judgment biases. Handbook of Behavioral Economics: Applications and Foundations 2 (2:69-186), Eds. Elsevier.
Besley, T., Ghatak, M., 2018. Prosocial motivation and incentives. Annu. Rev. Econom. 10, 411–438.
Bilén, D., Dreber, A., Johannesson, M., 2021. Are women more generous than men? A meta-analysis. J. Econ. Sci. Assoc. 7, 1–18.
Bohren, J.A., Imas, A., Rosenberg, M., 2019. The dynamics of discrimination: theory and evidence. Am. Econ. Rev. 109 (10), 3395–3436.
Bordalo, P., Coffman, K., Gennaioli, N., Shleifer, A., 2019. Beliefs about gender. Am. Econ. Rev. 109 (3), 739–773.
Bowles, S., Polanía-Reyes, S., 2012. Economic incentives and social preferences: substitutes or complements? J. Econ. Lit. 50 (2), 368–425.
Brañas-Garza, P., Capraro, V., Rascón-Ramírez, E., 2018. Gender differences in altruism on mechanical turk: expectations and actual behaviour. Econ. Lett. 170, 19–23.
Brooks, A.W., Huang, L., Kearney, S.W., Murray, F.E., 2014. Investors prefer entrepreneurial ventures pitched by attractive men. Proc. Natl. Acad. Sci. 111 (12), 4427–4431.
Buser, T., Gerhards, L., van der Weele, J., 2018. Responsiveness to feedback as a personal trait. J. Risk Uncert. 56 (2), 165–192.
Campos-Mercade, P., Mengel, F., 2023. Non-Bayesian statistical discrimination. Management Science.
Cappelen, A.W., Falch, R., Tungodden, B. (2019). The Boy crisis: experimental evidence on the acceptance of males falling behind. Working Paper.
Cason, T.N., Gangadharan, L., Grossman, P.J., 2022. Gender, beliefs, and coordination with externalities. J. Public Econ. 104744, 214.
Chakraborty, P., Serra, D. (2021). Gender differences in top leadership roles: does aversion to worker backlash matter? Working Paper.
Coffman, K., 2014. Evidence on self-stereotyping and the contribution of ideas. Q. J. Econ. 129 (4), 1625–1660.
Coffman, K., Collis, M., Kulkarni, L. (2020). Stereotypes and Belief Updating. Working Paper.
Coffman, K., Exley, C.L., Niederle, M., 2021. The role of beliefs in driving gender discrimination. Manag. Sci. 67 (6), 3551–3569.
Coutts, A., 2019. Good news and bad news are still news: experimental evidence on belief updating. Exp. Econ. 22 (2), 369–395.
Croson, R., Gneezy, U., 2009. Gender differences in preferences. J. Econ. Lit. 47 (2), 448–474.
Derks, B., Van Laar, C., Ellemers, N., 2016. The queen bee phenomenon: why women leaders distance themselves from junior women. Leadersh. Q. 27 (3), 456–469.
Eagly, A.H., Makhijani, M.G., Klonsky, B.G., 1992. Gender and the evaluation of leaders: a meta-analysis. Psychol. Bull. 111 (1), 3–22.
Eckel, C., Gangadharan, L., Grossman, P.J., Xue, N., Chaudhuri, A., 2021. The gender leadership gap: insights from experiments. A Research Agenda For Experimental Economics, Ed. Edward Elgar.
Eil, D., Rao, J.M., 2011. The good news-bad news effect: asymmetric processing of objective information about yourself. Am. Econ. J. Microecon. 3 (2), 114–138.
Erkal, N., Gangadharan, L., Koh, B.H., 2020. Replication: belief elicitation with quadratic and binarized scoring rules. J. Econ. Psychol. 81, 102315.
Erkal, N., Gangadharan, L., Koh, B.H., 2022. By chance or by choice? Biased attribution of others' outcomes when social preferences matter. Exp. Econ. 25 (2), 413–443.
Erkal, N., Gangadharan, L., Koh, B.H. (2023). Gender biases in performance evaluation: the role of beliefs versus outcomes. Working Paper.
Fenske, J., Castagnetti, A., Sharma, K. (2020). Attribution bias by gender: evidence from a laboratory experiment. Working Paper.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.
Funk, K.D., Jensen, U.T., Molina, A.L., Stritch, J.M., 2021. Does leader gender matter for performance evaluations? Evidence from two experiments. Public Manag. Rev. 1–19.
Gangadharan, L., Jain, T., Maitra, P., Vecci, J., 2016. Social identity and governance: the behavioral response to female leaders. Eur. Econ. Rev. 90, 302–325.
Gangadharan, L., Jain, T., Maitra, P., Vecci, J., 2019. Female leaders and their response to the social environment. J. Econ. Behav. Organ. 164, 256–272.
Glick, P., Fiske, S.T., 1996. The ambivalent sexism inventory: differentiating hostile and benevolent sexism. J. Pers. Soc. Psychol. 70 (3), 491–512.
Glick, P., Fiske, S.T., 2001. An ambivalent alliance: hostile and benevolent sexism as complementary justifications for gender inequality. Am. Psychol. 56 (2), 109–118.
Glick, P., Fiske, S.T., Mladinic, A., Saiz, J.L., Abrams, D., Masser, B., Adetoun, B., Osagie, J.E., Akande, A., Alao, A., Brunner, A., Willemsen, T.M., Chipeta, K., Dardenne, B., Dijksterhuis, A., Wigboldus, D., Eckes, T., Six-Materna, I., Exposito, F., Moya, M., Foddy, M., Kim, H.J., Lameiras, M., Sotelo, M.J., Mucchi–Faina, A., Romani, M., Sakalli, N., Udegbe, B., Yamamoto, M., Ui, M., Ferreira, M.C., Lopez Lopez, W., 2000. Beyond prejudice as simple antipathy: hostile and benevolent sexism across cultures. J. Pers. Soc. Psychol. 79 (5), 763–775.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.
Grether, D.M., 1980. Bayes rule as a descriptive model: the representativeness heuristic. Q. J. Econ. 95 (3), 537–557.
Grossman, P.J., Eckel, C., Komai, M., Zhan, W., 2019. It pays to be a man: rewards for leaders in a coordination game. J. Econ. Behav. Organ. 161, 197–215.
Heilman, M.E., Chen, J.J., 2005. Same behavior, different consequences: reactions to men's and women's altruistic citizenship behavior. J. Appl. Psychol. 90 (3), 431–441.
Hossain, T., Okui, R., 2013. The binarized scoring rule. Rev. Econ. Stud. 80 (3), 984–1001.
Jampol, L., Zayas, V., 2021. Gendered white lies: women are given inflated performance feedback compared with men. Pers. Soc. Psychol. Bull. 47 (1), 57–69.
Miller, D.T., Ross, M., 1975. Self-serving biases in the attribution of causality: fact or fiction? Psychol. Bull. 82 (2), 213–225.

Möbius, M.M., Niederle, M., Niehaus, P., Rosenblat, T.S., 2022. Managing self-confidence: Theory and experimental evidence. Manag. Sci. 68 (11), 7793–7817.
Niederle, M., Kagel, J.H., Roth, A.E., 2016. Gender. The Handbook of Experimental Economics (2:481-562), Eds. Princeton University Press, New Jersey.
Reuben, E., Timko, K., 2018. On the effectiveness of elected male and female leaders and team coordination. J. Econ. Sci. Assoc. 4 (2), 123–135.
Sarsons, H. (2019). Interpreting signals in the labor market: evidence from medical referrals. Working Paper.
Sarsons, H., Gërxhani, K., Reuben, E., Schram, A., 2021. Gender differences in recognition for group work. J. Polit. Econ. 129 (1), 101–147.
Solnick, S.J., 2001. Gender differences in the ultimatum game. Econ. Inq. 39 (2), 189–200.
Swim, J.K., Sanna, L.J., 1996. He's skilled, she's lucky: a meta-analysis of observers' attributions for women's and men's successes and failures. Pers. Soc. Psychol. Bull. 22 (5), 507–519.