

Choosing and Using Information in Evaluation Decisions

Katherine B. Coffman, Scott Kostyshak, Perihan O. Saygin

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Choosing and Using Information in Evaluation Decisions

Abstract

Most studies of gender discrimination consider how male versus female candidates are assessed given otherwise identical information about them. But, in many settings of interest, evaluators have a choice about how much information to acquire about a candidate before making a final assessment. We use a large controlled experiment to explore how this type of endogenous information acquisition amplifies discriminatory outcomes in a simulated hiring environment. Across evaluators, we vary the composition of candidate pools, exploring not only environments where men outperform women on average but also environments with no gender difference or with a female advantage. Perhaps surprisingly, we observe no gender discrimination overall: conditional on their likelihood of being qualified, male and female candidates receive indistinguishable evaluations. But, we observe important differences across candidate pools. Candidates belonging to an advantaged group—the gender with the performance advantage in the pool—receive significantly better evaluations than equally qualified candidates in pools with no gender gap in performance. Similarly, candidates belonging to a disadvantaged group—the gender with a performance disadvantage in the pool—receive significantly worse evaluations relative to equally qualified candidates in pools with no gender gap in performance. This “relative advantage” bias appears in initial assessments, influences how evaluators update their beliefs about a candidate after acquiring more information, and persists in final evaluations. This bias has a significantly larger impact on evaluations when evaluators endogenously acquire information compared to treatments where we exogenously provide it, in part because we observe significant under-acquisition of information. We show that this bias leads to two important types of mistakes: evaluators miss out on talented candidates from disadvantaged groups and over-select less talented candidates from advantaged groups.

JEL-Codes: J160, J710.

Keywords: discrimination, information acquisition, beliefs, belief updating, stereotypes.

Katherine B. Coffman
Harvard Business School, Boston / MA / USA
kcoffman@hbs.edu

Scott Kostyshak
Department of Economics, University of
Florida, Gainesville / FL / USA
skostyshak@ufl.edu

Perihan O. Saygin
Department of Economics, University of
Florida, Gainesville / FL / USA
psaygin@ufl.edu

March 20, 2024

Pre-registered: AEARCTR-0010969. We received funding from Harvard Business School. We would like to thank May Hong, Yoona Kim, Emma Ronzetti, and Douglas Turner for excellent research assistance, and Lucas Coffman, Markus Eytting, Nagore Iriberry, James Peck, Matthew Rabin, David Sappington, Joshua Schwartzstein, Andrei Shleifer, and Sevgi Yuksel for their helpful feedback and suggestions. We benefited from comments and questions from seminar audiences at Wharton, Ohio State, Cornell, NYU Abu Dhabi, Purdue, Florida State, and Bonn as well as conference participants at the Motivation and Incentives Conference at Harvard Business School, the 2023 Spring Behavioral and Experimental Economics Research Conference, the 2023 BRIQ Beliefs Workshop, the 2023 ESA Europe Meeting, and the 2023 SITE Gender Conference.

1 Introduction

Evaluations play an important role in determining economic outcomes. In academic contexts, teachers provide grades to students, admissions officers select new cohorts of university students, advisors identify promising graduate students, and editors and referees decide which papers will be published. In professional contexts, hiring managers choose candidates and supervisors review performance. Even outside of school and work, life trajectories are shaped by the discretionary judgements of evaluators such as mortgage lenders, criminal justice officers, and doctors.

Across these evaluation contexts, social scientists have explored the prevalence, sources, and consequences of discrimination. Perhaps the most common approach to documenting discrimination is to compare the treatment of two otherwise identical candidates, isolating a role for gender (or race, age, etc.) in driving disparate outcomes. For instance, audit studies such as the seminal work by Bertrand and Mullainathan (2004) send identical resumes to different job openings, varying only the name attached to the resume. They find that resumes receive significantly fewer callbacks when they are labeled with a Black-sounding name rather than a white-sounding name. Studies of this type provide clean, compelling evidence of discrimination: *given* identical information about two candidates, those candidates are treated differently based upon their race or gender.

Our project builds on this important body of work by exploring what happens when information is not *given* but instead must be *acquired*. We consider a decision-maker who must choose how much to learn about a candidate before making their evaluation. For instance, a hiring manager must choose how many references to call before deciding whether to interview the candidate, a loan officer must decide how many years of employment history to review carefully before deciding whether to extend a mortgage, or a doctor must decide how many tests to order before making a diagnosis. Our hypothesis is that these endogenous decisions about how much information to acquire have the potential to fuel discrimination.

We study this question within the context of a simulated hiring experiment. Evaluators are presented with limited information about a candidate. They make an initial assessment of that candidate's likely quality. Then, they decide whether to acquire additional information about the candidate or instead make a final evaluation. Acquiring additional information is costly, but

only modestly so. We use this environment to study (i) evaluators' initial beliefs about candidate quality, (ii) when evaluators choose to acquire more information, (iii) how evaluators update their beliefs in response to that information, and (iv) evaluators' final assessments of candidate quality. We ask how each of these decisions depends upon candidate gender and the characteristics of the pool of candidates. In particular, we ask how the extent of gender discrimination depends upon the underlying gender gaps in performance in the candidate pool.

In a preliminary study, we collect performance data from a group of workers via the online platform Prolific. These workers serve as the candidates in our main evaluation study. For each candidate, we measure performance on a brief multiple-choice math and science test. Each candidate also answers some brief questions about themselves, including their gender. Based upon their performance, we classify each candidate as either Above-the-Bar or Below-the-Bar.

In our main study, we recruit new participants from Prolific to serve as evaluators. The task of evaluators is to accurately classify candidates as Above or Below-the-Bar. For each candidate, an evaluator sees a brief profile that includes the gender of the candidate. In our endogenous treatment, evaluators choose whether to receive additional information about the candidate prior to making their classification. This additional information takes the form of noisy binary signals about the candidate's performance. Signals are imperfect but informative, with Above-the-Bar candidates generating a positive signal more often than Below-the-Bar candidates. Evaluators can choose to acquire up to 5 signals about a given candidate before making their classification. In our exogenous treatment, evaluators must view all 5 signals, sequentially, before making their classification. Comparisons across these two treatments isolate the role of endogenizing information acquisition. This design also allows us to separately consider both evaluator selection into receiving more information and, at the same time, biased responses to information absent these selection effects.

We incentivize evaluators to make accurate judgements by offering a bonus for a correct final classification of a candidate. We take steps to encourage information acquisition by assigning only a small cost to each signal (5 cents) and by minimizing the time cost of viewing additional signals. We also ask for the evaluator's initial belief of the candidate being Above-the-Bar, based purely on the candidate profile, and then elicit an updated belief after each additional signal acquired.

This design yields a wealth of interesting data to explore. From an evaluation standpoint, we can analyze initial assessments, how beliefs are updated, and the accuracy of final classifications. From an information standpoint, we can unpack the endogeneity of information acquisition, asking what predicts the decision to seek more information about a candidate and how this impacts final classifications. Our goal is to understand how biases emerge from the choice and use of information.

A critical piece of our design is the construction of four different candidate pools for evaluators to consider. All pools have 50 candidates, 25 male and 25 female. But, the pools vary in the share of male and female candidates within the pool that are Above-the-Bar: 60% of men and 40% of women; 60% of both men and women; 40% of both men and women; and, finally, 40% of men and 60% of women. We will refer to these pools as 60 – 40, 60 – 60, 40 – 40, and 40 – 60, respectively.

We provide evaluators with the relevant shares for their pool, and this information remains available to them throughout their participation. We choose to provide these shares instead of relying on the (potentially biased) prior beliefs of evaluators so that we can induce exogenous variation in beliefs about gender differences. This is critical for understanding what role beliefs about population-level gender differences play in driving discriminatory behavior. By exploring how gender discrimination varies across pools, we can back out the extent to which relative comparisons between men and women are biasing evaluator decision-making.

Discrimination on the basis of relative comparisons is deeply tied to a recent literature on the importance of contrast across groups in shaping beliefs. The Bordalo, Coffman, et al. (2016) stereotypes model predicts that beliefs about a group are biased toward types that are diagnostic or representative relative to a reference group—those types that are *relatively* more frequent in the target group compared to the reference group.¹ This framework makes a clear prediction for our setting. Consider the case in which 60% of men are Above-the-Bar. Stereotyping predicts that beliefs about men will be more favorable when they are in a pool where only 40% of women are Above-the-Bar than when they are in a pool where 60% of women are also Above-the-Bar. Evaluator beliefs for a given group would thus depend not only on the true proportion of that group that is Above-the-Bar, but also on the relative advantage that group has compared to the

¹Evidence consistent with this model of stereotyping has been found in a variety of contexts. See Alesina, Miano, and Stantcheva (2023), Bordalo, Coffman, et al. (2019), Carlana (2019), Hebert (2020), and Coffman (2014), among others.

other. We test this prediction in our setting by considering how the assessment of a candidate of a given gender depends upon their group’s relative advantage (or disadvantage) within the candidate pool.

In this same spirit, Esponda, Oprea, and Yuksel (2023) document that contrast between groups influences how new information is perceived. Relying on an abstract setting with green and orange groups, they show that evaluator perceptions of a noisy visual signal depend on differences in priors across groups: the same visual signal is perceived as more favorable when it is associated with the advantaged group. Our project advances this important line of work by considering how stereotyping shapes not only how evaluators use information, but also how they choose how much of it to acquire. Relative to Esponda et al. (2023), we study a setting where there is arguably less wiggle room in how a signal could be perceived. And, by using a less abstract setting, we explore the explanatory power of this important channel for gender discrimination.

Collapsing across the four candidate pools, we find no evidence of discrimination against female or male candidates in either treatment. Evaluators assess female candidates no less favorably than male candidates, both in terms of initial assessments and final classifications.

However, we observe significant and substantial evidence of contrast effects: evaluator assessments are biased by comparisons across groups. Holding fixed a candidate’s prior likelihood of being Above-the-Bar, a candidate is treated significantly more favorably when their gender has a relative advantage. For example, men are significantly more likely to be classified as Above-the-Bar in the 60 – 40 pool, where a greater share of men are top performers, than in the 60 – 60 pool, where both men and women have the same chance of being a top performer. Importantly, female candidates also benefit from relative advantage when it is their group that is advantaged: conditional on having a 60% chance of being Above-the-Bar, they are evaluated significantly more favorably when male candidates have only a 40% chance of being Above-the-Bar relative to when both groups have a 60% chance of being Above-the-Bar. Just as candidates with a 60% chance of being Above-the-Bar are viewed more positively when their group has the relative advantage, candidates with a 40% chance of being Above-the-Bar are viewed more negatively when their group has the relative disadvantage: evaluators provide more negative assessments to low-prior candidates from the unequal pools than to those same candidates in the 40 – 40 pools.

Relative advantage has a sizable impact on initial assessments, prior to the receipt of any signals: the gap in initial assessments between high and low-prior candidates is 75% larger when evaluators face candidate pools with unequal groups.² Relative advantage continues to matter in final assessments, even after evaluators have had the chance to acquire more information. The gap in final classifications between high and low-prior candidates is more than 50% larger when evaluators face candidate pools with unequal groups.

Final evaluations are biased, in part, because of under-acquisition of information. Despite the low cost of information, in nearly 25% of cases, evaluators choose to receive no signals about the candidate and the median number of signals acquired is 2.³ There is significant selection into information. The initial assessments of evaluators who choose to receive no signals about the candidate are significantly more biased by relative advantage on average.

In the endogenous treatment, high-prior candidates from the advantaged group are significantly more likely to be classified as Above-the-Bar (67%) compared to high-prior candidates in the 60 – 60 pool (61%). And, evaluators are quicker to classify them as such, requiring significantly fewer signals on average before deciding to make their final evaluation. Similarly, low-prior candidates from the disadvantaged group are significantly less likely to be classified as Above-the-Bar (40%) compared to low-prior candidates in the 40 – 40 pool (46%), and evaluators require significantly more signals on average before making this positive determination when the candidate is disadvantaged. Note that these findings rely on comparisons of candidates *with the same prior likelihood of being Above-the-Bar*, who produce the same signals in expectation. These otherwise identical candidates are treated differently depending upon whether their group happens to be advantaged or disadvantaged relative to the other.

Relative advantage shapes not only how much information evaluators choose to acquire, but also how they use the information they receive. By shifting focus to the exogenous treatment, we can study biases in belief updating absent selection into information. Consider a high-prior candidate. When an evaluator views a negative signal about that candidate, they remain significantly more

²We compute this by contrasting the comparisons of the average assessment of low and high-prior candidates from the 40 – 40 and 60 – 60 pools to the assessments of low and high-prior candidates from the 60 – 40 and 40 – 60 pools.

³By contrast, a Bayesian decision-maker would always acquire at least 3 signals per candidate. We show this formally in Section 5.1

optimistic when the candidate is from a relatively advantaged group (60 – 40 relative to 60 – 60). Casually speaking, relative advantage seems to provide high-prior candidates more benefit of the doubt. On the flip side, when an evaluator views a positive signal for a low-prior candidate, they update their beliefs significantly less favorably when that candidate is relatively disadvantaged (40 – 60 relative to 40 – 40). Informally, low-prior candidates seem to have a harder time convincing evaluators of their high quality when they are at a relative disadvantage.

While overall accuracy of final classifications is high, endogenizing information acquisition significantly increases the rate of mistakes, from 16% in the exogenous treatment to 25% in the endogenous treatment. Not surprisingly, mistakes are more common for atypical candidates: Above-the-Bar candidates from low-prior groups and Below-the-Bar candidates from high-prior groups are significantly less likely to be classified correctly. These mistakes are more common (i) in the endogenous treatment, and (ii) when groups are unequal. That is, a talented candidate from a low-prior group is least likely to be identified as such when their group is disadvantaged relative to the other, and when their evaluator gets to choose whether to seek out more information. Similarly, evaluators are less likely to recognize that a candidate from a high-prior group is actually Below-the-Bar when that candidate is from a relatively advantaged group and getting additional information is not required.

Together, these results illustrate the reinforcing role of relative advantage. Contrast between groups makes it more likely that high-prior candidates are classified as Above-the-Bar and that low-prior candidates are classified as Below-the-Bar. Allowing for endogenous information acquisition exacerbates the impact of this form of stereotyping, with group advantage reducing the number of signals evaluators acquire before reaching a stereotypical judgment. While these judgements are correct on average, stereotyping overlooks the significant heterogeneity within groups, at the expense of talented candidates from disadvantaged groups (and to the benefit of less talented candidates from advantaged groups).

2 Related Literature

Our work contributes to a growing literature on belief-based discrimination, under which potentially biased beliefs drive discriminatory behavior (Bordalo, Coffman, et al., 2016; Bohren, Imas, and Rosenberg, 2019; Bohren, Haggag, et al., 2019; Bordalo, Coffman, et al., 2019; Coffman, Exley, and Niederle, 2021). Much of this work considers fixed information sets: the researcher asks, holding fixed a given set of information about a candidate, are male and female candidates evaluated differently?⁴ We build on this important body of work by expanding the decision problem to consider endogenous acquisition of information, asking how evaluators choose whether to get more information and how evaluators use that information to update their beliefs.

Our study of endogenous information acquisition is related to the idea of attention discrimination introduced by Bartoš et al. (2016). They consider how much effort evaluators choose to exert when reading an application, finding that evaluators spend more effort scrutinizing minorities in a lemon-dropping market, and less effort considering minorities in a cherry-picking market. Building on this idea, Cavounidis, Lang, and Weinstein (2023) explore the idea of selective monitoring, where an employer is more likely to monitor and subsequently punish Black employees relative to white. A related literature explores how lack of information may perpetuate discrimination; for instance, past discrimination against a stigmatized group may lead to under-sampling, depriving decision-makers of the information needed to de-bias their beliefs.⁵

Even closer to our work, Eytting (2023) uses a series of online experiments to study how motivated reasoning shapes information acquisition and discriminatory behavior. He finds that individuals are more likely to seek out information that matches their motivation, using this to justify discriminatory behavior. Relative to this past work, we introduce an important comparison across endogenous and exogenous information treatments. Our design and analysis are informed by cognitive, rather than motivated, explanations for behavior; indeed, our results suggest that cognitive biases alone are enough to generate discrimination through biased acquisition and biased processing of information.

⁴In their recent working paper, Bohren, Hull, and Imas (2022) refer to this as “direct” discrimination, distinguishing it from systemic forms of discrimination that may give rise to differences in candidate qualifications prior to the evaluation stage.

⁵See the theoretical approaches of Komiyama and Noda (2020) and Lepage (2021) as well as experimental evidence from Gupta (2023) and Fonseca and McCrea (2023).

Our work is also connected to a wealth of recent literature on biases in belief updating, particularly as they relate to gender and discrimination. One strand of this work documents that stereotypes shape how men and women update their beliefs about themselves in response to noisy signals (Möbius et al., 2022, Coffman, Collis, and Kulkarni, 2023, Coffman, Ugalde Araya, and Zafar, 2024).⁶ The second relevant strand of this work looks at biased processing of information about others.⁷ In a stylized experiment on racial bias and hiring, Rackstraw (2022) shows that decision-makers give more weight to stereotype-consistent signals when updating their beliefs. Using evidence from both abstract and gender-based experiments, Campos-Mercade et al. (2023) show that employer conservatism (under-reaction to information) leads to more discrimination against members of disadvantaged groups relative to a Bayesian benchmark. Moving to the field, Egan, Matvos, and Seru (2022) show asymmetry in how misconduct is punished among financial advisors, with women facing greater punishment than men, inconsistent with a Bayesian response. Similarly, Sarsons (2017) finds that referrals to female surgeons, relative to male surgeons, fall more sharply after a negative patient outcome.

We build on this excellent recent work by studying endogenous information acquisition and biased information processing in a unified framework, with treatments that allow us to tease apart potential mechanisms. Our findings are both remarkably consistent with and shed more light on many of these past findings. We find support for both biased acquisition and processing of information, driven not by bias against women or motivated reasoning, but rather by a contrast-based form of stereotyping that exaggerates differences across groups.

3 Experimental Design

We study evaluator decision-making in a controlled environment that allows for causal identification of key factors that predict behavior. To do so, we construct a simulated labor market using candidates and evaluators recruited from Prolific. Below, we describe the design. The full

⁶Our design allows us to study a similar concept in the context of evaluating others, identifying the impact of gender stereotyping on belief updating.

⁷A flood of recent work has begun to unpack the various mechanisms that may contribute to these belief biases. For instance, Ruzzier and Woo (2023) explore confirmation bias, a la Rabin and Schrag (1999), Campos-Mercade and Mengel (2023) consider conservatism, Schwartzstein and Sunderam (2021) consider endogenously chosen models for organizing data, and Bordalo, Conlon, et al. (2023b) and Bordalo, Conlon, et al. (2023a) consider selective attention and memory. Our results will be most consistent with stereotyping as a driver (Bordalo, Coffman, et al. (2016)).

experimental materials are available in the Appendix; the experiment received IRB approval from Harvard Business School and was pre-registered (AEARCTR-0010969).

3.1 Candidate Sourcing

We field candidates by running a preliminary study with 200 participants.⁸ Our goal was to collect performance data and individual characteristics that could be used to build candidate profiles. Participants begin the study by taking a 10-question math and science skills test. Each multiple-choice question appears on a separate page of the study, in a random order. They have up to 20 seconds to answer each question; the time limit is designed to discourage looking up answers on the internet. We incentivize participants by offering 20 cents in additional compensation for each question answered correctly.

Following completion of the test, participants answer 25 simple questions about their preferences. Each question has only two possible answers. For example, participants are asked whether they prefer beach vacations or mountain vacations, whether they prefer spring or fall, etc. The goal is to have information about each potential candidate that can serve as profile “filler,” additional information about the person that increases evaluator engagement and differentiates individuals beyond gender. We intended for these questions to be unrelated (both in perception and in reality) to performance on the test and to gender, simplifying our analysis of evaluator decisions.

After collecting the candidate data, we used the distribution of performance on the test to classify each participant as having demonstrated Above-the-Bar performance (top-half of performers) or Below-the-Bar performance (bottom-half of performers). Using a binary measure of candidate performance simplifies the design and empirical analysis. We use the Above-the-Bar terminology because we felt it would be intuitive in a hiring context and that it would be perceived (accurately) as binary.

⁸The study was advertised as 10 minutes, with \$2 in guaranteed pay and the possibility of incentive pay. It was conducted on Prolific in June 2022.

Table 1: Candidate Pools

Probability that Candidate from Group is Above-the-Bar			
		Female Candidate	
		40%	60%
Male	40%	40-40	40-60
Candidate	60%	60-40	60-60

Note: This table presents the construction of four different pools with the corresponding share of candidates who are Above-the-Bar for female and male candidates.

3.2 Constructing Candidate Pools

We use the data from the candidate survey to construct pools of candidates for evaluators to assess. We build four different pools. Each pool has exactly 50 members: 25 women and 25 men. Across pools, we vary the frequency of men and women within the pool who are Above-the-Bar. Table 1 displays the design. In the 60 – 60 pool, both 60% of men and 60% of women within the pool are Above-the-Bar. In the 60 – 40 pool, 60% of men but only 40% of women are Above-the-Bar, while in the 40 – 60 pool, 40% of men and 60% of women are Above-the-Bar. Finally, in the 40 – 40 pool, only 40% of both men and women are Above-the-Bar.

We build a candidate profile for each candidate within the pool. This profile consists of their gender and their answers to exactly four preference questions from the candidate study: whether they prefer a beach vacation or a mountain vacation, whether they prefer lions or tigers, whether they prefer light or dark colors, and whether they prefer spring or fall.

3.3 Evaluator Study

In our main evaluation study, evaluators are tasked with assessing whether candidates were Above-the-Bar based upon their profile and additional signals of performance. We provide evaluators with information about the candidate study. We explain that the candidates they will evaluate are real Prolific participants who completed a past study for us. We explain that these participants each completed a test that contained math and science questions, and that, based upon their score on this test, we classified each candidate as either Above-the-Bar or Below-the-Bar. We explain that, as evaluators, their job is to use information about a candidate to guess whether that candidate

was Above-the-Bar.

We randomly assign evaluators to one of the four pools. In addition, across subject, we vary whether the evaluator can choose whether to receive additional signals of performance (endogenous) or if instead the signals are provided with certainty (exogenous). Evaluators assess five candidates from their assigned pool, drawn at random with replacement.

3.3.1 Pool Information

We provide each evaluator with specific information about their candidate pool. We inform them of the overall share of candidates who were Above-the-Bar within their pool (60% for the 60 – 60 pool, 50% for the 60 – 40 and 40 – 60 pools, and 40% for the 40 – 40 pool). We tell them that there are exactly 25 men and 25 women in the candidate pool. And, most critically, we provide them with the share of men and women within their pool that were Above-the-Bar. We communicate this information with a pie chart. They see two pie charts side-by-side, one for men and one for women, with green wedges corresponding to the share of men (women) who are Above-the-Bar. Following the pie charts, we summarize the information in a table. Throughout their evaluations, evaluators can see this table of information about the candidate pool by reading the Instructions Summary at the bottom of each page. Figures A.1 and A.2 in the Appendix present examples.⁹

The goal is to provide clear, accurate, accessible information about the relevant candidate pool. In doing so, we aim to dictate evaluator beliefs of gender differences in performance.¹⁰ The pools allow us to induce exogenous variation in these beliefs.

For ease of exposition, we use the term **candidate prior**, or sometimes just **prior**, to refer to the share of candidates of a given gender who are Above-the-Bar within a pool. Because there are

⁹We also provide them with a breakdown according to each of the other profile characteristics. For example, we inform them what share of candidates who preferred mountain vacations were Above-the-Bar versus what share of candidates who preferred beach vacations were Above-the-Bar. We do the same for spring versus fall, lions versus tigers, and light colors versus dark colors. Note that by design, within each pool, the share of Above-the-Bar candidates of each of these preferences types is very similar and close to the pool average. For example, within the 40 – 40 pool, we aimed to have approximately 40% of both beach vacation types and mountain vacation types be Above-the-Bar.

¹⁰We do not observe participants' ex-ante beliefs about gender differences in this task. Past work suggests that for a math and science test of this type, which carries a more masculine stereotype, participants are likely to believe that men outperform women on average, see Bordalo, Coffman, et al., 2019 and Coffman, Exley, et al., 2021. While we do not observe ex-ante beliefs of gender differences, our approach (if we take our estimates at face-value) allows us to predict the behavior of an evaluator with any particular belief about the gender difference in the likelihood of being Above-the-Bar.

just two possible candidate priors (40 or 60), we write **low-prior** to refer to a candidate prior of 40, and **high-prior** to refer to a candidate prior of 60.¹¹

3.3.2 Initial Candidate Assessment

For each evaluation, the computer chooses one candidate at random from the pool of 50 candidates and shows the candidate profile to the evaluator. Figure A.3 provides a screenshot of an example candidate profile. The evaluator then guesses the likelihood that the candidate is Above-the-Bar. To indicate their answer, they use a slider that ranges from 0 percent likely to 100 percent likely; we use color-coding (moving from red to green) and intuitive labels on top of precise likelihoods (moving from not at all likely to extremely likely) to improve comprehension.

3.3.3 Information Acquisition

We explain to evaluators that, in order to help them make a better guess, they can pay to acquire (**endogenous** treatment) or will be given (**exogenous** treatment) additional information about the candidate. Additional information about candidates takes the form of noisy signals of performance. The signals are simple and binary, either **positive** or **negative**. These signals are randomly generated by the computer, with the probability of a positive signal dependent upon whether the candidate is Above-the-Bar. For Above-the-Bar candidates, the probability of a positive signal is 0.75; for Below-the-Bar candidates, the probability of a positive signal is 0.25. We explain these probabilities to evaluators using both percentages and expected frequencies. We use color-coding (green and red) and icons (plus sign or minus sign) to increase comprehension.

In the endogenous treatment, after completing the initial likelihood slider for a candidate, evaluators must decide whether to “Make Their Evaluation” or “Receive More Information.” If they choose to receive more information, they pay a small cost (5 cents, explained in more detail below) and receive one signal. Then, they guess again the likelihood that the candidate is Above-the-Bar using another slider. This process repeats, with evaluators being asked again whether they want to “Make Their Evaluation” or “Receive More Information.” Evaluators can choose to receive

¹¹Note that these semantic choices abstract away from a candidate’s membership in groups defined by characteristics other than their gender (such as beach vacation lovers, or dark color fans), though those characteristics will be accounted for in our empirical approach. Throughout, again for ease of exposition, the reader can interpret **group** as defined by gender.

up to five signals about each candidate.

The exogenous treatment operates identically, except instead of choosing whether to receive each signal, the evaluator must view all 5 signals for each candidate. Just as in the endogenous treatment, signals are shown sequentially and the evaluator must complete a new likelihood slider after each signal.

3.3.4 Final Evaluations

When an evaluator in the endogenous treatment chooses to forgo additional information and make their evaluation, they are asked to make a final binary classification of the candidate. They indicate whether they believe the candidate is Above-the-Bar, or not. Evaluators in the exogenous treatment make this evaluation after 5 signals for the candidate. Following their final evaluation, evaluators receive no feedback before moving to the next candidate.

In addition, to discourage evaluators in the endogenous treatment from skipping signals to save time in the study, we introduced a time penalty for each non-acquired signal.¹² The time penalty was designed to be similar to viewing and responding to a signal, both in terms of format and in terms of length of time. For each non-acquired signal for a candidate, the evaluator was shown a random number between 0 and 100, had to indicate whether that number was above or below 50, and then had to drag a slider to the chosen number.

3.3.5 Payments

All evaluators receive \$3.50 in guaranteed pay for completing the 20-minute study. In addition, we pay 1 out of every 10 participants a bonus based on their evaluation of the candidates. At the conclusion of the study, the computer randomly chooses 1 of the 5 candidates that the participant evaluated to determine the bonus payment. If the evaluator correctly classified the chosen candidate, they earn \$7 in additional bonus payment. If they incorrectly classified that candidate, they earn 50 cents in additional bonus payment. In the endogenous treatment, for each signal acquired for the chosen candidate, we deduct 5 cents from the bonus payment.

¹²For example, an evaluator who chose to view two signals for a candidate would be asked to complete three (5–2) time penalty tasks following their final evaluation of that candidate. We mention these filler tasks in the instructions.

In addition, evaluators chosen to earn incentive pay also earn money depending on their answer to one randomly-chosen likelihood slider question. One of their slider answers is chosen as the “slider-that-counts.” If their slider-that-counts is within 10 percentage points of the true probability, they receive an additional bonus payment of \$1.

4 Hypotheses and Empirical Approach

We study how evaluators use candidate gender in their assessments and how endogenizing the information acquisition process amplifies (or mitigates) any bias we observe. In this section, we lay out an empirical approach for analyzing these research questions.

Previous literature documents discrimination against women in many field contexts of interest (see Bertrand and Duflo, 2017 for an overview). However, in many cases, it is difficult to determine what drives this discrimination. One particular challenge is that, in part due to past discrimination, many of these environments feature significant gender gaps; as a result, we disproportionately observe evaluators making decisions in contexts where they believe (either accurately or inaccurately) that men have a greater likelihood of being qualified on average. This can make it hard to identify whether discrimination against women is driven by their status as a disadvantaged group—a group with a *relatively* lower likelihood of being qualified.

Our controlled experiment can tease out this channel. Our approach separates bias based purely upon a candidate’s gender from a more general form of bias that stems from (potentially implicit) comparisons across groups. In our setting, bias based purely upon a candidate’s gender would take the form of male candidates receiving more favorable assessments than female candidates, holding fixed other candidate-level characteristics. If, on the other hand, it is less about gender per se and more about relative comparisons, we would expect that candidate assessments depend not on gender but rather on how the candidate’s group compares to the other group.

We refer to this key measure as a candidate’s **relative advantage**, defined as the difference in priors across groups within the pool, signed so that positive values favor the candidate’s own group. Thus, a male candidate has a relative advantage of 20 within the 60 – 40 pool, of –20 within the 40 – 60 pool, and of 0 within the 40 – 40 and 60 – 60 pools; a female candidate’s relative

advantage is -20 , 20 , and 0 within those same pools, respectively.

We rely on a workhorse specification that predicts an evaluator’s assessment of a candidate from the candidate’s gender and relative advantage. We do so controlling for the candidate prior (40 or 60) and an indicator for whether the evaluator was assigned to the exogenous treatment. We cluster standard errors at the evaluator level. Formally, we predict an evaluation of a candidate i by evaluator j :

$$Evaluation_{i,j} = \beta_0 + \beta_1 Male_i + \beta_2 RelAdv_i + \beta_3 Prior_i + \beta_4 Exog_j + \mathbf{X}_i + \epsilon_{i,j} \quad (1)$$

where $Male_i$ is an indicator for candidate i being male, $RelAdv_i \in \{20, 0, -20\}$ is the candidate’s relative advantage within the pool, $Prior_i \in \{40, 60\}$ is the share of candidates of i ’s gender who are Above-the-Bar, and $Exog_j$ is an indicator for the evaluator’s treatment assignment. A positive β_1 represents gender bias, while a positive β_2 represents a bias that stems from contrast between groups. We can add interactions of the exogenous treatment indicator with $Male_i$ and $RelAdv_i$ to study how the magnitudes of these biases depend upon whether information acquisition is endogenous.

Recall that evaluators see four characteristics of each candidate in addition to gender: preference between fall/spring, preference between beach/mountain vacation, preference between light/dark colors, and preference between lions/tigers. Our control vector, \mathbf{X}_i , contains candidate i ’s relative advantage on each of these four dimensions.¹³ We control for these other relative advantages to rule out that differences among men and women along these other dimensions could drive our results.

An $Evaluation_{i,j}$ can correspond to different types of assessments. First, we observe the initial likelihood estimate the evaluator provides for the candidate, after seeing the candidate’s profile but before receiving any signals. We refer to this as the evaluator’s **initial assessment**. We

¹³For instance, suppose within a pool that 42% of candidates who liked lions but only 38% of candidates who liked tigers were Above-the-Bar, and candidate i liked lions. Then, we would define this candidate’s relative advantage on the lion/tiger dimension as 4. A candidate within this same pool who instead preferred tigers would have a relative advantage of -4 on this dimension. Relative advantage on the other dimensions is determined analogously. Each relative advantage enters the specification independently, so that each candidate has a relative advantage defined in terms of gender, plus four additional relative advantage terms corresponding to the four preference characteristics. Note that because we designed the pools to minimize differences in performance across these other dimensions, we have only minimal variation in relative advantage on these dimensions. This makes it difficult for us to benchmark the impact of relative advantage as defined by gender against relative advantage defined by these other dimensions.

also observe the evaluator’s updated likelihood assessment after each additional signal they view. Finally, we observe the evaluator’s final (binary) classification of the candidate as either Above-or Below-the-Bar. For evaluators in the exogenous treatment, this final classification is made after all five signals are viewed. In the endogenous treatment, evaluators make their final classification whenever they choose to stop receiving signals for the candidate and move to this stage. We refer to this binary classification as the evaluator’s **final evaluation**.

We then study information acquisition choices for evaluators in the endogenous treatment. We explore when evaluators choose to acquire another signal, and how these decisions depend upon a candidate’s gender and/or relative advantage. To explore this, we predict the decision to acquire signal t , where $t \in \{1, 2, 3, 4, 5\}$. We estimate the following specification:

$$Acquire_{i,j,t} = \beta_0 + \beta_1 Male_i + \beta_2 RelAdv_i + \beta_3 Prior_i + \beta_4 Likelihood_{i,j,0} + \mathbf{X}_i + \epsilon_{i,j,t} \quad (2)$$

where $Acquire_{i,j,t}$ is an indicator that takes 1 if evaluator j chose to receive signal t for candidate i , conditional on having viewed signal $t - 1$. We control for an evaluator’s initial assessment of a candidate, $Likelihood_{i,j,0}$, accounting for any impact of relative advantage or candidate gender that influences initial beliefs. We can run this specification separately by signal history. We will define **confirming signals**: confirming signals are consistent with the candidate prior (negative signals for low-prior candidates and positive signals for high-prior candidates). With this framework, we can explore whether candidate gender or relative advantage plays a larger role in decisions after confirming signals, when the accumulated evidence about the candidate tells a consistent story, or whether instead these biases are larger when the evidence is more mixed.

Finally, we study biases in belief updating. We restrict attention to only evaluators in the exogenous treatment. This cleanly isolates responses to information, shutting down any selection into receiving more information. For these exogenous treatment evaluators, we predict the updated likelihood assessment of evaluator j for candidate i after observing t signals, where $t \in \{0, 1, 2, 3, 4, 5\}$. To focus on the role of how beliefs are updated, we control for an evaluator’s initial assessment of a candidate, $Likelihood_{i,j,0}$. Formally,

$$Likelihood_{i,j,t} = \beta_0 + \beta_1 Male_i + \beta_2 RelAdv_i + \beta_3 Prior_i + \beta_4 Likelihood_{i,j,0} + \mathbf{X}_i + \epsilon_{i,j,t} \quad (3)$$

This allows us to ask whether, conditional on initial assessment, evaluators assess male candidates or candidates with a greater relative advantage more favorably after observing new information. Again, we can run this specification separately by candidate prior and signal history. This allows us to explore whether candidate gender or relative advantage plays a smaller role in belief updating after confirming signals, when there is less uncertainty about candidate quality.

We close by asking how much evidence evaluators require before making a final evaluation, and whether these requirements seem to be a function of a candidate’s gender and/or relative advantage. To explore this, we predict the total number of signals acquired by an evaluator before they decide to make their final evaluation. To allow for simple comparisons, we start by considering low and high-prior candidates separately.

We estimate the following specification, restricting attention to only low (or high) prior candidates:

$$TotalSignals_{i,j} = \beta_0 + \beta_1 Male_i + \beta_2 RelAdv_i + \mathbf{X}_i + \epsilon_{i,j} \quad (4)$$

where $TotalSignals_{i,j}$ is the total number of signals acquired by evaluator j before making a final evaluation for candidate i . We make the important distinction of splitting the sample based upon whether candidate i was classified as Above-the-Bar or Below-the-Bar by evaluator j . This allows us to ask how quickly evaluators classify different candidates as Above or Below-the-Bar.

Consider the hypotheses for low-prior candidates. First, we focus on candidates who were classified as Below-the-Bar. In this model, $\beta_1 > 0$ would indicate that evaluators required more signals before classifying low-prior male candidates as Below-the-Bar relative to low-prior female candidates. Similarly, $\beta_2 > 0$ would indicate that evaluators required more signals before classifying low-prior candidates from the 40 – 40 pool as Below-the-Bar relative to low-prior candidates from the 40 – 60 pool. Put differently, we can ask whether low-prior female candidates and/or relatively disadvantaged candidates are quicker to be classified as Below-the-Bar; do evaluators require less evidence before dismissing these candidates as Below-the-Bar? Analogously, by estimating this same specification for the set of low-prior candidates who were classified as Above-the-Bar, we can ask whether evaluators are *slower* to classify low-prior female candidates and/or relatively disadvantaged candidates as Above-the-Bar; do evaluators demand more evidence before being willing to classify these candidates as Above-the-Bar? Note that restricting to candidates from low-

prior groups rules out that any differences are driven by different signal realizations. All candidates in these specifications have the same true likelihood of being Above-the-Bar and therefore produce the same signals in expectation.

A similar set of hypotheses can be tested for high-prior candidates. First, we can focus on high-prior candidates who were classified as Above-the-Bar, asking whether evaluators require fewer signals before classifying male candidates (relative to female candidates) or advantaged candidates (relative to candidates in the 60 – 60 pool) as Above-the-Bar. And, on the flip side, we can restrict attention to high-prior candidates who were classified as Below-the-Bar, hypothesizing that evaluators may be slower to classify male candidates and advantaged candidates as Below-the-Bar.

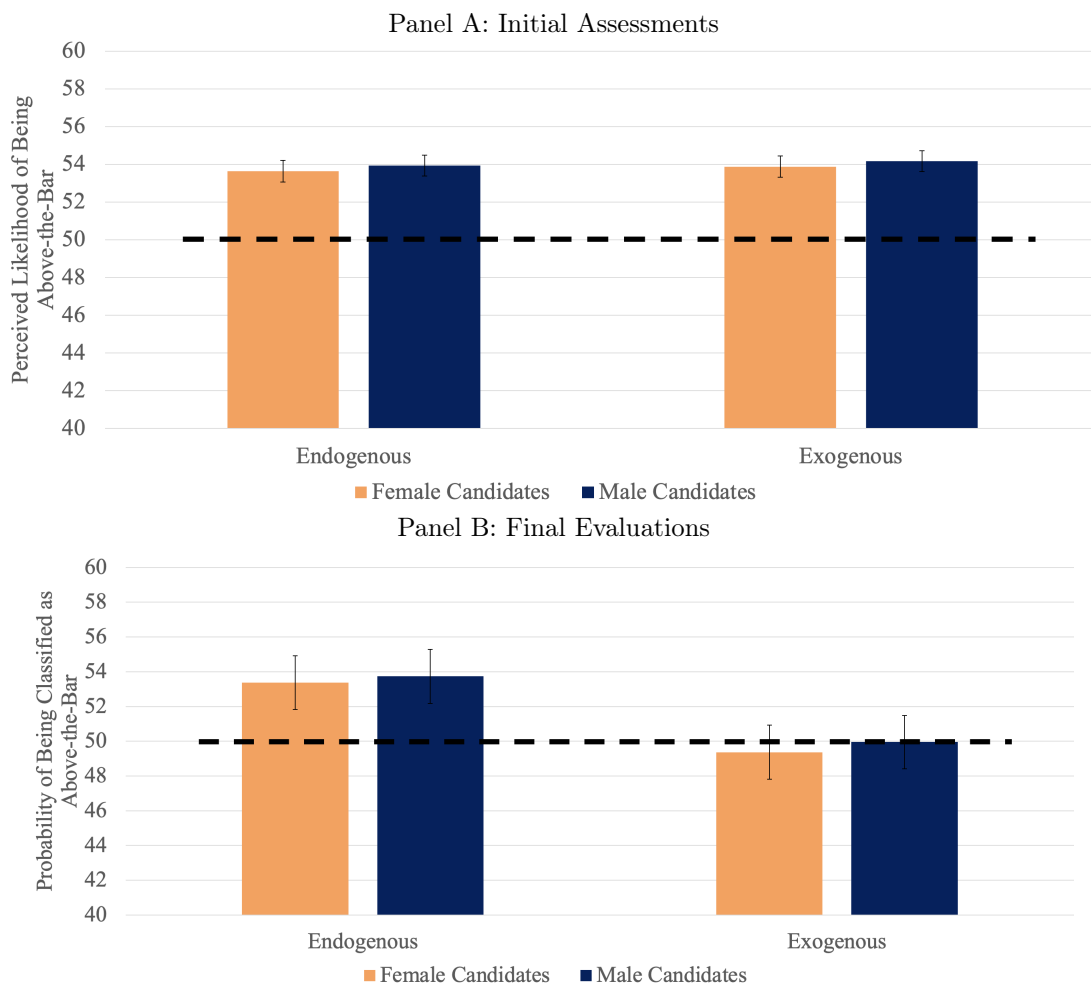
5 Results

We start by providing an overview of our evidence on assessments, exploring discrimination on the basis of gender and relative advantage. Figure 1 presents the average assessment of male and female candidates. Panel A presents initial assessments, while Panel B shows the share of male and female candidates who are classified as Above-the-Bar in final evaluations. We find no evidence that evaluators discriminate on the basis of gender. Evaluators provide very similar assessments of male and female candidates—initially and after acquiring information. This is true independent of whether the information was acquired endogenously or exogenously.

Initially, evaluators believe that both male and female candidates have approximately a 54% chance of being Above-the-Bar on average. In the endogenous treatment, 54% of male candidates and 53% of female candidates are ultimately classified as Above-the-Bar in final evaluations, tracking initial assessments; this gender difference is not significant. We see that evaluators in the exogenous treatment make significantly less optimistic final evaluations; more information seems to bring evaluations closer to the truth, on average. There continues to be no significant gender difference, with approximately 50% of both male and female candidates classified as Above-the-Bar.

Figure 2 presents the same data, but now comparing assessments on the basis of group differences. Panel A presents initial assessments. We show initial assessments for low-prior groups,

Figure 1: Initial Beliefs and Final Evaluations by Gender



Note: This figure presents the average assessment of male and female candidates, for both treatment groups. Panel A shows initial assessments, while Panel B shows final assessments. The dashed lines show the true percentage of candidates in that group in the pool that are Above-the-Bar. A 95% confidence interval is shown at the top of each bar.

comparing candidates from the 40 – 40 pool, in which members of both groups have an equal chance of being Above-the-Bar, to low-prior candidates in the 60 – 40 or 40 – 60 pools, who are at a performance disadvantage relative to the other group. We observe that low-prior candidates are assessed significantly less favorably when their group is relatively disadvantaged. Candidates from high-prior groups are assessed significantly more favorably if their group holds a relative

performance advantage compared to when both groups are equal.

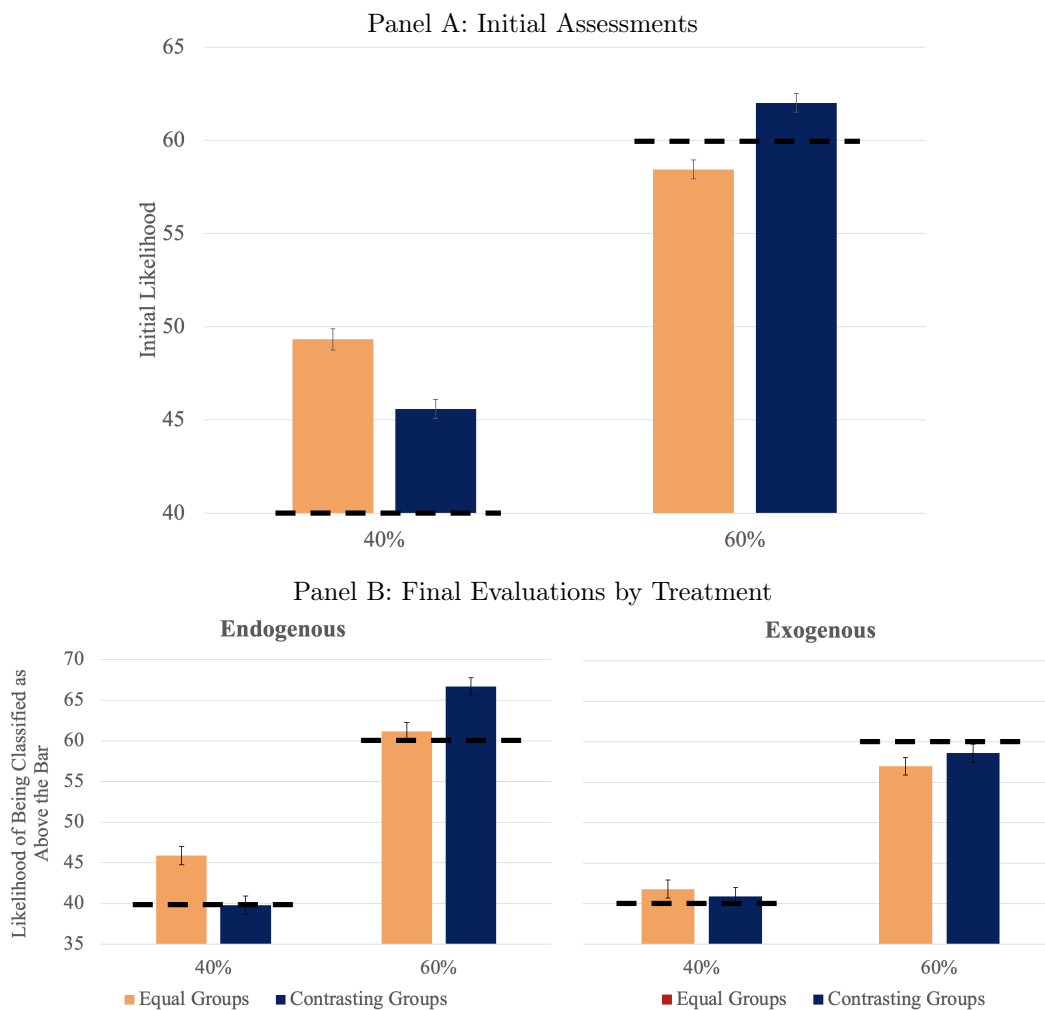
Panel B examines these same comparisons, now focusing on final evaluations by treatment. Within the endogenous treatment, relative performance advantage continues to play a significant role in final evaluations. Nearly 46% of low-prior candidates are classified as Above-the-Bar when the groups are equal, compared to only 40% of low-prior candidates from disadvantaged groups. Relative advantage benefits high-prior candidates in the endogenous treatment: 67% of high-prior candidates are viewed as Above-the-Bar when their group has the performance advantage. Only 61% are viewed as Above-the-Bar when the groups are equal.

Final evaluations depend on relative advantage in the endogenous treatment. But, as Figure 2 reveals, this does not hold true in the exogenous treatment. When evaluators must acquire all five signals about each candidate, final evaluations are more accurate on average and there is less of a role for group comparisons in shaping perceptions of candidates.

Table 2 formalizes these patterns. We predict an evaluator’s assessment of a candidate—either initial likelihood estimate, final likelihood estimate, or final binary classification—from our two key variables: whether the candidate is male and the candidate’s relative advantage. We control for candidate prior (40 or 60) as well as the candidate’s relative advantage in terms of the other profile characteristics. We include round fixed effects and cluster standard errors at the evaluator level. In Columns 1, 3, and 5, we simply include an indicator for the exogenous treatment. In Columns 2, 4, and 6, we interact the exogenous treatment with male and relative advantage, asking if their impact is smaller when information is required.

The results are consistent with our takeaways from Figures 1 and 2. Evaluators do not assess male candidates significantly more positively than female candidates, but relative advantage matters. We estimate that, initially, one point of relative advantage increases the believed likelihood of that candidate being Above-the-Bar by 0.16 percentage points ($p < 0.01$, Column 1). Acquiring information on average reduces this bias, shrinking the impact of a point of relative advantage by roughly half (to 0.08pp) in final likelihood assessments ($p < 0.01$, Column 3). In terms of final binary evaluations, we estimate that one point of relative advantage increases a candidate’s

Figure 2: Initial Beliefs and Final Evaluations by Relative Advantage



Note: This figure presents the average assessments broken down by prior likelihood and equal/contrasting grouping, for both treatment groups. Panel A shows initial assessments, while Panel B shows final evaluations. The dashed lines show the true percentage of candidates in that group in the pool that are Above-the-Bar. A 95% confidence interval is shown at the top of each bar.

Table 2: The Impact of Relative Advantage on Assessments

OLS Predicting Initial Beliefs, Final Beliefs, and Final Evaluations

	Initial Slider		Final Slider		Final Eval	
	(1)	(2)	(3)	(4)	(5)	(6)
Relative Advantage	0.157*** (0.0177)	0.164*** (0.0208)	0.0779*** (0.0272)	0.105*** (0.0323)	0.148*** (0.0398)	0.267*** (0.0499)
Male Candidate	0.0105 (0.292)	0.128 (0.419)	0.145 (0.531)	0.440 (0.705)	0.267 (0.789)	0.342 (1.141)
Group Prior	0.456*** (0.0282)	0.456*** (0.0282)	0.550*** (0.0393)	0.550*** (0.0393)	0.778*** (0.0558)	0.778*** (0.0558)
Exogenous	0.242 (0.368)	0.357 (0.464)	-2.726*** (0.531)	-2.438*** (0.758)	-3.854*** (0.757)	-3.808*** (1.083)
Rel Adv x Exog		-0.0145 (0.0217)		-0.0524 (0.0377)		-0.235*** (0.0565)
Male x Exog		-0.230 (0.585)		-0.578 (1.060)		-0.103 (1.571)
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes	Yes	Yes
No. Observations	16000	16000	16000	16000	16000	16000
No. Clusters	3200	3200	3200	3200	3200	3200
Adj. R2	0.168	0.168	0.0411	0.0411	0.0409	0.0419

Notes: Standard errors are reported in parentheses and clustered at the evaluator level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In the first two columns, the dependent variable is the likelihood of the candidate being Above-the-Bar chosen by the evaluator in the first slider. Columns 3 and 4 report the regressions on the likelihood of the candidate being Above-the-Bar chosen by the evaluator in the last slider before making their final decision. The last two columns report the regressions of the indicator variable taking value 1, if the final evaluation of the candidate was Above-the-Bar, and 0 otherwise. All regressions include round fixed effects and candidate advantage by filler characteristics.

likelihood of being classified as Above-the-Bar by 0.15pp ($p < 0.01$, Column 5).¹⁴

When we turn to treatment differences, we see that the exogenous provision of information reduces the bias induced by group differences. One point of relative advantage increases an evaluator’s final likelihood slider for the candidate by 0.11pp in the endogenous treatment ($p < 0.01$, Column 4), but by only 0.05pp in the exogenous treatment (not statistically significantly different

¹⁴Note that, overall, our data on beliefs suggest a high-level of comprehension and attention. In 92% of cases, evaluators update their beliefs in the correct direction after observing a signal. More than half of our evaluators update their beliefs in the correct direction every time, and 90% of evaluators update their beliefs in the correct direction at least 75% of the time. Their stated beliefs are also highly consistent with their final evaluations. Define an internally consistent final evaluation as either (i) classifying the candidate as Above-the-Bar and having a final slider likelihood greater than or equal to 50%, or (ii) classifying the candidate as Below-the-Bar and having a final slider likelihood less than or equal to 50%. By this definition, 97% of final evaluations are internally consistent.

than 0, but we also cannot reject that the impact is the same across the two treatments). The differences are more stark in terms of final binary evaluations. Relative advantage has a significant and sizable impact in the endogenous treatment: we estimate that a candidate from the advantaged group is approximately 5pp more likely to be deemed Above-the-Bar relative to the same candidate when both groups are equal ($p < 0.01$, Column 6, computed by multiplying the coefficient on relative advantage, 0.267, by the amount of relative advantage, 20). But, we estimate no significant impact of relative advantage on final evaluations in the exogenous treatment.

Our analysis reveals that evaluators' initial assessments depend, in part, upon comparisons across groups. When information is provided exogenously, relative differences across groups no longer predict final evaluations. But, when evaluators have the choice of whether and how much information to acquire, final evaluations remain biased by comparisons across groups. Evaluators treat candidates more favorably when their group is advantaged relative to the other, and less favorably when their group is disadvantaged relative to the other, even after the chance to acquire individualizing information about the candidate.

In the following sections, we consider two reasons why access to information may fail to debias evaluators. First, we consider under-acquisition of information. Then, we consider biased processing of information.

5.1 Information Acquisition

In this section, we analyze the information acquisition decisions made by evaluators in the endogenous treatment. Figure A.4 in the Appendix presents the distribution of number of signals acquired by pool, revealing no significant differences in the number of signals acquired across pool type. Importantly, this suggests that the mere existence of differences across groups did not alter the amount of information sought by evaluators on average, perhaps reducing concerns that effort, focus, or attention varied across treatment. On average, evaluators acquire 2.3 signals about a candidate. In 25% of evaluator-candidate observations, evaluators choose to receive no signals

about the candidate. The median number of signals acquired is 2.¹⁵

This represents substantial under-acquisition of information. To see this, consider a Bayesian counterfactual. Our risk-neutral Bayesian holds a prior belief equal to the candidate prior (so, 40 for a low-prior candidate and 60 for a high-prior candidate) and uses Bayes' rule to update their beliefs. When deciding whether to acquire another signal, they weigh the gain in final evaluation accuracy that a signal (and the option to buy additional signals) would provide against the cost of the signal (5 cents) and choose to buy a signal if it would increase their earnings in expectation. We can compare the behavior of our participants to how this Bayesian would behave. Appendix Figure A.8 presents the results, showing the total number of signals per candidate acquired by our evaluators compared to the total number of signals acquired by the Bayesian in these same cases. The Bayesian always buys at least 3 signals; our evaluators acquire fewer than 3 signals in 54% of cases, suggesting significant under-acquisition on average. We can also use a Bayesian benchmark to compute the expected dollar value of purchasing another signal at the moment a decision-maker chooses to stop acquiring signals. In approximately half of cases, acquiring another signal had an expected value of more than 50 cents (given the evaluator's reported likelihood at the moment they stopped acquiring signals). In a third of cases, the expected value was more than \$1.

Evaluators in our endogenous treatment are leaving money on the table, both relative to this Bayesian benchmark and relative to evaluators in the exogenous treatment who were forced to acquire information. Even if we factor in a 5 cent signal cost in both treatments, evaluators in the exogenous treatment earn significantly more money from their final evaluations. It is somewhat striking that we see significant under-acquisition of information even when signals are relatively cheap in terms of both money and time.

5.1.1 When and Why Do Evaluators Under-Acquire Information?

What predicts the decision to seek more information about a candidate? First, we show that more biased evaluators—in terms of their propensity to rely on differences across groups in their initial

¹⁵Figure A.6 in the Appendix presents the average total number of signals acquired by candidate order. We do find that participants acquire fewer signals for the first candidate they assess. We speculate that participants may under-estimate the time-cost or annoyance of the filler task for skipped signals initially. After directly experiencing the filler task following the first candidate, they may revise downwards the net cost of a signal, leading them to acquire more signals for candidates 2 - 5. We show that our results are robust to excluding decisions for the first candidate in Figure A.7 and Tables A.2, A.3, and A.4 in the Appendix.

assessments—are less likely to seek additional information. Drawing on the specification of Table 2, Column 1, Table 3 predicts an evaluator’s **initial** assessment of a candidate from the candidate’s gender and their group’s relative advantage. In Column 1, we consider the data for evaluators who chose to make a final evaluation of the candidate immediately, without acquiring any signals for that candidate. In Column 2, we consider the data from all other evaluators in the endogenous treatment, those that went on to acquire anywhere between one and five signals for the candidate. In both columns, we are focused on the initial assessments of evaluators. We are asking whether, prior to receiving any additional information about the candidate, evaluators who choose to receive more information look different than evaluators who make immediate decisions. Indeed, that seems to be the case. We estimate that evaluators who choose to make their final evaluation immediately are significantly more reliant on group differences in their initial assessments than evaluators who go on to seek more information. One point of relative advantage increases a candidate’s initial assessment by 0.14pp for evaluators who choose more information, compared to increasing it by 0.28pp for evaluators who make their final evaluations without any more information. We present an interacted model in Column 3, finding that the extent to which group differences impact initial assessments varies significantly with the evaluator’s subsequent information choice ($p < 0.05$).

Column 4 shows that evaluators who choose not to receive more information also produce more biased final evaluations. For evaluators who do not seek out information, we estimate that every point of relative advantage increases the chance of a candidate being classified as Above-the-Bar by 0.8pp ($p < 0.01$). For a sense of magnitudes, this translates to a candidate from a high-prior group being 16pp more likely to be classified as Above-the-Bar if they are in a pool with some low-prior candidates rather than in a pool with all high-prior candidates. Among evaluators who seek out more information before making a final decision, relative advantage has no significant impact on final evaluations ($p < 0.01$ for difference across evaluator types).

Next, we focus on evaluators who did choose to receive more information and explore what predicts the decision to receive another signal about a candidate. Recall that we define a confirming signal as a signal that is consistent with the candidate’s prior: a positive signal for a candidate from a high-prior group, or a negative signal for a candidate from a low-prior group (see Section 4 for more details).

Table 3: Selection Into Acquiring Information
 OLS Predicting Initial Likelihoods and Final Evaluations by Subsample

	Immediate Evaluation	Got More Info	Pooled	
	Initial Slider	Initial Slider	Initial Slider	Final Eval
Relative Advantage	0.28*** (0.056)	0.14*** (0.025)	0.24*** (0.043)	0.78*** (0.085)
Male Candidate	-0.77 (1.04)	0.41 (0.42)	-0.78 (1.05)	-3.26 (2.19)
Group Prior	0.34*** (0.083)	0.46*** (0.040)	0.43*** (0.038)	0.78*** (0.079)
Got More Info			-4.89*** (0.84)	-13.2*** (1.75)
Got More Info x Rel Adv			-0.093** (0.041)	-0.70*** (0.088)
Got More Info x Male			1.16 (1.12)	4.43* (2.53)
Round FE	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes
No. Observations	1982	5988	7970	7970
No. Clusters	717	1424	1594	1594
Adj. R2	0.16	0.18	0.18	0.070

Notes: Standard errors are reported in parentheses and clustered at the evaluator level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable in all columns is the likelihood of the candidate being Above-the-Bar chosen by the evaluator in the first slider before receiving any signals. The first column include the sample of observations where the evaluation was made without acquiring any information. The second column includes the observations where evaluation was made after receiving at least one signal. The third and fourth columns pool all observations and estimates the effect of receiving any signals on the initial likelihood assessment and final evaluation, respectively. All regressions includes controls for round fixed effects and candidate advantage by filler characteristics. All three columns include endogenous treatment only.

Intuitively, we expect that receiving a confirming signal should reduce the likelihood that an evaluator chooses to receive another signal. For example, suppose an evaluator is assessing a candidate from a high-prior group. The evaluator chooses to receive a signal, and the signal is positive. Now, the evaluator knows the candidate both comes from the high-prior group and has generated a positive signal, both consistent with the candidate being Above-the-Bar. This should make the evaluator feel more ready to reach a final evaluation, classifying the candidate as Above-the-Bar, relative to a situation where that same high-prior candidate has generated a negative signal. This logic leads us to expect that, in general, receiving a confirming signal should reduce the likelihood of acquiring another signal.

Table 4 tests this hypothesis. We predict an evaluator’s decision to acquire another signal from an indicator for whether the evaluator has just received a confirming signal. In Panel A, Columns 1–4, we analyze decisions after the first, second, third, and fourth signals, respectively. Note that there is selection into each additional signal, potentially in terms of evaluator characteristics and signal histories, and a correspondingly smaller sample size as we consider later signal realizations. With that note of caution, we investigate whether confirming signals reduce the likelihood of receiving more information. For early signals, we see evidence consistent with the hypothesis. We estimate that receiving a confirming first signal reduces the likelihood of pursuing another signal by 3pp ($p < 0.01$), while receiving a confirming signal as the second signal (independent of whether the first signal was confirming or not), reduces the likelihood of pursuing a third signal by 8pp ($p < 0.01$). We do not find this effect for the third and fourth signals. The final column pools over all signals, continuing to predict the decision to receive more information from whether the directly preceding signal was confirming. We estimate that, on average, a confirming signal reduces the likelihood of acquiring another signal by 3pp ($p < 0.01$).

We interpret this result as illustrating that confirming signals make it “easier” for evaluators to reach a decision about a candidate. This motivates the important question of how contrast between groups interacts with this effect. Consider our previous example, with an evaluator assessing a candidate from a high-prior group who has just received a confirming (positive) signal. Table 4 revealed that this evaluator will be less likely to seek out additional information (relative to having observed a negative signal). In our next analysis, we ask whether this effect is amplified when there are differences across groups. Consider a high-prior candidate for whom an evaluator has seen a positive signal. We ask whether the impact of that confirming signal will be larger when the candidate’s group is advantaged (60 – 40 pool) relative to when the two groups are equal (60 – 60 pool). That is, does the candidate being in the advantaged group make it even more likely that the evaluator feels convinced by the confirming signal, forgoes additional information, and classifies the candidate as Above-the-Bar?

Considering the case of the low-prior candidate helps to crystallize the hypothesis and motivate our empirical approach. Suppose an evaluator is assessing a low-prior candidate and has just seen a negative signal. Our hypothesis is that it is even more likely that the evaluator feels convinced

Table 4: Endogenous Information Acquisition

OLS Predicting the Decision to Get More Info After...

Panel A: Overall					
	1st Signal	2nd Signal	3rd Signal	4th Signal	Pooled-Last Signal
Confirming Signal	-0.031*** (0.0092)	-0.079*** (0.013)	-0.016 (0.016)	-0.0068 (0.020)	-0.027*** (0.0053)
Round FE	Yes	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes	Yes
No. Observations	5988	5045	3633	2287	18515
No. Clusters	1424	1287	1138	826	1424
Adj. R2	0.043	0.018	0.0029	0.00012	0.0050

Panel B: Comparing Pools with Equal Versus Contrasting Pools					
	1st Signal	2nd Signal	3rd Signal	4th Signal	Pooled-Last Signal
Confirming Signal	-0.054*** (0.013)	-0.10*** (0.019)	-0.025 (0.024)	-0.031 (0.030)	-0.042*** (0.0078)
Equal Groups	-0.026 (0.016)	-0.014 (0.021)	0.0081 (0.028)	0.013 (0.033)	-0.0095 (0.0082)
Equal Groups x Conf. Signal	0.046** (0.018)	0.047* (0.025)	0.019 (0.032)	0.044 (0.040)	0.029*** (0.011)
Round FE	Yes	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes	Yes
No. Observations	5988	5045	3633	2287	18515
No. Clusters	1424	1287	1138	826	1424
Adj. R2	0.043	0.018	0.0029	0.0017	0.0052

Notes: Standard errors are reported in parentheses and clustered at the evaluator level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions includes controls for relative advantage, male candidate indicator, group prior, initial likelihood assessment, round fixed effects, and candidate advantage by filler characteristics. All three columns include endogenous treatment only. The dependent variable is an indicator variable for deciding to receive a signal after the corresponding signal in each column for the corresponding subsample. The last column pools all observations in each round in the long format.

by this confirming signal when the candidate’s group is disadvantaged relative to the other group (40 – 60) compared to when both groups are equal (40 – 40 pool). Pulling the two examples together, we are hypothesizing that confirming signals have a larger impact—in terms of reducing the likelihood of pursuing more information—when the two groups are contrasting relative to when the two groups are equal.

Panel B of Table 4 explores this question, augmenting the specifications of Panel A. We include an indicator for whether the candidate pool featured equal groups (40 – 40 or 60 – 60), instead of contrasting groups (60 – 40, 40 – 60), and we interact this indicator with having just observed a confirming signal. We find that contrast between groups does amplify the impact of confirming signals. Evaluators are 5pp less likely to get more information after seeing a confirming first signal when the groups are contrasting (Column 1, $p < 0.01$). This effect is eliminated when the two groups are equal ($p < 0.05$ for difference across pool types). Responses after the second through fourth signals are directionally consistent with this same pattern, though more noisily estimated. Column 5 considers all of the data. We estimate that receiving a confirming signal (relative to a disconfirming signal) reduces the likelihood of seeking more information by 4pp when groups are contrasting ($p < 0.01$), but by only 1pp when groups are equal ($p < 0.01$ for difference across pool types).

This finding suggests that group differences and confirming signals reinforce each other. A positive signal for a high-prior candidate is particularly convincing when their group is also advantaged relative to the other group; a negative signal for a low-prior candidate is particularly convincing when their group is also disadvantaged relative to the other group.

Our results highlight two ways in which differences across groups shape information acquisition decisions. First, evaluators who choose not to pursue more information are more influenced by group differences on average, both in terms of their initial assessments and in terms of their final evaluations. Second, receiving a confirming signal reduces the likelihood that an evaluator seeks out an additional signal, significantly more so when there are differences across the groups. Both of these findings illustrate how a reliance on group differences in assessments can be reinforced when evaluators must actively seek out additional information. Of course, the other way in which group differences may be reinforced is through biased processing of signals. We explore this in the

next section.

5.2 Biased Belief Updating

Evaluators provide beliefs after every signal they acquire, allowing us to investigate how individuals update their beliefs in response to new information. Our focus is understanding whether candidate gender and/or relative advantage predicts updating. While this analysis could be done on the full sample, in the main text we analyze only evaluators in the exogenous treatment. This rules out selection into additional information, isolating the role for biased processing. In the Appendix, we replicate these results with the full sample reported in Table A.1; while the biases we document are stronger in the exogenous treatment, absent selection, we observe similar patterns when considering the full sample.

To provide intuition and motivate a more unified analysis, we begin by analyzing the data separately for low and high-prior candidates. Table 5, Panel A considers assessments of low-prior candidates. We predict the believed likelihood of the candidate being Above-the-Bar after different signal histories, controlling for relative advantage, an indicator for a male candidate, and our standard controls. In addition, we control for the individual evaluator’s initial assessment; any impact we observe of relative advantage on subsequent beliefs is on top of this initial bias.¹⁶ Columns 1 and 2 consider beliefs after a positive and negative signal, respectively. We estimate that relative advantage has a significant impact on belief updating, but only when the signal is positive. When a low-prior candidate receives a negative signal, it does not matter whether the groups were equal or their group was disadvantaged (Column 2). But, if a low-prior candidate receives a positive signal, the candidate gets a significantly smaller boost from the positive signal when they are disadvantaged ($p < 0.01$, Column 1). When we consider updating after two signals, both positive, both negative, or one of each, we find a directionally similar pattern, though no significant differences. It seems to be the case that a low-prior candidate benefits less from positive news when they are disadvantaged.

¹⁶Prior to receiving any signals, evaluators believe low-prior candidates have a 2.2pp lower chance of being Above-the-Bar when they are in the disadvantaged group, relative to the equal groups case ($p < 0.01$).

Panel B turns attention to high-prior candidates.¹⁷ Columns 1 and 2 consider beliefs after the first signal, conditional on initial assessment. The results are flipped relative to low-prior candidates. For high-prior candidates, relative advantage has a significant impact on belief updating after a negative signal, but not after a positive signal. After a negative signal, evaluators maintain significantly more optimistic beliefs for a high-prior candidate when they are advantaged relative to when the two groups are equal (Column 2, $p < 0.05$). We see a similar protective impact of relative advantage after two signals. Conditional on initial assessments, beliefs are 3.6pp more favorable for high-prior candidates with two negative signals when they are advantaged relative to equal (Column 5, $p < 0.01$). The effect of relative advantage is smaller when the signals are mixed (at 1.5pp, Column 4, $p < 0.05$), and non-existent when both signals are positive.

Together, the results for low and high-prior candidates seem to tell a clear story. When evaluators observe early confirming signals, relative advantage does not play a significant role in shaping how beliefs are updated. But, when early signals are disconfirming, relative advantage shapes how evaluators respond to the mixed picture. Low-prior candidates seem to get less of a boost out of a positive signal when they are disadvantaged, relative to when the groups are equal; high-prior candidates seem to get more of the benefit of the doubt after a bad signal when they are advantaged, relative to when the groups are equal. These patterns are inconsistent with any Bayesian story for behavior, as candidates with the same prior probability of being Above-the-Bar and the same signal history are assessed differently on the basis of an irrelevant comparison group.

In Table 6, we use the insights from responses to early signals to build and analyze a more general model of behavior. We want to understand the impact of relative advantage depending upon whether the evidence has been mostly confirming (in terms of consistency with the prior). To do so, we predict an evaluator’s belief about a candidate after signal t from the share of the t signals that have been confirming. For a low-prior candidate, this is the share of negative signals seen so far; for a high-prior candidate, this is the share of positive signals seen so far. To explore whether biases are larger when the evidence has been more mixed, we interact this share of confirming signals variable with relative advantage. A negative interaction term suggests that

¹⁷In terms of initial assessments, prior to receiving signals, evaluators believe high-prior candidates have a 3.2pp greater chance of being Above-the-Bar when they are in the advantaged group relative to when both groups are equal ($p < 0.01$). We control for these initial assessments in the analysis presented.

Table 5: Biased Belief Updating
 OLS Predicting Beliefs After Given Signal History

Panel A: Only Low-Prior Sample (Exogenous):					
	1 Positive	1 Negative	2 Positives	2 Mixed	2 Negatives
Relative Advantage	0.111*** (0.0351)	-0.00154 (0.0356)	0.0522 (0.0644)	0.0308 (0.0347)	-0.0410 (0.0494)
Male Candidate	0.766 (0.601)	-0.259 (0.625)	0.417 (1.001)	0.380 (0.634)	-0.909 (0.879)
Initial Likelihood	0.699*** (0.0250)	0.614*** (0.0406)	0.481*** (0.0383)	0.634*** (0.0322)	0.501*** (0.0434)
Round FE	Yes	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes	Yes
No. Observations	1809	2166	1043	1543	1389
No. Clusters	959	1042	719	894	864
Adj. R2	0.540	0.402	0.263	0.465	0.244
Panel B: Only High-Prior Sample (Exogenous):					
	1 Positive	1 Negative	2 Positives	2 Mixed	2 Negatives
Relative Advantage	-0.00456 (0.0251)	0.0936** (0.0404)	0.0207 (0.0418)	0.0748** (0.0326)	0.175*** (0.0644)
Male Candidate	-0.241 (0.455)	0.409 (0.705)	-0.182 (0.724)	0.0725 (0.586)	1.271 (1.130)
Initial Likelihood	0.722*** (0.0254)	0.665*** (0.0406)	0.544*** (0.0360)	0.701*** (0.0289)	0.443*** (0.0486)
Round FE	Yes	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes	Yes
No. Observations	2184	1871	1457	1515	1083
No. Clusters	1050	978	863	895	728
Adj. R2	0.628	0.383	0.391	0.522	0.146

Notes: Standard errors are reported in parentheses and clustered at the evaluator level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table reports the OLS results predicting beliefs about the likelihood of being Above-the-Bar for only the evaluators in the exogenous treatment. We split the sample by candidate prior and run the analysis on the subsamples including only low-prior candidates and only high-prior candidates in Panel A and B, respectively. The dependent variable in the first two columns is the evaluator's believed likelihood of the candidate being Above-the-Bar after receiving the first signal. The following columns report the regressions on the believed likelihood of the candidate being Above-the-Bar after receiving two signals. All regressions include round fixed effects and candidate advantage by filler characteristics.

relative advantage benefits a candidate relatively less when the evidence has been more consistent with the candidate's prior

Table 6. Biased Belief Updating Beyond Two Signals (Exogenous Treatment Only)

OLS Predicting Beliefs After Given Signal History

	Low Prior		High Prior	
	Belief after Signal t	Above-the-Bar?	Belief after Signal t	Above-the-Bar?
Relative Advantage	0.074* (0.045)	0.22** (0.099)	0.16*** (0.048)	0.27*** (0.10)
Male Candidate	1.04 (0.75)	1.46 (1.83)	1.43* (0.80)	2.45 (1.85)
Share of Confirming Signals	-0.47*** (0.013)	-1.25*** (0.025)	0.50*** (0.011)	1.29*** (0.020)
Sh. of Conf. Signals x Rel. Adv.	-0.0012 (0.00076)	-0.0037** (0.0015)	-0.0019*** (0.00068)	-0.0029** (0.0014)
Sh. of Conf. Signals x Male	-0.022* (0.012)	-0.017 (0.026)	-0.016 (0.011)	-0.038 (0.024)
Round FE	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes
No. Observations	19875	3975	20275	4055
No. Clusters	1180	1180	1198	1198
Adj. R2	0.51	0.61	0.51	0.62

Notes: Standard errors are reported in parentheses and clustered at the evaluator level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the perceived likelihood of the candidate being Above-the-Bar after receiving t signals in Columns 1 and 3 for low and high-prior candidates, respectively. Columns 2 and 4 include only last round, final evaluation round, where the dependent variable is whether the candidate is classified as Above-the-Bar for low and high-prior candidates, respectively. All columns include observations from the exogenous information acquisition treatment only. All regressions include round fixed effects and candidate advantage by filler characteristics as well as initial likelihood chosen by the participant before any signals.

Columns 1 and 2 consider low-prior candidates. The first column predicts the evaluator's belief of a candidate's likelihood of being Above-the-Bar after signal t . As we did in Table 5, we control for the evaluator's initial assessment of the candidate, so that any estimated impact of relative advantage on later beliefs can be interpreted as an impact on updating behavior. The negative coefficient on the share of confirming signals tells us that, as expected, as the share of negative signals increases, evaluator beliefs are more pessimistic. We do not observe a significant interaction of relative advantage with the share of confirming signals, suggesting that bias against disadvantaged groups is, on average, no greater when the evidence has been more mixed ($p=0.107$).¹⁸ In

¹⁸To give a sense of effect sizes, we estimate that moving from all positive signals to all negative signals reduces the believed likelihood of being Above-the-Bar by 47pp for a candidate in the 40 – 40 pool (coefficient on confirming share in Column 1). The (insignificant) negative interaction term suggests that this reduction is approximately 49pp for a low-prior candidate in the 40 – 60 pool.

Column 2, we bring this same empirical approach to analyzing final classifications, asking whether relative advantage plays a larger role in final evaluations when signals are less in line with the candidate prior. Indeed this seems to be the case. As the share of confirming signals increases (negative signals for low-prior candidates), relative advantage plays a significantly smaller role in how that candidate is classified.

Columns 3 and 4 present the same analysis for high-prior candidates. In these specifications, a larger share of confirming signals indicates more positive signals for the candidate. We see that, as expected, this translates into a greater believed likelihood of the candidate being Above-the-Bar after signal t and a greater chance of the candidate ultimately being classified as Above-the-Bar. Relative advantage also strongly predicts positive assessments, both in terms of beliefs after signal t and in final classifications. However, we see a significant and sizable negative interaction. High-prior candidates benefit less from being relatively advantaged when most signals have been positive compared to when signals have been more mixed.

Our data shows that initial assessments are biased by group differences. These same group differences predict both the decision to get more information and reactions to new information. In particular, group comparisons predict how evaluators react to mixed evidence. When the picture seems clear, i.e., information is more consistent with the candidate prior, evaluators update their beliefs similarly independent of the candidate's relative advantage. On the other hand, group comparisons play a significant role in the case of exceptions, i.e., when information is less consistent with the candidate prior. Being from an advantaged group seems to mitigate the impact of a negative signal. And, positive signals seem to carry less weight for candidates from disadvantaged groups. These forces have implications for the accuracy of evaluations, particularly for candidates who are not typical of their pool.

5.3 Accuracy of Final Evaluations

We have shown that group differences impact both the decision to acquire more information and how that information is incorporated into beliefs. What is the cumulative impact of these decisions on final evaluations? To answer this question, we must focus on the endogenous treatment where both of these factors are at work. Figure 3 presents CDFs of the number of signals acquired

before an evaluator makes their final decision, splitting the data by candidate advantage. Panel A considers only high-prior candidates, showing the speed with which they are classified as Above-the-Bar (left) or Below-the-Bar (right).¹⁹ We observe that evaluators are quicker to classify advantaged candidates as Above-the-Bar, relative to candidates from the 60 – 60 pool (Panel A – left). Thus, not only are advantaged candidates more likely to be classified as Above-the-Bar (67% of high-prior candidates in the 60 – 40 pool compared to 61% of candidates in the 60 – 60 pool), they are also classified this way more quickly. Similarly, evaluators are slower to classify advantaged candidates as Below-the-Bar (Panel A – right).

Panel B presents this same analysis for low-prior candidates, revealing similar patterns. Evaluators require more signals before classifying a disadvantaged candidate as Above-the-Bar compared to a low-prior candidate from the 40 – 40 pool (left); they require fewer signals before classifying that disadvantaged candidate as Below-the-Bar (right). Disadvantaged candidates are more likely to be classified as Below-the-Bar relative to other low-prior candidates (60% versus 54%), and they are classified this way more quickly. Across candidate types, evaluators are quicker to classify candidates in line with group priors, particularly so when the groups are contrasting rather than equal.

Overall, evaluators are quite accurate in their final evaluations. Within the exogenous treatment, 84% of final evaluations are correct. Accuracy is significantly lower in the endogenous treatment at 76% ($p < 0.01$ when comparing across treatments).

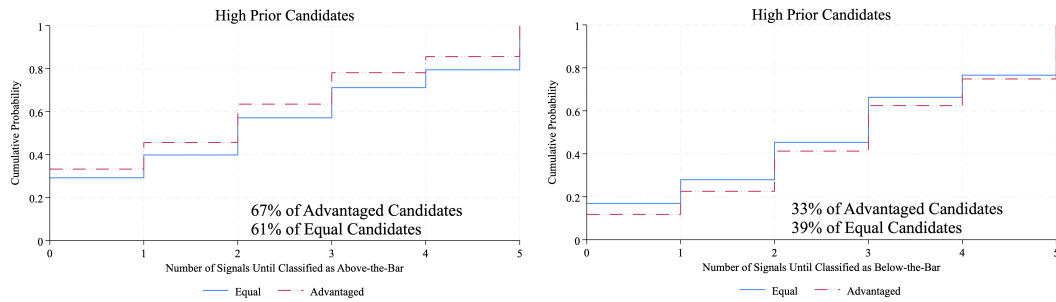
How do group differences impact the accuracy of final evaluations? On average, they have no significant impact: evaluators classify approximately the same proportion of candidates accurately independent of whether the groups are equal (40 – 40 or 60 – 60 pools) or contrasting (60 – 40 or 40 – 60 pools).²⁰ But, there is important heterogeneity. Group differences make it more likely that “typical” candidates are accurately classified. Above-the-Bar candidates from high-prior groups are more likely to be correctly classified as Above-the-Bar when they are advantaged (relative to the equal groups case); Below-the-Bar candidates from low-prior groups are more likely to be correctly classified as Below-the-Bar when they are disadvantaged (relative to the equal groups case). At

¹⁹Note that by speed here we mean the number of signals before a final evaluation is reached.

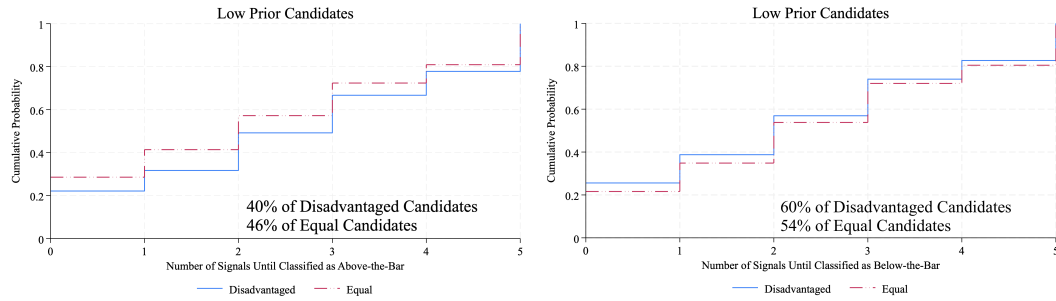
²⁰Consistent with the treatment differences described in the previous paragraph, this is approximately 84% of candidates in the exogenous treatment and 76% of candidates in the endogenous treatment.

Figure 3: Speed of Classifications

Panel A: High-Prior Candidates



Panel B: Low-Prior Candidates



Note: Panel A presents the distribution of the number of total signals viewed before a high-prior candidate is classified as Above-the-Bar (left) or Below-the-Bar (right). The percentages reported within the figure are the percentage of high-prior candidates of each type that are classified as Above-the-Bar (left) or Below-the-Bar (right). Panel B presents the same analysis for low-prior candidates.

the same time, group advantage makes it significantly less likely that an evaluator successfully classifies candidates that are not typical of their group. Figure 4 demonstrates this important pattern. We graph the proportion of correct evaluations for these atypical candidates, splitting the data by treatment assignment and candidate pool.

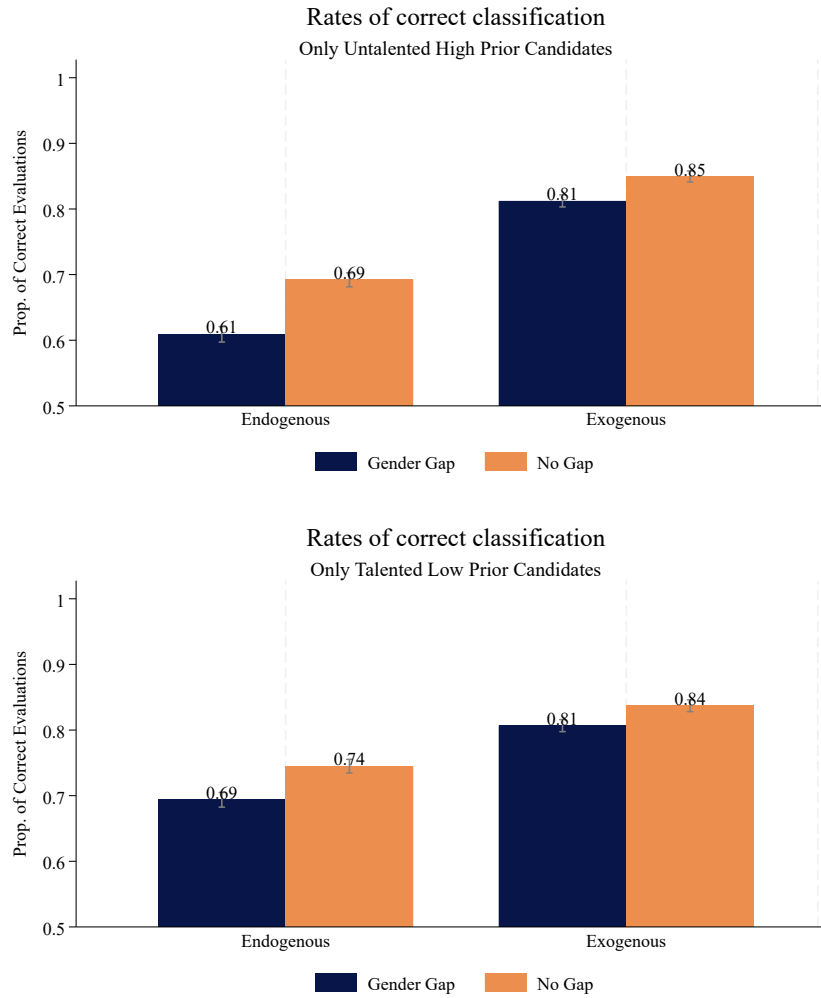
Not surprisingly, atypical candidates are on average harder to successfully classify, with overall success rates at 76% compared to 83% for candidates that are typical of their pool. The more important finding is that these mistakes are significantly more likely when the groups are different. In both treatments, contrast across the groups increases the share of incorrect classifications of atypical candidates. Consistent with the overall pattern, the rate of correct classifications of atypical candidates is lower in the endogenous treatment compared to the exogenous treatment. And, the impact of differences across groups on the likelihood of mistakes is nearly twice as large in the endogenous treatment. Within the exogenous treatment, atypical candidates are classified incorrectly approximately 3pp more often when there are differences across groups; in the endogenous treatment, this gap is 7pp.

The combination of requiring information acquisition and removing contrast between groups has a significant impact on the rate at which atypical candidates are classified correctly. In the exogenous 60–60 treatment, Below-the-Bar high-prior candidates are correctly classified 25pp more often than in the endogenous treatments with contrast across groups. Similarly, Above-the-Bar candidates from low-prior groups are 16pp more likely to be correctly identified when evaluators must receive information and there is no contrast between groups compared to the endogenous treatments with contrast across groups.

6 Discussion

We explore how evaluators choose and use information as they assess candidates. Overall, male and female candidates are assessed quite similarly by evaluators; we find no evidence that evaluators, in general, are biased against female candidates, either in terms of their initial assessments, how they update their beliefs, or in what final evaluations they submit. Instead, we find that relative advantage—the extent to which the candidate’s gender outperforms the other gender within the

Figure 4: Rate of Correct Classifications of Atypical Candidates



Note: The top graph shows the likelihood of correctly identifying a talented candidate from the low-prior group while the bottom graph shows the likelihood of correctly identifying an untalented candidate from the high-prior group. We show these likelihoods by candidates' relative advantages in their pools both in the exogenous and endogenous information acquisition treatments. Both types of mistakes are more common in the endogenous treatment and in particular when the high-prior candidate has a relative advantage and low-prior candidate has a relative disadvantage.

candidate pool—plays a significant role in decision-making. Conditional on a candidate’s likelihood of being Above-the-Bar, evaluators provide significantly more favorable absolute assessments of the candidate when that candidate has a relative advantage within the pool.

Relative advantage shapes initial assessments and how evaluators update their beliefs in response to new information. As a result, evaluators are quicker to classify high-prior candidates as Above-the-Bar when they are at a relative advantage, requiring fewer signals. They are similarly quicker to dismiss low-prior candidates as Below-the-Bar when they are at a relative disadvantage.

We find that relative advantage plays a larger role when information is more mixed. Evaluators are more likely to give the benefit of the doubt to high-prior candidates with early negative signals when they are from an advantaged group; low-prior candidates from the disadvantaged group benefit less from early positive signals.

Evaluators are significantly less likely to make accurate assessments when information must be endogenously acquired. We observe significant under-acquisition of information, despite the fact that information is cheap in our setting both in terms of monetary and time costs. This under-acquisition of information fuels stereotypical judgments: compared to the exogenous treatment, evaluators in the endogenous treatment are significantly less likely to accurately identify an Above-the-Bar candidate from a low-prior group, particularly when their group is disadvantaged. Similarly, evaluators in the endogenous treatment are significantly less likely to accurately identify a Below-the-Bar candidate from a high-prior group, particularly when their group is advantaged.

Although our experimental design involves only one time period, our results may have dynamic implications. Consider a context where men in the candidate pool are more likely to be qualified than women in the candidate pool, due to direct or systemic discrimination or other forces. In this world, an evaluator may hold an (accurate) prior like one in our study, that 40% of women are Above-the-Bar and 60% of men are Above-the-Bar. We can use the estimates from our setting to provide a rough approximation of how an evaluation process might unfold. If the evaluator’s prior on differences across groups biases decisions as in our experiment, endogenous information acquisition and biased processing could lead to women having approximately a 40% chance of being hired (classified as Above-the-Bar) and men having approximately a 65% chance of being hired (5pp higher than the prior).

If performance is noisy, it might be difficult to elicit the true type of the candidates even after the candidate is hired. In the next time period, the evaluator might think that the outcomes of the previous time period provide the least biased prior, i.e., that 40% of women are Above-the-Bar and 65% of men are Above-the-Bar. This could set into motion a dynamic process through which beliefs of group differences are further exaggerated over time, fueling disparities in evaluations. While of course this is only a speculative back-of-the-envelope calculation, it suggests a channel through which gender gaps could be perpetuated and amplified. Endogenous information acquisition can lead to judgments that reinforce, or even expand, baseline gender gaps.

Our findings suggest that individuals belonging to disadvantaged groups may find themselves facing additional hurdles in order to convey their capabilities or merit effectively. This is particularly true in environments where the process of acquiring information is endogenous. In such settings, individuals from disadvantaged backgrounds may need to exert significantly more effort to demonstrate their competence or qualifications convincingly. They must strive to send multiple positive signals in order to be *seen* or persuade evaluators which also requires perseverance and determination.

Making additional information on candidates available has been proposed as a way to reduce discrimination based upon biased beliefs (Bohren, Imas, et al., 2019). Our results suggest a potential obstacle for this path, particularly in the case where decision-makers must actively choose to receive information. Evaluators in our study significantly under-acquire information on individuals, despite the signals being readily available, cheap, and informative. Future work should explore the reasons why individuals do not take advantage of this information. Money and time costs are arguably trivial in our setting. Risk preferences also seem to be an unlikely explanation.²¹ It could be that evaluators aim to avoid the cognitive cost of incorporating new information into their beliefs and/or that they under-estimate the value of this additional information.

While we use the exogenous treatment as an important comparison group in our study, it also suggests a potentially promising intervention for remedying discrimination in these contexts.

Our results suggest that requiring evaluators to acquire a certain amount of information about all

²¹For instance, consider the decision to acquire the first signal. The evaluator is choosing between a lottery that pays \$7 with probability P and \$0.5 with probability $(1 - P)$, and a lottery that pays \$6.95 with probability $P + Y$ and \$0.45 with probability $(1 - P - Y)$. It is not clear that a risk averse decision-maker who properly values the informational content of the signal (the gain in probability Y) would prefer the second lottery to the first.

candidates could reduce bias and increase the accuracy of judgments. While the practicality of this intervention might vary across settings of interest, future research should consider how reducing evaluator discretion around information acquisition impacts discrimination in the field. On a similar note, our results suggest that interventions focusing on reducing the cost of information might not be as effective.

References

- Alesina, Alberto, Armando Miano, and Stefanie Stantcheva (2023). “Immigration and Redistribution”. In: *The Review of Economic Studies* 90.1, pp. 1–39.
- Bartoš, Vojtěch et al. (2016). “Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition”. In: *American Economic Review* 106.6, pp. 1437–1475.
- Bertrand, Marianne and Esther Duflo (2017). “Field Experiments on Discrimination”. In: *Handbook of Economic Field Experiments* 1, pp. 309–393.
- Bertrand, Marianne and Sendhil Mullainathan (2004). “Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination”. In: *American Economic Review* 94.4, pp. 991–1013.
- Bohren, J Aislinn, Kareem Haggag, et al. (2019). *Inaccurate Statistical Discrimination: An Identification Problem*. Tech. rep. National Bureau of Economic Research.
- Bohren, J Aislinn, Peter Hull, and Alex Imas (2022). *Systemic Discrimination: Theory and Measurement*. Tech. rep. National Bureau of Economic Research.
- Bohren, J Aislinn, Alex Imas, and Michael Rosenberg (Oct. 2019). “The Dynamics of Discrimination: Theory and Evidence”. In: *American Economic Review* 109.10, pp. 3395–3436. DOI: 10.1257/aer.20171829. URL: <https://www.aeaweb.org/articles?id=10.1257/aer.20171829>.
- Bordalo, Pedro, Katherine Coffman, et al. (July 2016). “Stereotypes”. In: *The Quarterly Journal of Economics* 131.4, pp. 1753–1794. ISSN: 0033-5533. DOI: 10.1093/qje/qjw029. eprint: <https://academic.oup.com/qje/article-pdf/131/4/1753/30636967/qjw029.pdf>. URL: <https://doi.org/10.1093/qje/qjw029>.
- (Mar. 2019). “Beliefs about Gender”. In: *American Economic Review* 109.3, pp. 739–73. DOI: 10.1257/aer.20170007. URL: <https://www.aeaweb.org/articles?id=10.1257/aer.20170007>.
- Bordalo, Pedro, John Conlon, et al. (2023a). *How People Use Statistics*. Tech. rep. National Bureau of Economic Research.
- (2023b). “Memory and Probability”. In: *The Quarterly Journal of Economics* 138.1, pp. 265–311.

- Campos-Mercade, Pol and Friederike Mengel (2023). “Non-Bayesian Statistical Discrimination”. In: *Management Science*.
- Carlana, Michela (2019). “Implicit Stereotypes: Evidence from Teachers’ Gender Bias”. In: *The Quarterly Journal of Economics* 134.3, pp. 1163–1224.
- Cavounidis, Costas, Kevin Lang, and Russell Weinstein (Sept. 2023). “The Boss is Watching: How Monitoring Decisions Hurt Black Workers”. In: *The Economic Journal* 134.658, pp. 485–514. ISSN: 0013-0133. DOI: 10.1093/ej/uead079. eprint: <https://academic.oup.com/ej/article-pdf/134/658/485/56444311/uead079.pdf>. URL: <https://doi.org/10.1093/ej/uead079>.
- Coffman, Katherine (2014). “Evidence on Self-Stereotyping and the Contribution of Ideas”. In: *The Quarterly Journal of Economics* 129.4, pp. 1625–1660.
- Coffman, Katherine, Manuela Collis, and Leena Kulkarni (2023). “Stereotypes and Belief Updating”. In: *Journal of the European Economic Association* Forthcoming.
- Coffman, Katherine, Christine Exley, and Muriel Niederle (2021). “The Role of Beliefs in Driving Gender Discrimination”. In: *Management Science* 67.6, pp. 3551–3569. DOI: 10.1287/mnsc.2020.3660. eprint: <https://doi.org/10.1287/mnsc.2020.3660>. URL: <https://doi.org/10.1287/mnsc.2020.3660>.
- Coffman, Katherine, Maria Paola Ugalde Araya, and Basit Zafar (2024). “A Dynamic Investigation of Stereotypes, Belief Updating, and Behavior”. In: *Economic Inquiry*, forthcoming.
- Egan, Mark, Gregor Matvos, and Amit Seru (2022). “When Harry Fired Sally: The Double Standard in Punishing Misconduct”. In: *Journal of Political Economy* 130.5, pp. 1184–1248.
- Esponda, Ignacio, Ryan Oprea, and Sevgi Yuksel (2023). “Seeing what is Representative”. In: *The Quarterly Journal of Economics* 138.4, pp. 2607–2657. URL: <https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjad020/7181329>.
- Eyting, Markus (2023). “Why do we Discriminate? The Role of Motivated Reasoning”. In: *Working Paper*.
- Fonseca, Miguel and Ashley McCrea (2023). “The Role of Shortlisting in Shifting Gender Beliefs on Performance: Experimental Evidence”. In: *Working Paper*.
- Gupta, Neeraja (2023). “Essays in Behavioral and Experimental Economics”. PhD thesis. University of Pittsburgh.

- Hebert, Camille (2020). “Gender Stereotypes and Entrepreneur Financing”. In: *10th Miami Behavioral Finance Conference*.
- Komiyama, Junpei and Shunya Noda (2020). “On Statistical Discrimination as a Failure of Social Learning: A Multi-Armed Bandit Approach”. In: *arXiv preprint arXiv:2010.01079*.
- Lepage, Louis Pierre (2021). “Endogenous Learning, Persistent Employer Biases, and Discrimination”. In: *Persistent Employer Biases, and Discrimination (March 2, 2021)*.
- Möbius, Markus M. et al. (2022). “Managing Self-Confidence: Theory and Experimental Evidence”. In: *Management Science* 68.11, pp. 7793–7817.
- Rabin, Matthew and Joel L. Schrag (1999). “First Impressions Matter: A Model of Confirmatory Bias”. In: *The Quarterly Journal of Economics* 114.1, pp. 37–82.
- Rackstraw, Emma (2022). “Bias-Motivated Updating in the Labor Market”. In: *Available at SSRN 4278076*.
- Ruzzier, Christian A. and Marcelo D. Woo (2023). “Discrimination with Inaccurate Beliefs and Confirmation Bias”. In: *Journal of Economic Behavior & Organization* 210, pp. 379–390. ISSN: 0167-2681. DOI: <https://doi.org/10.1016/j.jebo.2023.04.018>. URL: <https://www.sciencedirect.com/science/article/pii/S0167268123001221>.
- Sarsons, Heather (2017). “Interpreting Signals in the Labor Market: Evidence from Medical Referrals”. In: *Job Market Paper*, pp. 141–145.
- Schwartzstein, Joshua and Adi Sunderam (2021). “Using Models to Persuade”. In: *American Economic Review* 111.1, pp. 276–323.

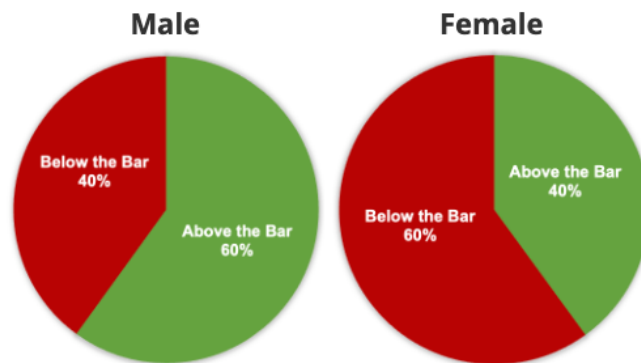
A Appendix

Figure A.1: Example of Pie Charts Used to Provide Pool Information

We asked each of the 50 candidates, "What is your sex?"



Here is the likelihood of the candidates being Above-the-Bar according to their answer to this question:



Note: Evaluator is assigned to a pool and then sees full info about the pool through a series of pie charts and a table indicating proportion of candidates Above-the-Bar by each answer. These figures are illustrating example pie charts for the 60 – 40 pool.

Figure A.2: Example of Summary Table Used to Provide Pool Information

Percent of Each Group Above-the-Bar

Male: 40%	Female: 40%
Prefers a Beach Vacation: 41%	Prefers a Mountain Vacation: 39%
Prefers Light Colors: 39%	Prefers Dark Colors: 41%
Prefers Spring: 39%	Prefers Fall: 41%
Prefers Lions: 41%	Prefers Tigers: 39%

This information also will appear on each page in the Instructions Summary.






Please click the arrow to begin evaluating candidates.

Note: Evaluator is assigned to a pool and then sees full info about the pool with a series of pie charts and a table indicating proportion of candidates Above-the-Bar by each answer. This figure illustrates an example summary table for the 40 – 40 pool.

Figure A.3: Example Candidate Profile

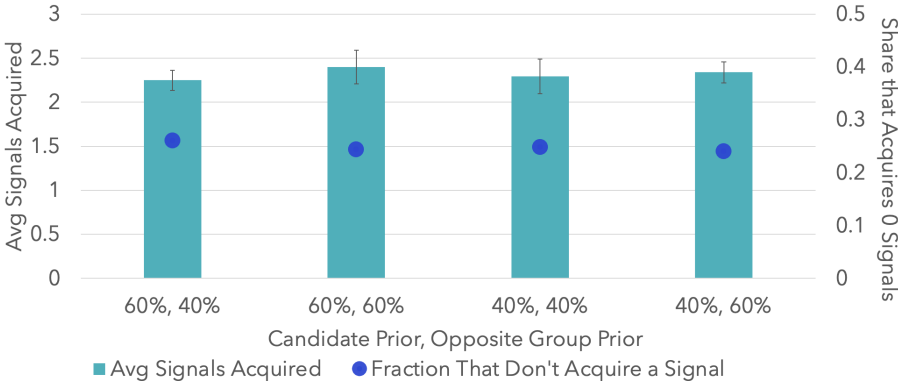
Candidate 1 out of 5

Below are Candidate 1's characteristics:

Prefers Lions 	<input type="radio"/>
Prefers Light Colors 	<input type="radio"/>
Prefers Spring 	<input type="radio"/>
Female 	<input type="radio"/>
Prefers A Beach Vacation 	<input type="radio"/>

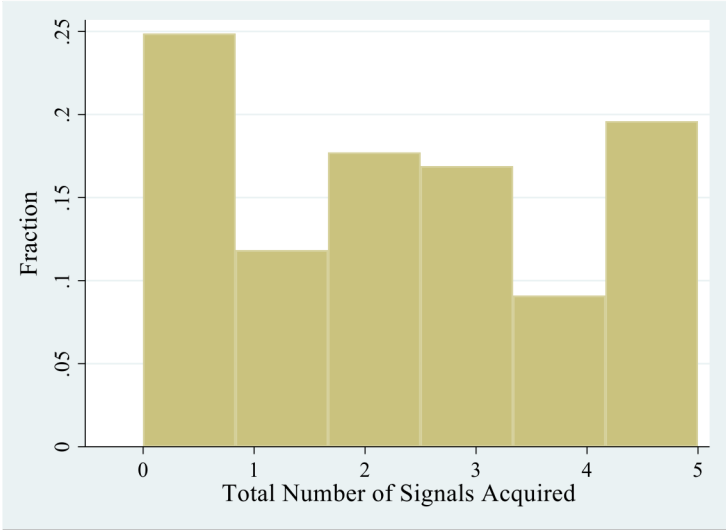
Note: This is a screenshot of an example candidate profile.

Figure A.4: Average number of signals acquired by pool – Endogenous Treatment Only



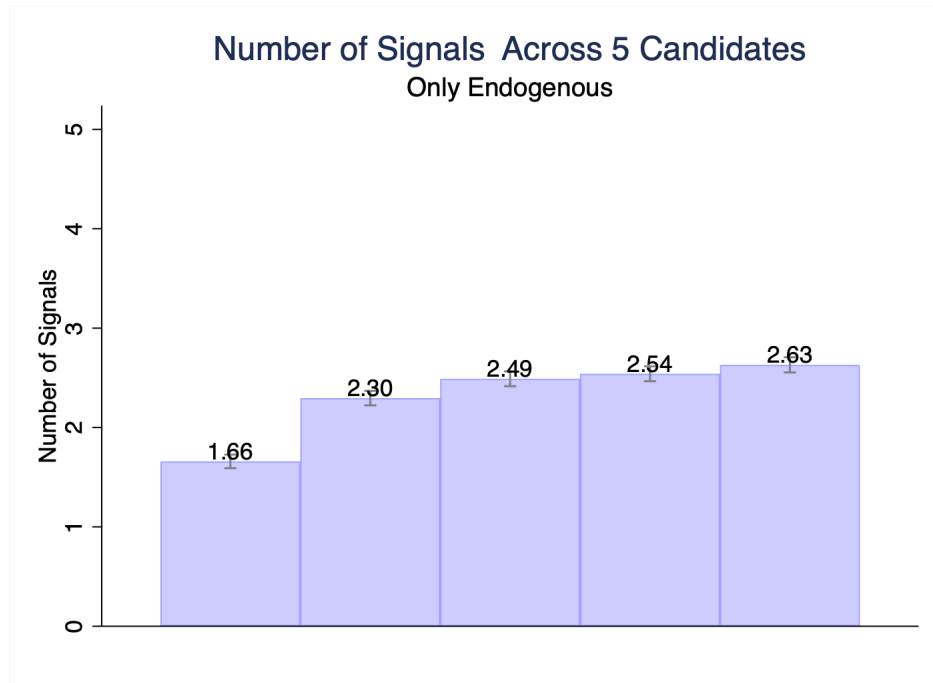
Note: This figure shows the average number of signals acquired by evaluators in the endogenous information acquisition treatment by the candidate pools, as well as the share of candidates for which an evaluator chose to receive 0 signals.

Figure A.5: Distribution of number of signals acquired by pool – Endogenous Treatment Only



Note: This is a histogram of the total number of signals acquired for each candidate-evaluator pair in the endogenous treatment.

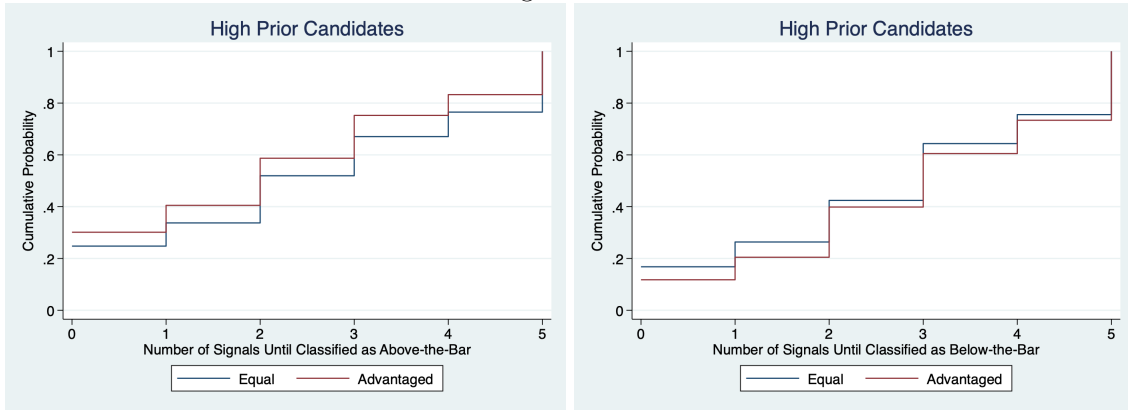
Figure A.6: Number of Signals Bought Across Candidates – Endogenous Treatment Only



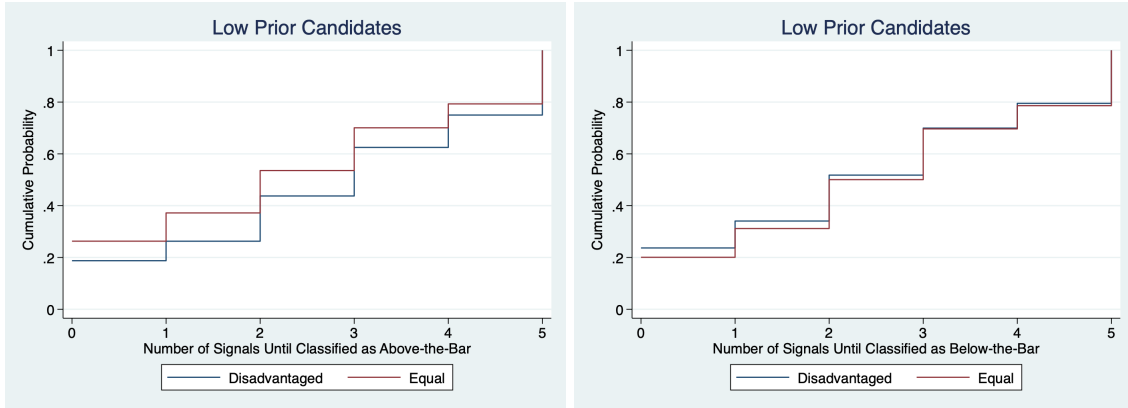
Note: This figure shows the average number of signals bought by the evaluators for each of the 5 candidates they evaluate by candidate order.

Figure A.7: Speed of Classifications: Robustness Excluding the First Round

Panel A: High-Prior Candidates

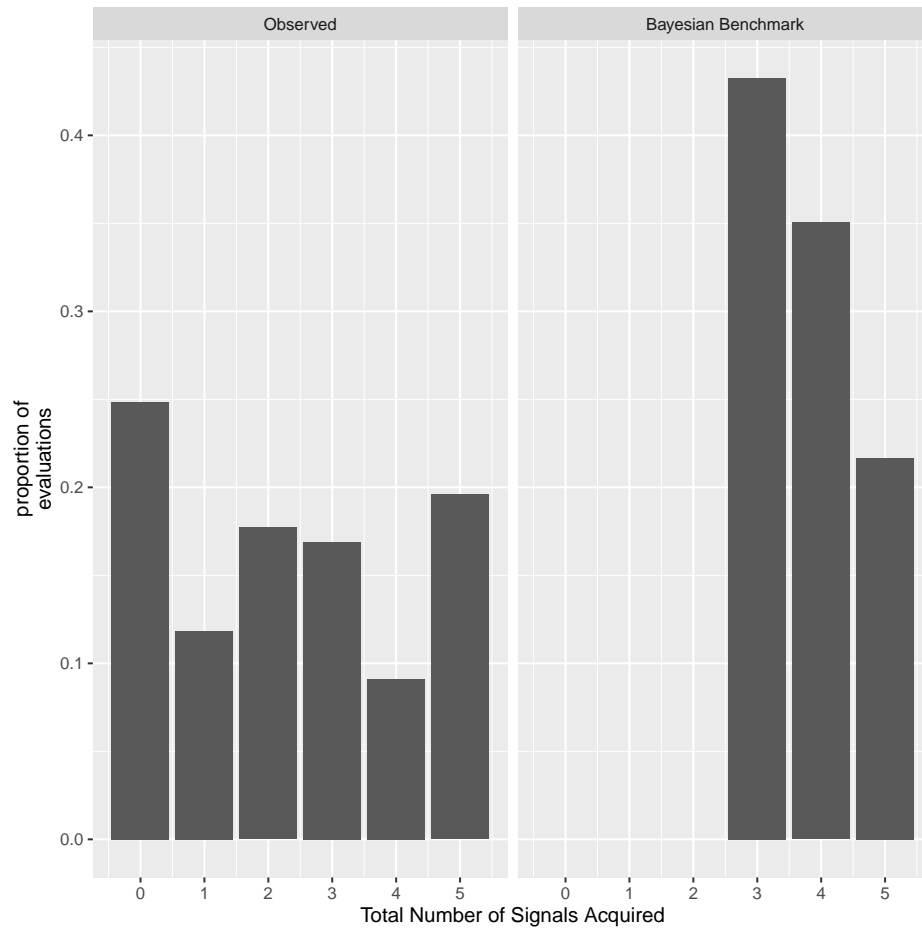


Panel B: Low-Prior Candidates



Note: This figure replicates Figure 3 dropping observations from the first candidate for each evaluator. Panel A presents the distribution of the number of total signals viewed before a high-prior candidate is classified as Above-the-Bar (left) or Below-the-Bar (right). Panel B presents the same analysis for low-prior candidates. When we include the round 1 in Figure 3, 61% of equal and 67% advantaged candidates were classified as above the bar among the high prior candidates. When we drop round 1 these rates are 60% and 65% respectively. For low prior candidates, including the round 1, 46% equal and 40% disadvantaged candidates are classified as Above-the-Bar. When we drop round 1, these rates remain mostly unchanged.

Figure A.8: Under-Acquisition of Information Relative to a Bayesian Benchmark



Note: The first panel displays the frequency of acquiring 0 - 5 signals in our endogenous treatment. The second panel displays the frequency with which a Bayesian decision-maker would acquire 0 - 5 signals in those same instances.

Table A.1: Updating After Signals by Splitting by Group Prior – OLS Predicting Beliefs After...

Panel A: Only Low-Prior Sample:						
	No Signals	1 Positive	1 Negative	2 Positives	2 Mixed	2 Negatives
Relative Advantage	0.150*** (0.0274)	0.107*** (0.0266)	-0.0229 (0.0266)	0.0880* (0.0487)	-0.00160 (0.0252)	-0.0835** (0.0399)
Male Candidate	0.679 (0.448)	0.231 (0.457)	-0.220 (0.472)	-0.523 (0.802)	0.247 (0.458)	-0.345 (0.708)
Exogenous	0.0637 (0.544)	-2.160*** (0.519)	1.930*** (0.520)	-6.547*** (0.914)	0.754 (0.467)	5.157*** (0.815)
Initial Likelihood		0.692*** (0.0200)	0.599*** (0.0284)	0.419*** (0.0319)	0.643*** (0.0246)	0.457*** (0.0332)
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes	Yes	Yes
No. Clusters	7882	3131	3794	1705	2479	2257
No. Observations	2326	1706	1871	1192	1497	1439
Adj. R2	0.0483	0.507	0.385	0.211	0.483	0.214
Panel B: Only High-Prior Sample:						
	No Signals	1 Positive	1 Negative	2 Positives	2 Mixed	2 Negatives
Relative Advantage	0.162*** (0.0237)	0.0181 (0.0194)	0.0756** (0.0296)	0.0393 (0.0320)	0.0801*** (0.0246)	0.123** (0.0498)
Male Candidate	-0.648 (0.401)	-0.0633 (0.339)	0.113 (0.517)	-0.0135 (0.562)	-0.00698 (0.448)	0.743 (0.873)
Exogenous	0.406 (0.476)	-2.722*** (0.395)	2.443*** (0.571)	-6.252*** (0.632)	-0.313 (0.487)	6.937*** (0.994)
Initial Likelihood		0.695*** (0.0202)	0.654*** (0.0282)	0.517*** (0.0294)	0.670*** (0.0240)	0.443*** (0.0383)
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes	Yes	Yes
No. Observations	8118	3905	3188	2408	2476	1750
No. Clusters	2381	1907	1745	1455	1537	1213
Adj. R2	0.0641	0.569	0.381	0.350	0.480	0.165

Notes: Standard errors are reported in parentheses and clustered at the evaluator level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table repeats the analysis in the main text Table 5 by including both exogenous and endogenous information acquisition groups. The dependent variable in the first column is the likelihood of the candidate being Above-the-Bar chosen by the evaluator in the first slider before receiving any signals. The following columns report the regressions on the likelihoods of being Above-the-Bar chosen on the sliders after receiving the corresponding first two signals in each column for the corresponding subsample. All regressions include round fixed effects and candidate advantage by filler characteristics.

Table A.2: The Impact of Relative Advantage on Assessments – Robustness Excluding the First Round

OLS Predicting Initial Beliefs, Final Beliefs, and Final Evaluations						
	Initial Slider		Final Slider		Final Eval	
	(1)	(2)	(3)	(4)	(5)	(6)
Relative Advantage	0.145*** (0.0183)	0.170*** (0.0215)	0.0775** (0.0302)	0.0912** (0.0355)	0.160*** (0.0446)	0.255*** (0.0550)
Male Candidate	-0.116 (0.313)	0.191 (0.447)	0.409 (0.590)	0.359 (0.786)	0.504 (0.879)	0.307 (1.263)
Group Prior	0.426*** (0.0289)	0.426*** (0.0289)	0.533*** (0.0439)	0.533*** (0.0439)	0.712*** (0.0633)	0.712*** (0.0633)
Exogenous	0.265 (0.377)	0.566 (0.484)	-2.338*** (0.598)	-2.390*** (0.846)	-3.006*** (0.864)	-3.220*** (1.221)
Rel Adv x Exog		-0.0478** (0.0228)		-0.0271 (0.0411)		-0.186*** (0.0625)
Male x Exog		-0.603 (0.626)		0.105 (1.177)		0.428 (1.751)
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes	Yes	Yes
No. Observations	12800	12800	12800	12800	12800	12800
No. Clusters	3200	3200	3200	3200	3200	3200
Adj. R2	0.152	0.152	0.0387	0.0386	0.0361	0.0367

Notes: This table tests the robustness of the Table 2 to the exclusion of the first round. Standard errors are reported in parentheses and clustered at the evaluator level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In the first two columns, the dependent variable is the likelihood of the candidate being Above-the-Bar chosen by the evaluator in the first slider. Columns 3 and 4 report the regressions on the likelihood of the candidate being Above-the-Bar chosen by the evaluator in the last slider before making their final decision. The last two columns report the regressions of the indicator variable taking value 1, if the final evaluation of the candidate was Above-the-Bar, and 0 otherwise. All regressions include round fixed effects and candidate advantage by filler characteristics.

Table A.3: Selection Into Acquiring Information – Robustness Excluding the First Round
 OLS Predicting Initial Likelihoods and Final Evaluations by Subsample

	Immediate Evaluation	Got More Info	Pooled	
	Initial Slider	Initial Slider	Initial Slider	Final Eval
Relative Advantage	0.29*** (0.064)	0.14*** (0.026)	0.25*** (0.049)	0.81*** (0.098)
Male Candidate	0.26 (1.22)	0.13 (0.45)	0.30 (1.21)	-2.71 (2.54)
Group Prior	0.31*** (0.093)	0.43*** (0.042)	0.40*** (0.039)	0.71*** (0.089)
Got More Info			-3.58*** (0.95)	-10.7*** (2.01)
Got More Info x Rel Adv			-0.090* (0.047)	-0.72*** (0.10)
Got More Info x Male			-0.18 (1.29)	3.59 (2.93)
Round FE	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes
No. Observations	1432	4944	6376	6376
No. Clusters	562	1405	1594	1594
Adj. R2	0.14	0.17	0.16	0.058

Notes: This table tests the robustness of the Table 3 to the exclusion of the first round. Standard errors are reported in parentheses and clustered at the evaluator level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable in all columns is the likelihood of the candidate being Above-the-Bar chosen by the evaluator in the first slider before receiving any signals. The first column include the sample of observations where the evaluation was made without acquiring any information. The second column includes the observations where evaluation was made after receiving at least one signal. The third and fourth columns pool all observations and estimates the effect of receiving any signals on the initial likelihood assessment and final evaluation, respectively. All regressions includes controls for round fixed effects and candidate advantage by filler characteristics. All three columns include endogenous treatment only.

Table A.4: Endogenous Information Acquisition – Robustness Excluding the First Round

OLS Predicting the Decision to Get More Info After...

Panel A: Overall					
	1st Signal	2nd Signal	3rd Signal	4th Signal	Pooled-Last Signal
Confirming Signal	-0.025*** (0.0095)	-0.076*** (0.013)	-0.0099 (0.017)	0.0033 (0.020)	-0.022*** (0.0057)
Round FE	Yes	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes	Yes
No. Observations	4944	4319	3183	2029	15871
No. Clusters	1405	1280	1119	809	1405
Adj. R2	0.010	0.0097	0.0016	-0.0022	0.0010

Panel B: Comparing Pools with Equal Versus Contrasting Pools					
	1st Signal	2nd Signal	3rd Signal	4th Signal	Pooled-Last Signal
Confirming Signal	-0.048*** (0.013)	-0.090*** (0.020)	-0.024 (0.025)	-0.021 (0.030)	-0.034*** (0.0082)
Equal Groups	-0.032** (0.016)	-0.0062 (0.021)	-0.0044 (0.029)	0.00045 (0.034)	-0.010 (0.0082)
Equal Groups x Conf. Signal	0.045** (0.019)	0.026 (0.027)	0.028 (0.034)	0.045 (0.041)	0.023** (0.011)
Round FE	Yes	Yes	Yes	Yes	Yes
Advantage by Fillers	Yes	Yes	Yes	Yes	Yes
No. Observations	4944	4319	3183	2029	15871
No. Clusters	1405	1280	1119	809	1405
Adj. R2	0.011	0.0090	0.0012	-0.0014	0.00095

Notes: This table tests the robustness of the Table 4 to the exclusion of the first round. Standard errors are reported in parentheses and clustered at the evaluator level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions includes controls for relative advantage, male candidate indicator, group prior, initial likelihood assessment, round fixed effects, and candidate advantage by filler characteristics. All columns include endogenous treatment only. The dependent variable is an indicator variable for deciding to receive a signal after the corresponding signal in each column for the corresponding subsample. The last column pools all observations in each round in the long format.