School of Economics Working Paper 2022-05

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

Nisvan Erkal *
Lata Gangadharan **
Boon Han Koh ***

*Department of Economics, University of Melbourne
**Department of Economics, Monash University
***School of Economics and Centre for Behavioural and Experimental Social Science, University of East Anglia

University of East Anglia

SCHOOL OF ECONOMICS

School of Economics
University of East Anglia
Norwich Research Park
Norwich NR4 7TJ
United Kingdom
www.uea.ac.uk/economics

Do women receive less blame than men? Attribution of outcomes in a prosocial setting[*]

Nisvan Erkal[†]
Lata Gangadharan[‡]
Boon Han Koh[§]

June 2022

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 a good outcome (high payoff). We observe gender biases in the attribution of bad outcomes. Bad outcomes of male (female) leaders are attributed more to their decisions (luck). These biases are driven by male evaluators and evaluators who are prosocial. We find no gender differences in the attribution of good outcomes. We conjecture that benevolent sexism may be driving our results.

Keywords: Gender biases; Beliefs; Attribution biases; Benevolent sexism; Social preferences; Laboratory experiments.
JEL Classification: C92, D91, J16

[*] We would like to thank 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).

[†] Department of Economics, University of Melbourne, VIC 3010, Australia. n.erkal@unimelb.edu.au.
[‡] Department of Economics, Monash University, VIC 3800, Australia. lata.gangadharan@monash.edu.
[§] Corresponding author. School of Economics and Centre for Behavioural and Experimental Social Science, University of East Anglia, NR4 7TJ, United Kingdom. b.koh@uea.ac.uk.

1 Introduction

This paper examines whether the outcomes of men and women are attributed differently in an environment where outcomes are determined by a combination of luck and unobservable actions. For instance, are the outcomes of women more likely to be attributed to luck than those of men? To answer this question, we focus on the evaluation of individuals in positions of power. Individuals in such roles 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 others.

We ask whether women in positions of power face different expectations than men in such roles. It is important to understand these expectations as they are likely to affect leaders' decision-making processes. Traditional gender roles or stereotypes would suggest that women in decision-making roles may be expected to act in a more altruistic manner (e.g., see Solnick, 2001; Heilman and Chen, 2005). In such an environment, gender biases in the attribution of outcomes may still emerge, but gender stereotypes may instead lead to evaluations that are in favor of female leaders. That is, if female leaders are expected to be more prosocial, then they may be evaluated in a more positive light for their outcomes, with their successes being attributed more to their actions and failures attributed more to luck.

We answer our core research question using a controlled laboratory experiment, which gives us the ability 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 can choose between a low investment and a high investment option. High investment is more costly to the leader, but it is more likely to generate a high return (i.e., a good outcome) for the group. Hence, leaders face a tradeoff between maximizing their own expected payoffs and those of the group members. The other group members serve as evaluators. They do not observe the leader's investment decisions, but the leader's gender is revealed to them. Based on this information, the evaluators report initial (prior) beliefs about the investment decisions made by the leader. They then report their updated (posterior) beliefs given the outcome of the investment chosen by the leader. To investigate the gender biases in the attribution of leaders' outcomes, we examine whether the evaluators' belief-updating process differs depending on whether they are faced with a male leader or a female leader.

Our results reveal that there are gender biases in the attribution of bad outcomes, but there are no gender differences in the attribution of good outcomes. Specifically, bad outcomes of male leaders are attributed more to their decisions while those of female leaders are attributed more to 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 for delivering bad outcomes. On the other hand, both male and female leaders receive similar credit for their good outcomes.

We observe that the gender biases are exhibited by male evaluators and evaluators who are prosocial. As a result, we conjecture that our findings may be driven by benevolent sexism. Evaluators, particularly male evaluators, may see a need to protect or defend female leaders by giving them the benefit of the doubt when observing a bad outcome, thus leading to a form of sexism that is motivated by good or prosocial intentions.[1]

Turning towards the implications of these results, the type of biases we detect in the evaluation process are important to understand because they may lead to differences in the type of feedback that leaders are given, thus denying female leaders the opportunity to receive the information they need to improve their performance in the future. They may also lead to gender differences in the incentives provided to leaders, which, in turn, may affect the choices that they make in the future.[2]

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, Collis, and Kulkarni (2020) focus on the role of gender stereotypes in driving individuals' prior and updated beliefs about their own ability. Bohren, Imas, and Rosenberg (2019), Fenske, Castagnetti, and Sharma (2020), Campos-Mercade and Mengel (2021), Coffman, Exley, and Niederle (2021), and Barron et al. (2022) focus on biases in the beliefs held about others.[3]

Our study contributes to this literature by examining biases in the attributions of leaders' outcomes through the lens of evaluators' belief-updating process. We specifically focus on an

[1] Benevolent sexism stereotypes women as affectionate and delicate individuals who need to be protected (e.g., Glick and Fiske, 1996; Glick et al., 2000). Although attitudes and behaviors which fall under benevolent sexism seem on the surface to favor women, researchers have argued that benevolent sexism can be as oppressive as hostile sexism (e.g., Glick and Fiske, 2001).

[2] For example, Erkal, Gangadharan, and Koh (2022b) investigate whether beliefs play a role in the determination of payments made to leaders. While they do not find biases in evaluators' beliefs, they still find gender differences in the payments which are driven by differences in the criteria used to determine these payments.

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

environment where prosocial preferences (and not ability) of leaders shape their decisions and outcomes.[4] This is an important paradigm to investigate gender biases in attributions as leadership often involves decision making for others, and prosocial motivation is regarded as an important attribute of leaders (Bénabou and Tirole, 2010). In such a domain, we find evidence of biases that seem to be in favor of women, and that these biases are driven by men. Hence, we uncover suggestive evidence that benevolent sexism may exist in our setup. Benevolent sexism has been shown to exist in other environments (e.g., Glick and Fiske, 1996; Glick et al., 2000), and it can be regarded as a more subtle form of discrimination. Importantly, despite the seemingly innocuous connotations of the term *benevolent*, benevolent sexism can prove to have long-term adverse consequences on women's labor market outcomes.

2 Experimental Design

The main task in the experiment is the leadership task, which is based on the design used in Erkal, Gangadharan, and Koh (2022a). 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.[5]

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.

[4] Our work builds on Erkal, Gangadharan, and Koh (2022a), who examine the relationship between the consensus effect and attribution biases in such an environment absent gender considerations.
[5] Instructions can be found in Appendix A. In the experiment, we refer to the evaluators as "Members".

Each participant makes decisions in five different investment tasks, where the returns to the two investment options vary across the tasks (see Table 1).[6] However, a similar trade-off exists across all five tasks: 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: Participants' roles and leader's gender are revealed. 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 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, 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.[7]

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).

---

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

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 good outcome was attained for the group) and conditional on the investment failing (i.e., a bad outcome was attained for the group).

Evaluators' reported beliefs are incentivized using the Binarized Scoring Rule (Hossain and Okui, 2013; Erkal, Gangadharan, and Koh, 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.

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 ECU 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 minutes, and participants earned 39.51 AUD on average.

In total, we analyze data from 350 participants.[8] Our main analysis will focus on evaluators' belief-updating behavior (see econometric framework presented in Section 3.2). Based on simulations using data from Erkal, Gangadharan, and Koh (2022a), our sample size allows us to detect a gender difference of about 0.25-0.3 in the attribution of outcomes (i.e., gender difference in the estimated values of γ_G and γ_B in equation 3 of Section 3.2).

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 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. $Q \in \{Q_L, Q_H\}$ is the outcome of the leader's chosen investment, providing a low return Q_L if the investment fails (bad outcome) and a high return Q_H if it succeeds (good outcome). This return is shared equally by the leader and the two evaluators. A high investment decision by the leader leads to a good outcome with probability p, while a low investment decision leads to a good outcome with probability $1 - p$. In our experiment, p = 0.75.

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

[8] 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.

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

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 a high investment if $EU(X) \geq EU(Y)$. Given the choice of parameters in our design, leaders face a trade-off between maximizing the expected net payoff to themselves and that to 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 X if $\beta \geq \beta^*$. That is, leaders will choose a high investment 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 the high investment. 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 good and a bad 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 evaluation of the leader's good and bad outcomes depend on the leader's gender. Gender stereotypes can bias beliefs through two channels. First, gender differences may emerge in evaluators' prior beliefs. Second, gender differences may emerge in posterior beliefs if male and female leaders' outcomes are attributed differently.[9]

Given the nature of our decision-making environment, it is gender stereotypes about prosocial preferences which may influence evaluators' beliefs. In their survey papers, Croson and Gneezy (2009), Niederle (2016), and Bilén, Dreber, and Johannesson (2021) report that, depending on the context, women tend to be more prosocial than men. If female leaders are perceived to be more prosocial than male leaders in our environment, then evaluators may hold

[9] In the context of self-evaluation, Bordalo et al. (2019) and Coffman, Collis, and Kulkarni (2020) show that stereotypes influence both belief formation and belief updating.

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 that the good outcomes of female leaders are attributed more to their investment decision as compared to those of male leaders, while their bad 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.[10] On the other hand, following discriminatory social norms, women may treat other women more harshly.[11] 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:

$$\operatorname{logit}\left(\sigma_j(X|Q)\right) = \operatorname{logit}(\mu_j) + I(Q = Q_H) \cdot \operatorname{logit}(p) + I(Q = Q_L) \cdot \operatorname{logit}(1-p), \qquad (2)$$

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

$$\operatorname{logit}\left(\hat{\sigma}_j(X|Q)\right) = \boldsymbol{\delta}\operatorname{logit}(\hat{\mu}_j) + \boldsymbol{\gamma_G}\, I(Q = Q_H) \cdot \operatorname{logit}(p) + \boldsymbol{\gamma_B}\, I(Q = Q_L) \cdot \operatorname{logit}(1-p) + \varepsilon_j, (3)$$

where ε_i captures non-systematic errors.

[10] Recent evidence by Cappelen, Falch, and Tungodden (2019) also shows that female spectators are more likely to consider males to have a productivity advantage and therefore evaluate their failure less favorably.
[11] 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, Van Laar, and Ellemers, 2016; Arvate, Galilea, and Todescat, 2018).

The specification in equation (3) nests the theoretical Bayesian benchmark as a special case with $\delta = \gamma_G = \gamma_B = 1$.[12] The parameters δ, γ_G, and γ_B therefore allow us to capture the emphasis that evaluators place on prior beliefs, good outcomes, and bad outcomes, respectively, when updating their beliefs. Focusing specifically on gender differences in γ_G and γ_B, our goal is to examine whether there are differences in the attribution of outcomes between male and female leaders. Specifically, in the context of our experiment, if γ_G (γ_B) is higher, then this implies that the evaluator is more likely to attribute the leader's good (bad) outcome to the leader's decision. If γ_G (γ_B) is lower, then the leader's good (bad) outcome is attributed more to luck.

4 Results

Our main research question is to understand how the good and bad outcomes of leaders are evaluated. To answer this question, we study whether evaluators attribute male and female leaders' outcomes to their investment decisions or luck.

4.1 Evaluators' Beliefs

In general, the evaluators in our sample behave in a pattern as predicted by Bayes' rule. Posterior beliefs for good outcomes are on average higher than those for bad outcomes (mean posterior beliefs across rounds are 51.2% and 31.9% for good outcomes and bad outcomes, respectively).[13] Table 2 presents ordinary least squares (OLS) estimates of equation (3) in Section 3 both at the pooled level (column 1) and separately by the leader's gender (columns 2 and 3). Column (1) reveals that evaluators suffer from base-rate neglect (test of $\delta = 1$: p-value < 0.001), attribute good outcomes more to luck than a Bayesian would on average (test of $\gamma_G = 1$: p-value = 0.064), but they are no different from a Bayesian in their attribution of bad outcomes (test of $\gamma_B = 1$: p-value = 0.495). Consequently, there is an overall asymmetry in the attribution of outcomes, where the leader's bad outcomes are attributed more to their

[12] Hence, we can also use this framework to examine systematic deviations from the Bayesian benchmark. See, e.g., Grether (1980), Ambuehl and Li (2018), Buser, Gerhards, and van der Weele (2018), Coutts (2019), Barron (2021), Erkal, Gangadharan, and Koh (2022a), and Möbius et al. (forthcoming).

[13] 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 the high investment. As shown in Figure 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, or if all their posterior beliefs are equal to their prior beliefs. 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, Gangadharan, and Koh, 2022a; Möbius et al., forthcoming). In Appendix C, we present as a robustness check the analysis with the full sample. Our main conclusions remain largely unchanged.

investment decisions and good outcomes are attribute more to luck on average (test of $\gamma_G = \gamma_B$: p-value = 0.083). This result of an overall asymmetry in the attribution of good and bad outcomes is consistent with Erkal, Gangadharan, and Koh (2022a).

We next focus on 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 good outcomes (p-value = 0.582). The direction of the estimates of γ_G in both columns suggests that the good 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 $\gamma_G = 1$ 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 evaluators differ in their evaluation of bad outcomes between male and female leaders (p-value = 0.018). Specifically, evaluators attribute the bad outcomes of male leaders more to their investment decisions than a Bayesian (test of $\gamma_B = 1$ in column 2: p-value = 0.059), but they are no different from a Bayesian in their attribution of the female leaders' bad outcomes (test of $\gamma_B = 1$ in column 3: p-value = 0.156). Hence, even though we do not observe an overall bias in the attribution of bad outcomes (column 1), the result at the pooled level masks a difference in the way bad outcomes are attributed between male and female leaders. We summarize as follows.

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

Our results suggest that, in the evaluation of bad outcomes, men are evaluated more unfavorably than women. The gender difference we observe in the updating behavior with respect to bad outcomes is consistent with expectation that women are more prosocial than men. That is, because women are expected to be more prosocial, their bad outcomes are attributed more to luck. However, we do not observe any gender difference in the evaluation of good outcomes.

We next explore whether similar gender differences exist in evaluators' prior beliefs. Specifically, we ask whether it is the case that evaluators expect women to be more prosocial. Panel (a) of Figure 1 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 = 0.250). Panel (b) shows the proportion of high investment decisions made by participants (in the role of leaders) in Stage 1. The figure reveals that 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).[14]

Overall, we do not find differences between male and female leaders both in evaluators' prior beliefs about leaders' decisions and in leaders' actual investment decisions. Beliefs are well calibrated in the sense that they are consistent with a lack of difference in male and female leaders' investment decisions. We summarize as follows.

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

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 equation (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 good outcomes between male and female leaders (p-values = 0.971 and 0.503, respectively).[15] On the other hand, while female evaluators are no different in their attribution of bad outcomes between male and female leaders (p-value = 0.863), male evaluators are more likely to attribute the bad outcomes of male leaders more to their investment decisions as compared to those of female leaders (p-value = 0.003). We also observe that female evaluators are consistently no different from the Bayesian benchmark in their attribution of bad outcomes regardless of the leader's gender (test of $\gamma_B = 1$ in columns 1 and 2: p-values = 0.803 and 0.664, respectively). On the other hand, relative to a Bayesian, male evaluators tend to attribute the bad outcomes of male leaders more to their investment decisions (test of $\gamma_B = 1$ in column 3: p-value = 0.021) and those of female leaders more to luck (test of $\gamma_B = 1$ in column 4: p-value = 0.060).

Hence, the gender difference in the attribution of bad outcomes we observe in Result 1 appears to be driven by male evaluators. We summarize our result as follows.

[14] Our conclusions from the non-parametric tests are also supported by regression analyses, as reported in Table B1 and Table B2 of Appendix B.

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

Result 3. *Male evaluators attribute the bad 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' bad outcomes. Both female and male evaluators are no different in their attribution of good outcomes between male and female leaders.*

We also 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 equation (3) separately for evaluators who chose low investment (columns 1 and 2) and evaluators who chose high investment (columns 3 and 4). We also analyze evaluators' updating behavior based on their prosocial preferences as defined according to their decisions in the dictator game, which was conducted as a separate task to the leadership task. The estimates, presented in Table B3 of Appendix B, are consistent with our main conclusions in Table 4.

The estimates in Table 4 reveal that both evaluators who chose low investment and those who chose high investment are no different in their attribution of good outcomes between male and female leaders (p-values = 0.481 and 0.318, respectively). Moreover, evaluators who are less prosocial (i.e., those who chose low investment) attribute leaders' good outcomes to luck relative to a Bayesian (p-values = 0.012 and 0.003 in columns 1 and 2, respectively), while evaluators who are more prosocial (i.e., those who chose high investment) are no different from a Bayesian in their attribution of good outcomes (p-values = 0.313 and 0.871 in columns 3 and 4, respectively).[16]

The gender biases in the attribution of bad outcomes are driven by evaluators who are more prosocial. Specifically, columns (3) and (4) reveal that prosocial evaluators are more likely to attribute the bad outcomes of female leaders to luck than those of male leaders (p-value = 0.002). Relative to a Bayesian, prosocial evaluators attribute bad outcomes of female leaders more to luck and those of male leaders more to their actions (p-values = 0.019 and 0.043, respectively). On the other hand, columns (1) and (2) reveal that evaluators who are less prosocial are no different in their attribution of bad outcomes between male and female leaders

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

(p-value = 0.481). These evaluators are also no different from a Bayesian in their attribution of leaders' bad outcomes (p-values = 0.218 and 0.887 in columns 1 and 2, respectively).

Hence, the gender difference in the attribution of bad outcomes we observe in Result 1 also appears to be driven by prosocial evaluators. We summarize our result as follows.

Result 4. *Prosocial evaluators attribute the bad outcomes of male leaders more to their investment decisions and those of female leaders more to luck, while less prosocial evaluators are no different in their attribution of male and female leaders' bad outcomes. Regardless of their own prosocial preferences, evaluators are no different in their attribution of good outcomes between male and female leaders.*

A possible explanation of our findings is benevolent sexism. Unlike hostile sexism, benevolent sexism tends to lead to behaviors toward women that are often characterized as prosocial.[17] This implies in our context that evaluators may see the need to protect female leaders and treat them more favorably, thus distorting their evaluation of bad outcomes in favor of women (Glick and Fiske, 1996).

5 Conclusion

In environments where the outcomes of leaders are determined by a combination of luck and unobservable actions, are successes and failures 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 women are held to a higher standard because they are expected to behave in a more prosocial manner.

We find an asymmetry between the evaluation of good outcomes and bad outcomes. Good outcomes are not treated differently between male and female leaders, suggesting that men and women are given the same credit for their successes (which is less than what a Bayesian would do). However, while the bad outcomes of male leaders are attributed more to

[17] According to Glick and Fiske (1996), hostile sexism is behavior that is in line with the classic definition of prejudice by Allport (1954). Such behavior is also consistent with the notion of animus or taste-based discrimination in economics (Becker, 1957). 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.

their decisions relative to a Bayesian, those of female leaders are attributed more to luck. Hence, in case of failure, men are assigned more blame than women.

We find that this gender bias in the evaluation of bad outcomes is driven by male evaluators and evaluators who are prosocial. Hence, one possible explanation for our findings is benevolent sexism. Male evaluators may see the need to treat women more favorably, thus giving them a greater benefit of the doubt in the face of failure. While our results suggest that the gender bias is acting against men and in favor of women in this instance, such biases in the evaluation process may still lead to adverse outcomes for women. These biases may potentially lead to distortions in the incentives provided to decision makers in positions of power and may harm the future actions taken by them. They may also imply that women are provided with less or different feedback than men, which may be crucial for the development of their careers. Future research can shed light on the distortionary impact that such biased evaluation processes can have on decision making by leaders.

References

Allport, G.W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Ambuehl, S., Li, S. (2018). Belief updating and the demand for information. *Games and Economic Behavior,* 109:21-39.

Arvate, P.R., Galilea, G.W., Todescat, I. (2018). The queen bee: A myth? The effect of top-level female leadership on subordinate females. *The Leadership Quarterly,* 29(5):533-548.

Barron, K. (2021). Belief updating: Does the 'good-news, bad-news' asymmetry extend to purely financial domains? *Experimental Economics,* 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*.

Becker, G.S. (1957). *The economics of discrimination*. Chicago: University of Chicago Press.

Bénabou, R., Tirole, J. (2010). Individual and corporate social responsibility. *Economica,* 77(305):1-19.

Bilén, D., Dreber, A., Johannesson, M. (2021). Are women more generous than men? A meta-analysis. *Journal of the Economic Science Association,* 7:1-18.

Bohren, J.A., Imas, A., Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review,* 109(10):3395-3436.

Bordalo, P., Coffman, K., Gennaioli, N., Shleifer, A. (2019). Beliefs about gender. *American Economic Review,* 109(3):739-773.

Buser, T., Gerhards, L., van der Weele, J. (2018). Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainy,* 56(2):165-192.

Campos-Mercade, P., Mengel, F. (2021). Non-Bayesian statistical discrimination. *Working Paper*.

Cappelen, A.W., Falch, R., Tungodden, B. (2019). The boy crisis: Experimental evidence on the acceptance of males falling behind. *Working Paper*.

Coffman, K. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics,* 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. *Management Science*.

Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics,* 22(2):369-395.

Croson, R., Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature,* 47(2):448-474.

Derks, B., Van Laar, C., Ellemers, N. (2016). The queen bee phenomenon: Why women leaders distance themselves from junior women. *The Leadership Quarterly,* 27(3):456-469.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics,* 14(4):583-610.

Erkal, N., Gangadharan, L., Koh, B.H. (2020). Replication: Belief elicitation with quadratic and binarized scoring rules. *Journal of Economic Psychology,* 81:102315.

Erkal, N., Gangadharan, L., Koh, B.H. (2022a). By chance or by choice? Biased attribution of others' outcomes when social preferences matter. *Experimental Economics,* 25(2):413-443.

Erkal, N., Gangadharan, L., Koh, B.H. (2022b). 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. *Experimental Economics,* 10(2):171-178.

Glick, P., Fiske, S.T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology,* 70(3):491-512.

Glick, P., Fiske, S.T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist,* 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. *Journal of Personality and Social Psychology,* 79(5):763-775.

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

Grether, D.M. (1980). Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly Journal of Economics*, 95(3):537-557.

Heilman, M.E., Chen, J.J. (2005). Same behavior, different consequences: reactions to men's and women's altruistic citizenship behavior. *Journal of Applied Psychology*, 90(3):431-441.

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

Miller, D.T., Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin,* 82(2):213-225.

Möbius, M.M., Niederle, M., Niehaus, P., Rosenblat, T.S. (forthcoming). Managing self-confidence. *Management Science*.

Niederle, M. (2016). Gender. In J.H. Kagel & A.E. Roth (Eds.), *The Handbook of Experimental Economics* (2:481-562). New Jersey: Princeton University Press.

Solnick, S.J. (2001). Gender differences in the ultimatum game. *Economic Inquiry*, 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. *Personality and Social Psychology Bulletin*, 22(5):507-519.

List of Tables

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 × 150 + 0.25 × 0) (expected return from Investment X) = 212.5 ECU, while the expected net payoff for each evaluator is given by (0.75 × 150 + 0.25 × 0) (expected return from Investment X) = 112.5 ECU.

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
δ : Logit(prior belief)	0.624	0.598	0.658	0.472
	(0.041)	(0.053)	(0.063)	
γ_G : Good outcome × logit(p)	0.783	0.854	0.725	0.582
	(0.117)	(0.210)	(0.104)	
γ_B : Bad outcome × logit($1-p$)	1.064	1.287	0.843	0.018
	(0.094)	(0.150)	(0.110)	
Observations	1,600	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 $I(\text{Good Outcome}) + I(\text{Bad Outcome}) = 1$, there is no constant term in the regression.

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	Male Leader (3)	Female Leader (4)	(3) vs. (4) p-value
δ : Logit(prior belief)	0.681	0.621	0.652	0.533	0.693	0.097
	(0.086)	(0.101)		(0.065)	(0.070)	
γ_G : Good outcome × logit(p)	0.812	0.820	0.971	0.915	0.630	0.503
	(0.168)	(0.166)		(0.404)	(0.138)	
γ_B : Bad outcome × logit($1 - p$)	0.962	0.923	0.863	1.631	0.753	0.003
	(0.150)	(0.177)		(0.262)	(0.128)	
Observations	410	420		370	400	
R-squared	0.619	0.560		0.479	0.551	

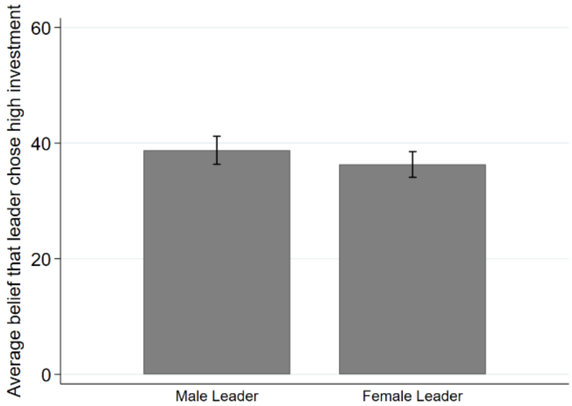
Robust standard errors clustered at the participant level in parentheses. This analysis excludes participants classified as inconsistent or non-updaters. Since the regression specification estimates parameters of an augmented Bayes' rule, no controls can be included as the presence of any controls would invalidate the interpretation of the parameters. Moreover, since $I(\text{Good Outcome}) + I(\text{Bad Outcome}) = 1$, there is no constant term in the regression.

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

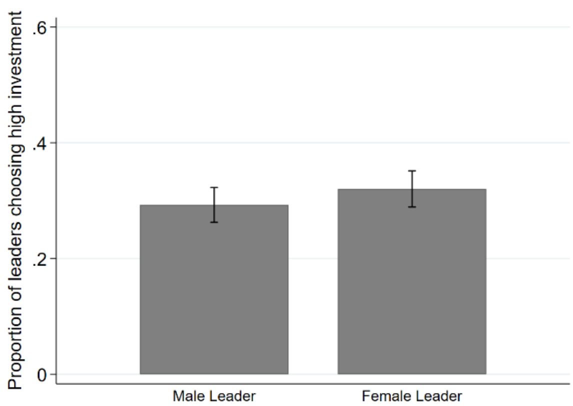
	Dependent variable: Logit(posterior belief)					
	Evaluator Chose Low Investment			Evaluator Chose High Investment		
Variables	Male Leader (1)	Female Leader (2)	(1) vs. (2) p-value	Male Leader (3)	Female Leader (4)	(3) vs. (4) p-value
δ : Logit(prior belief)	0.562	0.669	0.272	0.609	0.457	0.534
	(0.076)	(0.061)		(0.072)	(0.232)	
γ_G : Good outcome × logit(p)	0.471	0.638	0.481	1.434	0.970	0.318
	(0.204)	(0.119)		(0.427)	(0.182)	
γ_B : Bad outcome × logit($1-p$)	1.290	0.980	0.256	1.407	0.496	0.002
	(0.233)	(0.142)		(0.196)	(0.209)	
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 $I(\text{Good Outcome}) + I(\text{Bad Outcome}) = 1$, there is no constant term in the regression.

List of Figures



(a) Evaluator's average prior belief that leader has chosen high investment



(b) Proportion of high investment decisions by leaders

Figure 1: Evaluators' prior belief that the leader has chosen high investment and leaders' investment decisions, by the leader's gender