Public perceptions of discrimination at large firms

Patrick Kline, Evan K. Rose, and Christopher R. Walters*

March 27, 2026

Abstract

We analyze an online survey eliciting beliefs about discrimination and other aspects of recruiting conduct by large firms. One-third of respondents report having put off applying for a job due to fear of discrimination. We use respondents' comparisons of specific firms to construct measures of perceived discrimination, hiring discretion, and selectivity for each firm. Beliefs about the likely treatment of applicants with race and sex distinctive names correlate strongly with beliefs about broader notions of discrimination. Firms thought to prefer white job applicants are also viewed as favoring older workers, while perceptions of race and gender discrimination are unrelated. We compare these beliefs to experimental estimates of discrimination derived from a large resume correspondence study. Perceived gender discrimination strongly correlates with experimental estimates of gender contact gaps, while beliefs about racial discrimination are weakly related to experimental estimates of racial contact gaps. Respondents believe that selective firms discriminate more often and that managers have less hiring discretion at such firms. In contrast, the correspondence study evidence suggests that discretion is a positive predictor of racial discrimination and that much discrimination occurs at less selective firms.

Keywords: Discrimination, Beliefs, Ranking Methods

JEL Codes: J71, C11, C13

*Kline: UC Berkeley and NBER, pkline@econ.berkeley.edu; Rose: University of Chicago and NBER, ekrose@uchicago.edu; Walters: University of Chicago and NBER, crwalters@uchicago.edu. This project was approved by the Berkeley Committee for the Protection of Human Subjects. CPHS Protocol Number: 2021-09-14653.

1 Introduction

Employment discrimination by large employers has recently been the subject of detailed empirical study (Miller, 2017; Hoffman, Kahn and Li, 2018; Kline, Rose and Walters, 2022; Mocanu, 2022; Kline, Rose and Walters, 2024). This line of work finds that race and sex discrimination varies substantially among firms, with smaller biases among federal contractors, more discrimination in customer-facing sectors, and more discrimination in organizations that grant managers greater discretion in hiring. While recent survey evidence clarifies the role that race plays in shaping attitudes towards public policy (Alesina, Ferroni and Stantcheva, 2021; Haaland and Roth, 2023), less is known about what the public believes regarding discrimination by specific employers.

Are the employer biases documented in recent empirical work an open secret? Firms have strong legal and economic incentives to conceal any biases in their hiring processes. Yet workers invest considerable effort in learning about prospective employers, and large firms develop reputations—through social networks and, increasingly, online platforms—regarding their recruiting practices. Becker (1957) argued that the deleterious effects of discrimination on employment would be mitigated if employer biases were concentrated in a small share of firms, since minority-group workers could still find employment among other unprejudiced employers. However, if job seekers are unaware of which employers are biased, they may find it difficult to avoid discrimination, which can give rise to excess unemployment and lower wages (Black, 1995; Bowlus and Eckstein, 2002).

This paper examines public perceptions of hiring practices at 164 large U.S. employers using a survey of nearly 7,000 respondents. The survey oversamples job seekers and racial minorities to assess whether groups more likely to experience discrimination hold systematically different beliefs. Respondents are asked to rate a randomly selected set of five firms regarding their treatment of job applicants based on race, sex, and age. The firms studied are all members of the Fortune 500 and form a superset of those analyzed in a large resume correspondence experiment by Kline, Rose and Walters (2022).

To understand views on multiple forms of discrimination, respondents are randomly assigned to one of two survey arms, each with a distinct set of questions. The *names arm* asks how firms are likely to treat entry-level job applicants with different sorts of distinctive names. These questions can be viewed as eliciting beliefs about the likely effects of a correspondence experiment conveying race or gender through names, although such experiments are never mentioned in the survey. A second *conduct arm* directly elicits the likelihood that each firm will "discriminate against" a particular group, which captures views on discrimination beyond differential treatment of distinctive names.

Survey responses establish that the public perceives labor market discrimination to be pervasive. Roughly a third of respondents state that they have put off applying for a job out of fear of discrimination. This share rises to nearly 45% for Black respondents but exceeds 20% even for White respondents. These responses suggest discrimination looms large in the minds of job applicants of all racial groups.

We use two different approaches to aggregating perceptions of firm-level discrimination across respondents. The first approach simply averages beliefs about each firm coded on a 1 to 5 scale, resulting in a cardinal measure of discrimination perceptions. Our second approach converts cardinal responses into rank-ordered comparisons and aggregates them using a variant of the classic Borda (1784) score. This ordinal approach eliminates differences across respondents in the location and scale of responses, and has desirable properties for aggregating rankings in settings where the Independence of Irrelevant Alternatives (IIA) assumption is violated (Korba, Clémençon and Sibony, 2017; Shah and Wainwright, 2018). The cost of this approach is that only beliefs about relative conduct of firms are preserved. Reassuringly, however, we find that key results differ little across the cardinal and rank-ordered methods.

In both approaches, more widely dispersed firm-specific parameters indicate more agreement about differences in perceived conduct across firms. Accounting for sampling error, we find more dispersion for firm effects in perceived gender bias than corresponding estimates of race-related or age-related bias. This implies greater agreement about which

firms discriminate on the basis of gender than on the basis of race or age. Respondents also largely agree about which firms are more selective in hiring and which employers grant more discretion to hiring managers.

Next, we document three facts about the correlation structure of beliefs across several dimensions of firm conduct. First, beliefs about the treatment of applicants with racially distinctive names correlate strongly with beliefs about broader notions of racial discrimination. In other words, respondents tend to rank the racial preferences of firms similarly regardless of how questions about racial bias are posed. Second, firms that are thought to prefer white job applicants tend to also be seen as favoring older workers. However, beliefs about firm preferences over race are uncorrelated with beliefs about firms' gender preferences. Third, beliefs about which firms have strong racial, gender, or age preferences are highly correlated across respondents of different demographic groups. Although Black respondents have somewhat stronger beliefs about racial discrimination, their beliefs about which firms discriminate differ little from the beliefs of white respondents.

We then examine the extent to which these firm-level beliefs predict the experimental contact gaps in the large-scale correspondence experiment of Kline, Rose and Walters (2022). We find a statistically significant relationship between experimental race contact gaps and within-industry variation in perceived racial discrimination. However, between-industry variation in beliefs is not a significant predictor of racial contact gaps. Perceived gender discrimination strongly correlates with firm-specific gender contact gaps, both within and between industries. The link between experimental and survey-based measures of racial discrimination is stronger among current job seekers and workers who fear discrimination.

These findings on perceptions of discrimination suggest three key conclusions. First, gender discrimination in entry-level hiring appears to be an open secret: survey respondents largely agree on which firms discriminate by gender, and these views correlate strongly with experimental gender gaps. A plausible explanation for this state of affairs is that certain forms of gender discrimination, though illegal, are viewed as socially acceptable,

e.g., preferences for female employees at women's apparel stores. Awareness of such patterns may aid job-seekers by allowing them to target employers who will not penalize them based on gender, as in Becker's original theory.

Second, beliefs about racial discrimination are diffuse and only weakly related to experimental evidence. One interpretation of this finding is that beliefs about racial discrimination are often mistaken. In addition to generating potentially sizable psychic costs, mistaken beliefs can generate misallocation, a possibility bolstered by our finding that a large share of respondents report having put off applying to jobs out of fear of discrimination. Misallocation of this nature is an important feature of search models of discrimination with heterogeneous employer conduct (e.g., Black, 1995; Bowlus and Eckstein, 2002). In such models, uncertainty and mistaken beliefs about *which* employers discriminate can generate excess unemployment and equilibrium wage penalties, as job-seekers devote time and effort applying to discriminatory firms before finding an unbiased employer.

Third, the positive correlation between perceptions of firm selectivity and racial bias suggests the public believes discrimination takes place behind closed doors at elite institutions. Here, public beliefs depart sharply from the related academic evidence, which finds more discrimination at less profitable employers offering less attractive jobs as well as more bias among employers with higher levels of managerial hiring discretion. Whether this departure reflects mistaken beliefs among the public or incompleteness in academic attempts to measure discriminatory conduct is an interesting question for future research.

The rest of the paper is organized as follows. The next section describes the survey design and provides some descriptive statistics on the respondents. Section 3 introduces the two methods of aggregating beliefs. Section 4 uses these methods to summarize public views on discrimination.

2 Data

Our data come from a Qualtrics survey conducted between June and August of 2023. The survey asks each respondent a battery of questions about their perceptions of recruiting practices at five firms randomly selected from a list of 164 Fortune 500 companies. An example of the survey questions appears in Appendix A.

2.1 Survey sample

The survey was designed to oversample Black respondents and current job seekers. Qualtrics sampled respondents via two methods: probability sampling and convenience sampling. The recruited *probability sample* is a nationally representative panel of survey respondents who agree to fill out whatever survey they are assigned by Qualtrics. The remaining respondents are a *convenience sample* that specifically opted in to take our survey.

Both samples were screened to meet certain demographic guidelines. In both samples, we sought a 50/50 mix of Black and white respondents. Since Qualtrics' panel of probability sample respondents is nationally representative, we quickly exhausted their recruited pool of Black respondents. This required us to rely heavily on convenience sampling to meet our target sample sizes of 6,000 respondents. In the convenience sample, we additionally sought for 1/6th of the respondents to have a bachelor's degree or higher credential and roughly 50% to be actively looking for work, although we did not meet this latter target.

Table 1 shows some descriptive statistics on the survey analysis sample. This sample has already been screened for certain quality checks. First, Qualtrics removed some respondents as part of their usual screening procedures designed to weed out non-human respondents. Non-human respondents were primarily a concern for the convenience sample, as the probability sample was comprised of experienced survey takers for whom Qualtrics had a longer run record of responses. Next, we dropped respondents who failed an attention check that required them to recall the names of the companies they were

asked about in the survey. We also remove respondents who straight line answers to all the questions about firms described in the next section.

The total sample size is 6,515 respondents, 4,139 of whom were recruited via convenience sampling. While race and sex are roughly balanced in the recruited sample, the convenience sample skews heavily female. Overall, roughly two thirds of respondents are women. Nearly half of the overall sample self-identifies as Black and roughly 44% of the sample reports an age younger than 40. Most respondents who self-identified as neither Black nor White report membership in multiple racial groups.

While the modal respondent had some college, only about 29% had a bachelor's degree or higher. Roughly half of respondents report working as paid employees, with many listing their status as either looking for work or retired. Thirty-four percent of respondents report annual income above $60,000. For comparison, in the 2025 CPS-ASEC, the share was also 34% (U.S. Census Bureau, 2025).

2.2 Survey design

The survey consists of four main blocks: an introductory block on beliefs about company selectivity and desirability, two blocks on beliefs about discrimination, and a final concluding block. All respondents are shown the first and final block, while each respondent is shown either the second block or the third block, assigned randomly with equal probability.

Block zero: Respondent characteristics [all respondents]. Before beginning the survey, all respondents provide information on their demographic characteristics, employment status, and income.

Block one: Survey introduction [all respondents]. The survey begins with a vignette about hiring at entry-level jobs. All respondents are asked the following question.

> In this survey, we're interested in understanding people's perceptions of entry-

Table 1: Descriptive statistics

	All		Probability		Convenience	
	Count	Share	Count	Share	Count	Share
Gender						
Female	4406	0.676	1317	0.554	3089	0.746
Male	2109	0.324	1059	0.446	1050	0.254
Race						
Black	3217	0.494	1068	0.449	2149	0.519
White	2714	0.417	1069	0.45	1645	0.397
Other or Mixed Race	584	0.09	239	0.101	345	0.083
Hispanic						
Hispanic	694	0.107	275	0.116	419	0.101
Not Hispanic	5821	0.893	2101	0.884	3720	0.899
Age						
[18,25)	832	0.128	177	0.074	655	0.158
[25,40)	2023	0.311	734	0.309	1289	0.311
[40,65)	2558	0.393	959	0.404	1599	0.386
65+	1102	0.169	506	0.213	596	0.144
Marital Status						
Married	2224	0.341	1008	0.424	1216	0.294
Never Married	2850	0.437	870	0.366	1980	0.478
Other	1441	0.221	498	0.21	943	0.228
Education						
No High School Diploma	273	0.042	62	0.026	211	0.051
High School Diploma	1620	0.249	415	0.175	1205	0.291
Some College / Associate Degree	2725	0.418	1008	0.424	1717	0.415
Bachelor's / Graduate Degree	1897	0.291	891	0.375	1006	0.243
Employment						
Working As Paid Employee	3233	0.496	1369	0.576	1864	0.45
Working Self-employed	578	0.089	168	0.071	410	0.099
Not Working: Looking For Work	795	0.122	143	0.06	652	0.158
Not Working: Retired	993	0.152	425	0.179	568	0.137
Not Working: Other	916	0.141	271	0.114	645	0.156
Income						
Less Than $5,000	431	0.066	67	0.028	364	0.088
$5,000 To $30,000	1784	0.274	467	0.197	1317	0.318
$30,000 To $60,000	2048	0.314	672	0.283	1376	0.332
$60,000 To $100,000	1297	0.199	588	0.247	709	0.171
More Than $100,000	955	0.147	582	0.245	373	0.09
N. Of Respondents	6515		2376		4139	

Notes: This table presents summary statistics for respondents in the full sample, the probability sample (drawn from Qualtrics' recruited panel of potential respondents), and the convenience sample (who voluntarily opt in to the survey). All respondents were asked about their characteristics at the beginning of the survey.

> level jobs. Some examples of entry-level jobs include cashier, sales associate, delivery driver, and customer service representative. Have you held an entry-level job in the last 2 years? *[Yes/No/Prefer not to answer]*

In addition to soliciting information about the respondent's background, this question is meant to squarely frame the idea of an entry-level job in the respondent's mind.

We then ask a series of questions on a five point Likert (1932) scale about a randomly selected set of five firms. This set is constructed by first drawing one of three well-known "anchor" firms—KFC, Home Depot, or Coca-Cola—with equal probability. The remaining four firms are chosen without replacement from a list of 161 firms that nests the set of 108 firms studied in Kline, Rose and Walters (2022).

A first set of questions probes the respondent's beliefs about the selectivity and desirability of entry-level jobs at these five firms:

> In this section of the survey, we'd like to understand your beliefs about the hiring practices of five U.S. companies. We will ask you a series of questions about the likelihood that individuals are contacted for a job interview, hired for a job, and accept a job at these companies.

Respondents are then shown the following question, which asks them to rate the likelihood of a high-school graduate applying to an entry-level job at each company successfully passing an interview, which is our measure of perceived *selectivity*.

> Consider a **high-school graduate** with 1-2 years of relevant experience who has applied to an entry-level job.
>
> Please indicate the likelihood that an applicant would be able to successfully **pass an interview** with each of the following companies.

Responses are elicited for each of the five companies on a 5-point scale ranging from "very unlikely" to "very likely." Respondents also have the option to select "Don't know / prefer not to answer" for each firm.

We then ask the likelihood that this high-school graduate would accept the job if offered it, which is our measure of perceived *desirability*.

> Please indicate the likelihood that this applicant would **accept an entry-level job** at each of these companies, if offered one.

The survey subsequently branches into two arms, which are randomly assigned with equal probability. The *names arm* asks a series of questions about the likely treatment of distinctive names at the 5 firms. The *conduct arm* asks about the treatment of race, sex, and age at these firms. In order to avoid priming, each respondent is exposed to only one arm.

Block 2: Names arm [50% of sample]. The names arms measures the relative likelihood that each firm would contact and hire applicants to an entry-level job with different racially and gender-distinctive names. It begins by providing the following context:

> Now suppose two equally qualified applicants apply to these same five companies. As before, the applicants are both **high-school graduates with 1-2 years of relevant experience**, but the **applications use different names**. Each name is either distinctively white or black and either distinctively male or female.
>
> Here are some examples of such names:
> Distinctively **black** names: Lawanda and Reginald
> Distinctively **white** names: Allison and Edgar
> Distinctively **female** names: Charlotte and Ebony
> Distinctively **male** names: Matthew and Tremayne

Respondents are asked about both race and gender in random order. The first question in the names arm measures the relative likelihood that each firm would *contact* an applicant to an entry-level job. For race, for example, the question reads:

> For each company, please indicate how much more likely you believe the company is to **contact a job applicant** with a distinctively [**black/white**] name

than an applicant with a distinctively [white/black] name.

Responses are elicited on a 5-point scale ranging from "much less likely" to "much more likely," along with a "Don't know / prefer not to answer" option. Since the question is predicated on the notion that the applicants are equally qualified, it arguably elicits what respondents believe would result from a correspondence experiment.

The second question measures the relatively likelihood each firm would *hire* an applicant. For race, for example, it reads:

> For each company, please indicate how much more likely you believe the company is to **hire a job applicant** with a distinctively **[black/white]** name than an applicant with a distinctively [white/black] name.

Responses are elicited using the same structure as the first question.

The questions for gender are analogous (see the Appendix for the exact text). To examine whether the order the groups are presented is important, respondents are asked about the relative chances of a Black vs. White applicant or White vs. Black applicant (and likewise a male vs. female or female vs. male) with equal probability.

Finally, for both race and gender, respondents are asked to rate their confidence in their responses to the pair of questions, with answers ranging from "not at all confident" to "extremely confident."

Block 3: Conduct arm [50% of sample]. Instead of the names arm, half of respondents were randomized into the conduct arm. In this block, we ask about discrimination more broadly and again vary the potential target of discrimination (e.g., black or white workers) randomly with equal probability. The question is structured for race as follows:

> Please indicate how likely you think it is that each company below would **discriminate against** [**black/white**] job seekers.

Responses are elicited on a 5-point scale ranging from "very unlikely" to "very likely," as

in previous questions, alongside an opt-out option. Respondents also rate their confidence in the same way as in the names arm.

Respondents are also asked a variant of this question focused on gender and age discrimination, where the group discriminated against either is younger or older (below/above the age of 40), or male or female, chosen with equal probability in each case. The order of the questions is random.

Block 4: Conclusion [all respondents]. To conclude the survey, we ask all respondents two final questions tied to the labor economics literature on discrimination. The first question elicits beliefs about the level of discretion that hiring managers possess. The second is about whether discrimination has influenced the respondent's job search. The first question takes the usual Likert form:

> At most companies, hiring decisions are made by individual managers. However, companies differ in the amount of oversight placed on these hiring decisions. For each company, please indicate how likely you think it is that managers can **hire their preferred candidate without input** from colleagues or superiors.

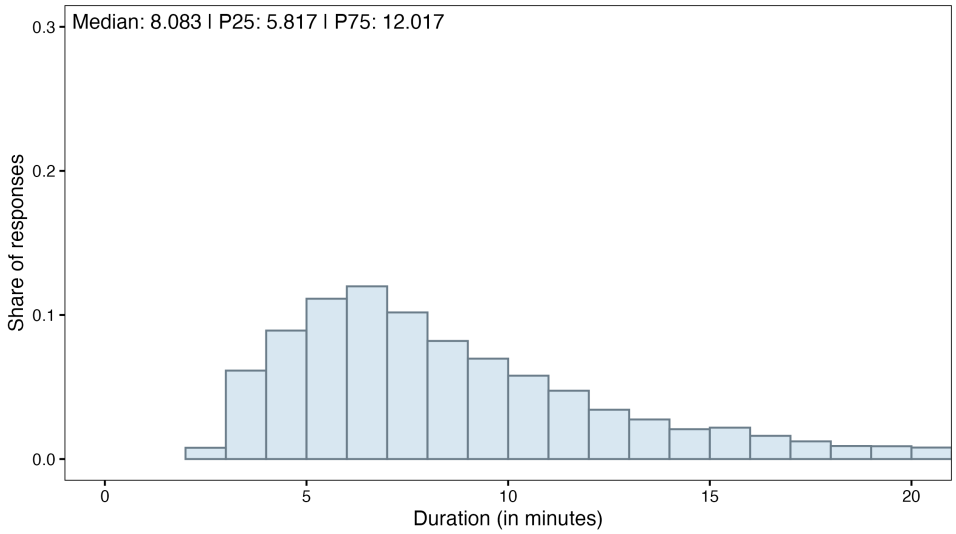
Finally, we asked respondents a simple yes/no question: "Have you ever put off applying for a job out of fear of discrimination?".

3 A first look at the survey data

Respondents in our estimation sample demonstrated engagement with the survey. Figure 1 reports the distribution of time spent on the survey. The median respondent spent 8 minutes completing the questions, the 25th percentile was 5.8 minutes, and the 75th percentile was 12.0 minutes.

Respondents also report familiarity with entry-level jobs and express a reasonable degree of confidence in their answers. Roughly 40% of respondents report having worked an

Figure 1: Distribution of survey completion times



Notes: This figure shows the distribution of time spent completing the survey for respondents in the final analysis dataset. The analysis dataset applies several quality controls, including dropping respondents who fail an attention check and respondents who straight-line all responses to key questions. See section 2.1 for further details on the sample.

entry-level job in the past two years. Table 2 reports the distribution of self-assessed confidence in answers to questions about the likely discriminatory behavior of five randomly selected firms. For each question, less than a third of respondents reported being "Not at all confident" or only "Slightly confident" in their answer. The share reporting "Not at all confident" was lowest (3.3 %) for the Discrimination Female (Conduct) question and highest (4.9%) for the Discrimination Black (Contact) question, suggesting respondents had stronger beliefs about relative patterns of firm gender discrimination than racial discrimination.

Table 2: Summary of respondent confidence assessments

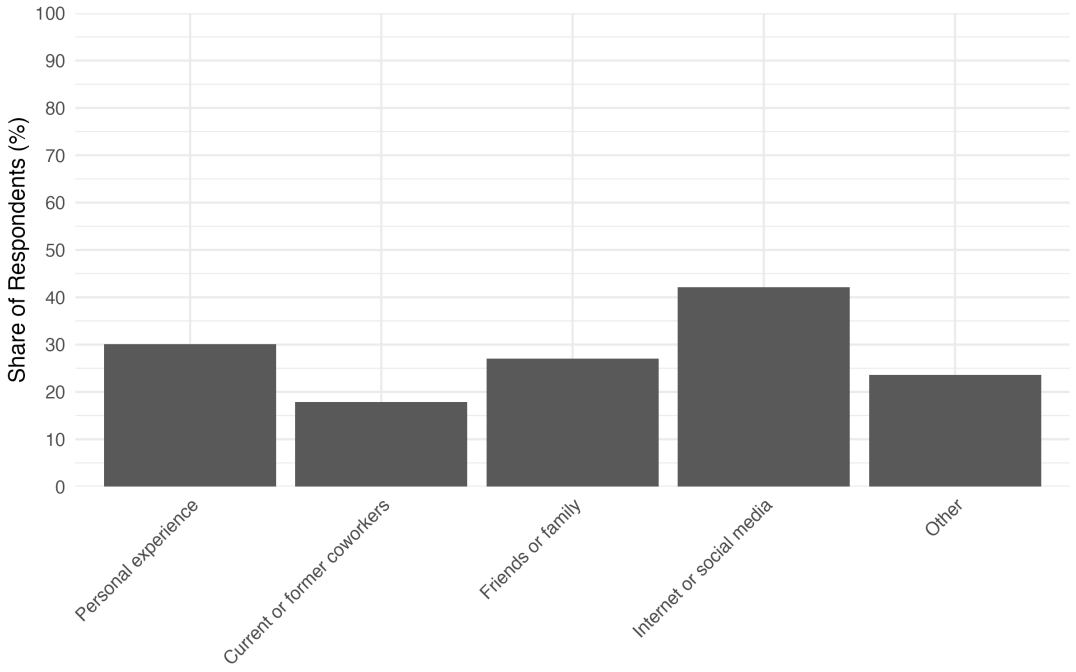
Measure	Not at all confident	Slightly confident	Somewhat confident	Very confident	Extremely confident	Missing
Discrimination Older (Conduct)	4.0	17.1	36.2	30.4	12.0	0.3
Discrimination Female (Conduct)	3.3	17.4	34.0	31.1	14.0	0.1
Discrimination Black (Conduct)	4.0	20.5	33.3	28.4	13.5	0.3
Discrimination Female (Contact)	4.3	22.8	36.6	26.2	9.9	0.3
Discrimination Black (Contact)	4.9	23.6	36.0	24.4	11.1	0.1

Notes: This table reports the distribution of respondents' confidence assessments for the primary questions in the names and conduct arms. As discussed in Section 2.2, confidence was assessed on a five point scale after each question in the conduct arm and after each pair of questions (for contact and for hiring likelihoods) in the names arm. Question text: How confident are you in your responses to the [question/two questions] above?

In addition to eliciting confidence about their answers, we asked respondents in the

conduct arm of the survey who stated that at least one of the companies was likely to discriminate against Black job seekers, where they obtained the information supporting this belief. Figure 2 reports the information sources these respondents listed. The most commonly reported sources of information were internet or social media, followed by personal experience, and family and friends.

Figure 2: Information source underlying racial discrimination perceptions



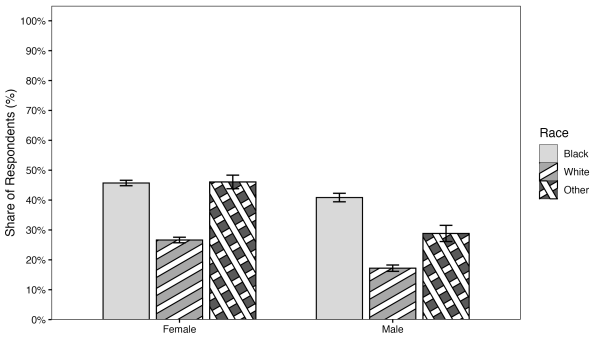
Notes: This figure shows the distribution of reported sources of information about firm-level discrimination. This question was only asked of respondents who indicated a company was very or somewhat likely to discriminate against Black job seekers in the conduct arm or indicated a company was much more or somewhat more likely to contact / hire a job seeker with a distinctively White name (or much less / somewhat less likely to contact / hire a job seeker with a distinctively Black name) in the names arm. Question text: Your responses indicated that you think [*company*] is likely to discriminate against Black applicants. What sources of information did you use to arrive at this conclusion?

A striking finding of the survey is that roughly one third of respondents report having put off looking for a job out of fear of discrimination. Figure 3 breaks down answers to this question by race and sex. Women are slightly more likely than men to report having been deterred from applying by discrimination. While a sizeable share of White respondents state that discrimination has affected their job search behavior, Black respondents are substantially more likely than White respondents to be deterred by discrimination.

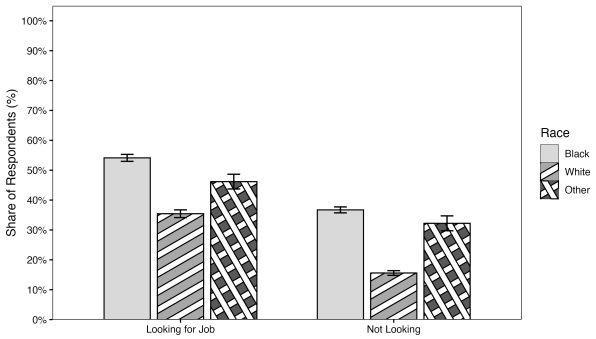
Another clear pattern, illustrated in the second panel, is that fear of discrimination is

Figure 3: Application deterred by fear of discrimination

(a) By sex



(b) By job search status

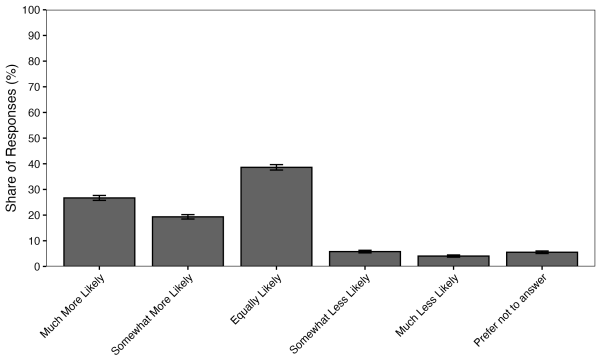


Notes: This figure reports the share of respondents who have ever been deterred from applying to a job due to fears of discrimination. Panel a breaks down respondents by race and sex, while Panel b breaks down responses by race and job search status. Question text: Have you ever put off applying for a job out of fear of discrimination?

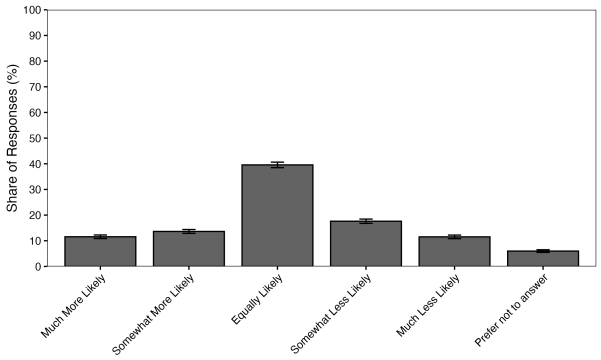
elevated among current job seekers of all races. More than half of Black respondents currently seeking a job state that fear of discrimination has deterred them from applying to a job. Roughly 1/3rd of white job seekers report the same.

Figure 4: Distribution of perceived contact discrimination

(a) Favor White vs. Black names



(b) Favor Black vs. White names



Notes: This figure reports the distribution of companies' perceived relative likelihood of contacting job seekers with different racially distinctive names. Only respondents in the names arm were asked this question. Panel (a) shows answers for the random subset of respondents asked about favoring White vs. Black names, while panel (b) shows answers for those shown the opposite structure. Question text: Please indicate how much more likely you believe the company is to contact a job applicant with a distinctively [White / Black] name than an applicant with a distinctively [Black / White] name.

Having established that perceived discrimination influences reported labor market search behavior, we now turn to examining beliefs about what sorts of discrimination are most prevalent. Figure 4 summarizes responses to the two variants of the question about racial discrimination from the survey's names arm, with the version listing White names first

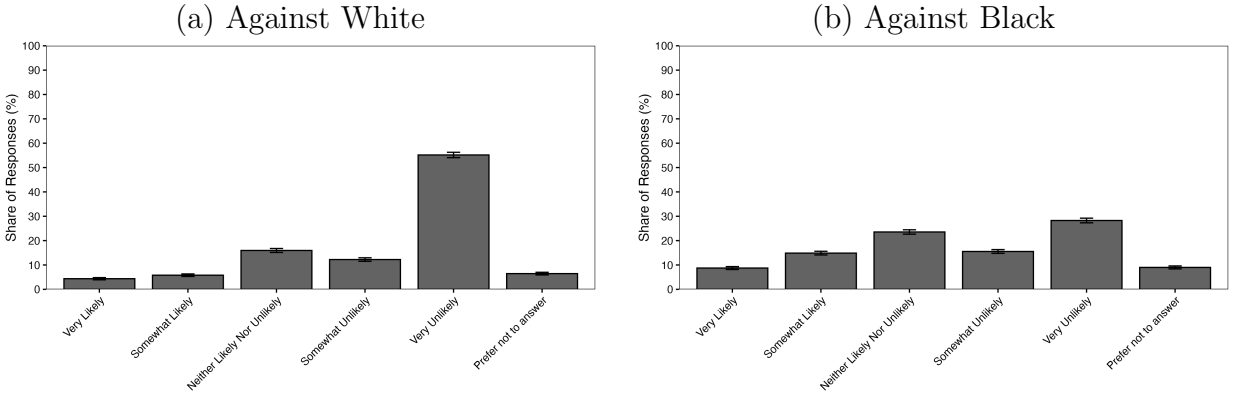
in panel (a) and the version listing Black names first in panel (b).

When distinctively White names are contrasted with distinctively Black names, a large share (28%) of respondents state that White names are much more likely to be contacted. However, when Black names are contrasted with White names, only about 11% say that Black names are much less likely to be contacted. Hence, there appears to be a non-negligible asymmetry in how respondents interpret the Likert scale across these two question framings.

Even larger asymmetries arise among variants of the conduct discrimination question summarized in Figure 5. Respondents are skeptical of the idea that White job seekers are discriminated against, with over half of respondents stating that such forms of discrimination are "very unlikely." This view accords with evidence from the large correspondence experiment analyzed by Kline, Rose and Walters (2022), who fail to reject the null hypothesis that no firms strictly prefer distinctively White to distinctively Black names. Because most respondents hold this view, there turns out to be little usable variation in this phrasing of the question. In contrast, conduct variants of the sex discrimination and age discrimination questions have lower shares of respondents claiming that discrimination is very unlikely and more who think discrimination is very likely (see Appendix A).

In what follows, we combine variants of each question in the names arm to extract as much information as possible. This is accomplished by reversing the Likert scale across question variants to make directions comparable, then pooling data for all variants. For the racial conduct question, this sort of pooling may be inappropriate because there is less signal in the question about discrimination against White job seekers. For consistency with our treatment of other questions, however, we continue to pool these two framings and explore robustness to estimating beliefs about discrimination against White and against Black respondents differently in robustness exercises. Pooling turns out to matter little, since variation in responses to the against Black framing tends to dominate.

Figure 5: Distribution of perceived racial discrimination likelihood



Notes: This figure reports the distribution of companies' perceived likelihood of discriminating against White job seekers (panel A) and against Black job seekers (panel B). Only respondents in the conduct arm were shown this question, with the against White vs. against Black framing chosen at random. Question text: Please indicate how likely you think it is that each company below would discriminate against [White / Black] job-seekers.

Since the Likert scores are based on subjects' interpretations of terms like "much more likely" and "somewhat more likely," it is plausible that not all respondents grade on the same scale. The asymmetries in responses to question framings documented above suggest further differences in scales across responses. Comparing the naive approach of simply averaging Likert scores for each firm to the rank-ordered approaches developed below allows us to examine sensitivity to respondent scale heterogeneity and potential framing effects.

4 Aggregation methods

We consider two approaches to aggregating responses to the questions in the names and conduct arms of the survey, as well as questions about firm selectivity and discretion. The first approach—Likert averaging—simply uses average responses on the 1-5 Likert scale for each firm. This approach exploits all the cardinal information embedded in numerical codings of Likert responses, but also relies on this common grading scale accurately capturing beliefs across respondents. The second approach—Borda ranking—relaxes this assumption by using only within-respondent ordinal comparisons as captured by Borda scores. While this method is more flexible, the cost is that only relative comparison across firms are recovered, resulting in the loss of cardinal information about the intensity of

perceptions.

4.1 Likert averaging

The Likert averaging approach first codes responses on a 1-5 scale (e.g., "Very likely" $\rightarrow$ 5, "Very unlikely" $\rightarrow$ 1). When combing two framings of the same question, we reverse the scale so that larger numbers always indicate more discrimination against the same group. We then take firm-level averages of responses for each question, or, equivalently, regress responses on a set of firm dummies using ordinary least squares (OLS). We refer to the sample average response for firm j, which we denote by $\hat{L}_j$, as the *Likert score*. For a given question, random assignment of choice sets to respondents guarantees that $\hat{L}_j$ is unbiased for its population average, which we denote by L_j.

Although the Likert scores are easy to construct, using them to understand beliefs can be challenging. Respondents may interpret the Likert options differently: a high-scoring firm for one respondent may be low-scoring for another even if their beliefs about conduct are identical. Even for the same respondent, the difference in beliefs about conduct between "very unlikely" and "somewhat unlikely" may not be the same as between "somewhat likely" and "very likely," although this method assigns both comparisons the same unit difference in scores.

These differences in grading scales across respondents and response categories generate two problems. One is that grading scale variation induces noise that inflates the standard errors on each firm's $\hat{L}_j$. A separate difficulty is that the units in which L_j is measured rely on the Likert cardinalization, which can be difficult to interpret when grading scales differ. The Borda scores help to address both of these difficulties.

4.2 Borda ranking

Borda (1784) scores were introduced as a positional voting system based on rank-ordered ballots that delivered global rankings. In our context, Borda scores are useful because they leverage *intra*-personal comparisons of firm Likert ratings in a way that differences

out respondent-specific rating thresholds.

To formally define Borda scores, let i index respondents and $C_i \subset \{1, \ldots, J\}$ denote respondent i's choice set, which is randomly assigned with $|C_i| = 5$. Write $j \succ_i k$ if respondent i ranks firm j strictly above firm k, meaning they assigned j a strictly greater Likert score than k, and $j \sim_i k$ if i is indifferent between them, meaning they assigned them the same Likert response. Define the "score" that firm j receives in the comparison with firm k for respondent i as

$$w_{ijk} = \mathbf{1}\{j \succ_i k\} + \tfrac{1}{2}\mathbf{1}\{j \sim_i k\}.$$

Note that ties yield a score of one half, while a strict "win" over another item yields a whole point.

Recall that each choice set C_i is comprised of a single anchor firm, drawn at random from the set $\mathcal{A}$, and four non-anchors drawn without replacement from the set $\mathcal{N}$. The Borda score is designed to measure the expected win rate of firm j against a randomly selected "opponent." For this to be a meaningful ranking metric, opponent strength should be equalized across all firms. However, our survey design prevents anchors from facing other anchors as opponents. To ensure anchors and non-anchors face opponents of equal expected strength, we work with a version of the Borda score that measures each firm's win rate against a randomly selected opponent in $\mathcal{N}$.

Define the *sample Borda score* for firm j as

$$\hat{B}_j = \underbrace{\frac{\mathbf{1}\{j \in \mathcal{N}\}}{2\,|\mathcal{N}|} + \frac{|\mathcal{N}| - \mathbf{1}\{j \in \mathcal{N}\}}{|\mathcal{N}|}}_{\text{non-anchor correction}} \underbrace{\frac{\displaystyle\sum_{i:j \in C_i} \sum_{k \in (C_i \cap \mathcal{N}) \setminus \{j\}} w_{ijk}}{\displaystyle\sum_{i:j \in C_i} (|C_i| - 1 - \mathbf{1}\{j \in \mathcal{N}\})}}_{\text{win rate}}.$$

The "win rate" term gives the share of times that respondents who could have ranked firm j higher than some other non-anchor firm, did so. Recall that in our experiment $|C_i| = 5$. Thus, the denominator of the win rate term, which counts the total number of

pairwise comparisons in which j could have been ranked above some non-anchor firm, will evaluate to 4 for anchors but only 3 for non-anchors. The "non-anchor correction" terms can be interpreted as granting each non-anchor a tie against itself and then rescaling the win rate accordingly. Thus, an anchor firm with 1 tie and 3 wins will have a $\hat{B}_j$ equal to a non-anchor with 3 wins.

For the population analogue, define the *pairwise win probabilities*

$$p_{jk} = \Pr(j \succ_i k \mid j, k \in C_i) \qquad \text{and} \qquad t_{jk} = \Pr(j \sim_i k \mid j, k \in C_i),$$

where the probabilities are taken over a randomly drawn respondent i conditional on both j and k appearing in their choice set. Now define the expected pairwise score

$$w_{jk} := \mathbb{E}[w_{ijk}] = p_{jk} + \tfrac{1}{2} t_{jk}.$$

With this notation, the *population Borda score* of firm j can be written

$$B_j := \mathbb{E}\left[\hat{B}_j\right] = \frac{\mathbf{1}\{j \in \mathcal{N}\}}{2\,|\mathcal{N}|} + \frac{|\mathcal{N}| - \mathbf{1}\{j \in \mathcal{N}\}}{|\mathcal{N}|} \mathbb{E}\left[\frac{\sum_{k \in (C_i \cap \mathcal{N}) \setminus \{j\}} w_{jk}}{|C_i| - 1 - \mathbf{1}\{j \in \mathcal{N}\}} \,\middle|\, j \in C_i\right],$$

where the remaining expectation averages over the random composition of the choice set. Since each non-anchor pair appears in the same choice set with equal probability, random assignment of non-anchors yields the simplification

$$B_j = \frac{1}{|\mathcal{N}|} \left[\tfrac{1}{2}\,\mathbf{1}\{j \in \mathcal{N}\} + \sum_{k \in \mathcal{N} \setminus \{j\}} w_{jk}\right],$$

Intuitively, the population Borda score grants each non-anchor a self-tie but otherwise treats them symmetrically to anchors.

Borda scores are attractive in our context because they are invariant with respect to differences in the location and scale of how respondents interpret Likert categories. They can be viewed as relaxations of alternative methods commonly used to analyze rank-

ordered lists and survey responses such as the rank-ordered logit model (Beggs, Cardell and Hausman, 1981; Hausman and Ruud, 1987; Allison and Christakis, 1994; Abdulkadiroğlu et al., 2020; Caldwell, Haegele and Heining, 2025). Logit methods rely on IIA, an assumption often found to be violated in actual choice data (Davidson and Marschak, 1959; Ballinger and Wilcox, 1997). In a context such as our survey, where choice sets have been randomly assigned, IIA is the assumption that the probability of ordering j higher than k is independent of which other firms, besides j and k, are included in the respondent's choice set.

Borda scores remain interpretable under a weaker condition known as strong stochastic transitivity (SST). As its name suggests, SST is a probabilistic generalization of transitivity: if j is likely to beat k, and k is likely to beat l, then j must be likely to beat l. To handle ties appropriately, we define SST in terms of the expected pairwise score w_{jk}. Note that $w_{jk} \geq 0.5$ if and only if j beats k at least as often as k beats j.

Definition 1 (Strong stochastic transitivity)**.** For all distinct firms j, k, ℓ, if

$$w_{jk} \geq 0.5 \quad \text{and} \quad w_{k\ell} \geq 0.5, \quad \text{then } w_{j\ell} \geq \max\{w_{jk}, w_{k\ell}\}.$$

While models predicted on IIA satisfy SST, the converse is not true. Many random utility models (e.g., multinomial probit) satisfy SST while violating IIA. Appendix C discusses a more general class of random utility models in which IIA fails due to context-dependence but SST continues to hold.

Korba, Clémençon and Sibony (2017) and Shah and Wainwright (2018) show that when SST holds, every firm faces every other firm as an opponent, and ties are impossible, the population Borda ordering coincides with the underlying preference ordering. In our design, anchors and non-anchors face different schedules of opponents. The "common-opponent" Borda score defined above corrects for this asymmetry by evaluating all firms against the same reference population ($\mathcal{N}$). This modification recovers the correct *global* ranking both anchor and non-anchor firms provided preferences satisfy SST and knife

edge indifferences are ruled out.

Proposition 1 (Global Consistency of Common-Opponent Borda under SST)**.** Assume:

1. **Strong Stochastic Transitivity (SST):** The pairwise scores $\{w_{jk}\}$ satisfy SST.
2. **No knife-edge indifference:** For all distinct firm pairs j, k, $w_{jk} \neq 0.5$.
3. **Symmetric Design:** The survey consists of two disjoint sets of firms, $\mathcal{A}$ (Anchors) and $\mathcal{N}$ (Non-anchors). Choice sets are formed by drawing one firm uniformly from $\mathcal{A}$ and four uniformly from $\mathcal{N}$.

Then, for *any* pair of distinct firms $j, k \in \mathcal{A} \cup \mathcal{N}$, the population Borda ranking coincides with the pairwise majority relation:

$$B_j > B_k \quad \Longleftrightarrow \quad w_{jk} > 0.5 \quad (\Longleftrightarrow p_{jk} > p_{kj}).$$

Proof. See appendix. □

4.3 Estimation

For Likert scores, we denote the average score across firms in our sample by $\bar{L} = J^{-1} \sum_{j=1}^{J} \hat{L}_j$, with corresponding population average $\mu_L = J^{-1} \sum_{j=1}^{J} L_j$. For Borda scores, we denote the sample average by $\bar{B} = J^{-1} \sum_{j=1}^{J} \hat{B}_j$ and the population average by $\mu_B = J^{-1} \sum_{j=1}^{J} B_j$. Both the estimated Likert scores $\hat{L}_j$ and the Borda scores $\hat{B}_j$ are consistent and asymptotically normal under standard regularity conditions. We estimate their associated asymptotic standard errors $\mathbb{V}[\hat{L}_j]^{1/2}$ and $\mathbb{V}[\hat{B}_j]^{1/2}$ analytically. Since the sampling variability in these standard error estimates is asymptotically negligible, we treat $V_{Lj} \equiv \mathbb{V}[\hat{L}_j]$ and $V_{Bj} \equiv \mathbb{V}[\hat{B}_j]$ as known.

4.3.1 Variances

To summarize the strength of beliefs about firms, we will be interested in the population variance

$$\sigma_L^2 = \frac{1}{J}\sum_{j=1}^{J}(L_j - \mu_L)^2, \quad \sigma_B^2 = \frac{1}{J}\sum_{j=1}^{J}(B_j - \mu_B)^2.$$

Asymptotically unbiased estimators of these variances can be obtained by subtracting the squared asymptotic standard errors as follows

$$\hat{\sigma}_L^2 = \frac{1}{J}\sum_{j=1}^{J}\left((\hat{L}_j - \bar{L})^2 - V_{Lj}\right), \quad \hat{\sigma}_B^2 = \frac{1}{J}\sum_{j=1}^{J}\left((\hat{B}_j - \bar{B})^2 - V_{Bj}\right).$$

While $\hat{\sigma}_L^2$ and $\hat{\sigma}_B^2$ are asymptotically unbiased, sampling error can lead to negative estimates in finite samples. To avoid such logical inconsistencies, we refine our variance estimates with a correction proposed by Katz (1961) that ensures the estimates are positive. The correction depends on the squared standard errors $\mathbb{V}[\hat{\sigma}_L^2]$ and $\mathbb{V}[\hat{\sigma}_B^2]$ of the variance estimators. We derive asymptotically unbiased estimators $\widehat{\mathbb{V}}[\hat{\sigma}_L^2]$ and $\widehat{\mathbb{V}}[\hat{\sigma}_B^2]$ of these quantities analytically using formulas given in Appendix E.

Define the z-scores $z_L = \hat{\sigma}_L^2/\widehat{\mathbb{V}}[\hat{\sigma}_L^2]^{1/2}$ and $z_B = \hat{\sigma}_B^2/\widehat{\mathbb{V}}[\hat{\sigma}_B^2]^{1/2}$, and let $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal pdf and cdf, respectively. The corrected estimators take the form:

$$\dot{\sigma}_L^2 = \hat{\sigma}_L^2 + \widehat{\mathbb{V}}[\hat{\sigma}_L^2]^{1/2}\frac{\phi(z_L)}{\Phi(z_L)}, \quad \dot{\sigma}_B^2 = \hat{\sigma}_B^2 + \widehat{\mathbb{V}}[\hat{\sigma}_B^2]^{1/2}\frac{\phi(z_B)}{\Phi(z_B)}.$$

Note that these formulas correspond to the mean of a normal distribution truncated from below at zero. Consequently, noisy estimates of the variance components are nudged up by the product of their standard error and an inverse Mills ratio evaluated at the z-score of the point estimate. When the standard error is small, the estimator collapses to the uncorrected estimate. Katz (1961) shows that these shape-constrained estimators are minimax and admissible under squared error loss.

4.3.2 Shrinkage

With $J = 164$ firms, the standard errors for any particular firm's Likert score or Borda score will generally be non-trivial. We therefore use shrinkage estimators when reporting collections of firm-level estimates.

Following Walters (2024), we implement an Empirical Bayes shrinkage procedure that accounts for precision dependence: the possibility that the distribution of latent firm parameters varies with the precision of their estimates. First, we remove the dependence of the mean on the standard error by fitting regressions of the estimates on their log standard errors. Let the mean-detrended residuals be:

$$\tilde{L}_j := \hat{L}_j - \hat{\gamma}_{L0} - \hat{\gamma}_{L1} \ln V_{Lj}, \quad \tilde{B}_j := \hat{B}_j - \hat{\gamma}_{B0} - \hat{\gamma}_{B1} \ln V_{Bj},$$

where $(\hat{\gamma}_{L0}, \hat{\gamma}_{L1}, \hat{\gamma}_{B0}, \hat{\gamma}_{B1})$ are the fitted regression coefficients.

Next, we allow the variance of the underlying signal to vary with precision. We estimate the following model of the conditional variance of the residuals by nonlinear least squares:

$$\mathbb{E}\left[\tilde{L}_j^2 \mid V_{Lj}\right] = V_{Lj} + \underbrace{\kappa_L \left(V_{Lj}\right)^{\delta_L}}_{\tau^2_{L,j}}, \quad \mathbb{E}\left[\tilde{B}_j^2 \mid V_{Bj}\right] = V_{Bj} + \underbrace{\kappa_B \left(V_{Bj}\right)^{\delta_B}}_{\tau^2_{B,j}}.$$

Here, $\tau^2_{\cdot,j}$ represents the estimated variance of the true signal for firm j. We use the fitted parameters $(\hat{\kappa}, \hat{\delta})$ to construct estimated signal variances $\hat{\tau}^2_{\cdot,j}$.

Finally, we construct shrinkage estimators by shrinking the residuals and adding back the estimated mean trends:

$$L_j^* = \hat{\gamma}_{L0} + \hat{\gamma}_{L1} \ln V_{Lj} + \left(\frac{\hat{\tau}^2_{L,j}}{\hat{\tau}^2_{L,j} + V_{Lj}}\right) \tilde{L}_j,$$

$$B_j^* = \hat{\gamma}_{B0} + \hat{\gamma}_{B1} \ln V_{Bj} + \left(\frac{\hat{\tau}^2_{B,j}}{\hat{\tau}^2_{B,j} + V_{Bj}}\right) \tilde{B}_j.$$

The shrinkage factors in parentheses represent the reliability ratios of the residuals. Adapting the shrinkage intensity to the specific noise profile of each unit—as achieved here by the variable reliability ratio—not only yields an improved linear predictor but also potentially improves the accuracy of rankings over the raw point estimates. In particular, if the underlying signals are normally distributed, then the ordering of the shrinkage estimators generally maximizes the probability of correctly ranking the items (Portnoy, 1982).

5 Signal properties

We begin with a summary of the variances of firm-level beliefs for each question. Then we move on to describe the relationship between questions. Finally, we use shrinkage methods to provide estimates of firms believed to have strong racial and gender preferences.

5.1 Variance components

The dispersion of Likert scores measures differences in average beliefs about conduct across firms. A larger standard deviation indicates more agreement across respondents in the sense that their responses for each firm tend to be similar. If all respondents answered randomly, then each firm's average score would collapse to the grand mean. The standard deviation of Likert scores also indicates the extent of perceived differences in conduct across firms, since it shares the same units as the underlying Likert scales.

Dispersion in Borda scores also measures the extent of rank-order agreement across respondents. At one extreme, if respondents assigned all firms the same Likert score (leading to pure ties) or each respondent was as likely to state that firm j discriminates more than firm k as the opposite, Borda scores would all collapse to one half. On the other hand, if respondents all shared the same global ordering of relative conduct across firms, Borda scores would be maximally dispersed: one firm would always win against all opponents, one would win against all but one, and so on.

Table 3 reports estimates of the dispersion in Likert and Borda scores. The column "Std Dev" reports the naive sample standard deviation of scores, while "Signal Std Dev" reports the square roots $\dot{\sigma}_L$ and $\dot{\sigma}_B$ of the minimax variance estimate that accounts for sampling error. The column "T-stat no signal" tests the null hypothesis that all firm-level scores are equal, so that the true signal variance is zero. For Borda, we also report the ratio of the signal variance to the maximal variance possible, which corresponds to the scores that result when all respondents share a common strict preference ordering.[1]

Table 3: Variance of firm-level beliefs

	Likert			Borda			
Outcome	Std Dev	Signal Std Dev	T-stat no signal	Std Dev	Signal Std Dev	T-stat no signal	Normed Variance
Discrimination Black (Conduct)	0.312	0.233	5.829	0.080	0.064	6.968	0.048
Discrimination Black (Contact)	0.336	0.296	10.643	0.102	0.093	13.158	0.102
Discrimination Black (Hire)	0.345	0.304	10.573	0.101	0.092	12.646	0.100
Discrimination Black (Pooled)	0.300	0.271	11.972	0.087	0.081	14.731	0.077
Discrimination Female (Conduct)	0.486	0.449	14.251	0.124	0.117	17.009	0.162
Discrimination Female (Contact)	0.483	0.460	16.863	0.127	0.121	18.496	0.172
Discrimination Female (Hire)	0.512	0.490	18.321	0.131	0.125	19.109	0.185
Discrimination Female (Pooled)	0.467	0.452	22.006	0.122	0.118	24.687	0.165
Discrimination Older (Conduct)	0.277	0.219	6.859	0.075	0.061	8.206	0.044
Firm Desirability	0.182	0.144	7.087	0.052	0.042	8.286	0.021
Firm Selectivity	0.637	0.629	37.614	0.158	0.156	40.275	0.290
Manager Discretion	0.470	0.455	21.190	0.113	0.110	21.764	0.143

Notes: This table reports estimates of the dispersion of Likert and Borda scores for several survey questions. "Discrimination Black (Conduct)" pools the two framings of the race discrimination question in the conduct arm. "Discrimination Black (Contact)" and "Discrimination Black (Hire)" refer to two questions in the names arm, with both framings (e.g., White vs. Black or Black vs. White) combined after appropriate recoding. "Discrimination Black (Pooled)" combines the questions underlying "Discrimination Black (Conduct)" and "Discrimination Black (Contact)". The "Discrimination Female" and "Discrimination Older" questions are defined analogously. Columns labeled "Std Dev" report the square root of the sample variance of estimates. "Signal Std Dev" reports the square root of asymptotically unbiased estimates of the variances of scores, or $\dot{\sigma}_L$ and $\dot{\sigma}_B$, respectively. "T-stat no signal" reports $\hat{\sigma}_L^2/\mathbb{V}[\hat{\sigma}_L^2]^{0.5}$ and $\hat{\sigma}_B^2/\mathbb{V}[\hat{\sigma}_B^2]^{0.5}$, which allows a t-test of the null hypothesis that the signal standard deviation is zero. "Normed variance" reports the ratio of the Borda signal variance to its theoretical maximum, which occurs when all respondents share a common strict preference ordering.

For Likert scores, accounting for sampling error yields a modest reduction in the estimated standard deviation. For each question, a t-test easily rejects the null hypothesis that the true signal standard deviation is zero. The estimated signal standard deviations of Likert scores for the racial discrimination questions range from 0.23 to 0.30, or roughly

[1]For n firms, population Borda scores in this case would be $\{0, 1/(n-1), 2/(n-1), ..., 1\}$. The variance of that array is $(n+1)/(12(n-1))$, which evaluates to approximately 0.0844 for $n = 164$.

a third of a Likert point. The standard deviations associated with questions about sex discrimination are higher, ranging from 0.45 to 0.49. The age discrimination question has a much smaller standard deviation of 0.22, indicating less agreement about firm-level conduct in this dimension.

Interestingly, the question on firm desirability has the smallest signal standard deviation of all measures, perhaps indicating that respondents found the question difficult to answer. In contrast, the questions about firm selectivity and manager discretion elicited as much or more agreement as the gender discrimination questions.

The Borda measures show similar patterns. As with Likert scores, we can decisively reject the null of zero variance. Signal variances are higher for sex discrimination than race discrimination questions, and smallest for age discrimination. The firm desirability measure exhibits the weakest signal, while firm selectivity exhibits the strongest. Despite large signal variances for many questions, the normed variances indicate that the distribution of Borda scores is typically far from its theoretical maximum implied by common strict orderings. This may reflect differences in beliefs across sub-populations, a question we return to below.

As noted in Section 3, these results pool responses to the different framings of questions in the names and conduct arm. Figure A3 shows that firm-level beliefs are tightly correlated when estimated in each framing separately with the exception of the conduct arm for race discrimination. Table A1 reports signal variances for each question and framing separately. Consistent with the limited dispersion in responses documented in Section 3, the signal variance is lowest for beliefs about discrimination against White workers in the conduct arm, with the t-test of no-signal failing to reject.

An important component of beliefs may reflect respondents' impressions of industry-level patterns as opposed to the conduct of specific firms. Table A2 explores the role of industry in generating the dispersion documented in Table 3 by computing the same statistics for industry-demeaned Likert and Borda scores (to capture dispersion within-industry) and industry-average scores (to capture between industry variation). For most questions, the

within- and between-industry components are comparable, although the signal standard deviation tends to be larger between-industry. For beliefs about discrimination against Black workers elicited in the conduct arm, for example, industry average beliefs capture 68% of the total variance. Between industry-variation is also important for beliefs about manager discretion and selectivity. We explore industry-level beliefs about conduct further below.

5.2 Correlation across questions

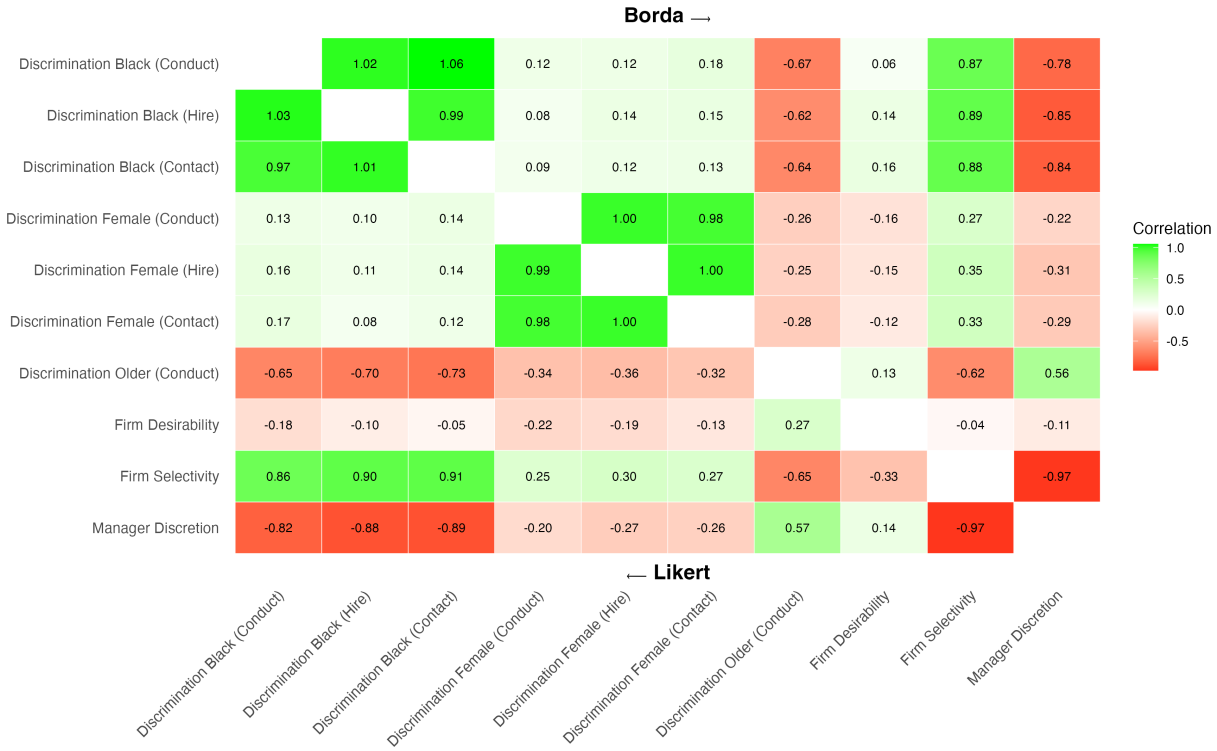
Figure 6 reports signal correlations of firm beliefs across questions. The lower triangle of the Figure reports the estimated signal correlation of the Likert scores. The upper triangle shows the correlation structure of the firm Borda scores. As described in the Appendix, these correlation estimates remove the influence of sampling error by using an estimate of the covariance matrix of the noise between the two questions.

Despite the different scales on which the Likert and Borda estimators measure beliefs, the correlation matrix turns out to be very nearly symmetric. This symmetry suggests the information content of Likert scores, which encode both interpersonal and intrapersonal comparisons, agrees closely with that of Borda scores, which exclusively aggregate intrapersonal comparisons.

Beliefs about racial discrimination are essentially perfectly correlated across questions. Evidently, respondent beliefs about the overall discriminatory conduct of firms align closely with their beliefs about the treatment of job applicants with racially distinctive names—that is, what would be measured in a typical audit study. The same pattern is found among the sex discrimination questions, which exhibit nearly perfect correlation as well.

Interestingly, the race and sex questions are only weakly correlated with each other. In contrast, companies believed to discriminate against Black candidates are perceived as being substantially less likely to discriminate against older candidates. Companies

Figure 6: Correlation of firm-level beliefs across questions



Notes: This figure reports the pairwise signal correlation in firm-level parameters across survey questions. Parameters are signed such that more positive values indicate discrimination against the group in the question label. Entries below the diagonal report the signal correlation in Likert scores. Entries above the diagonal give the signal correlation in Borda scores. In all cases the denominators of these correlations are formed based on the asymptotically unbiased variance estimators described in Section 4.3.1. For question pairs that are never asked of the same respondent (e.g., questions in the names and conduct arms) the sample covariance is used in the numerator. For covariances between pairs of questions asked of the same individual, we subtract an analytical estimate of the sampling covariance from the numerator.

believed to discriminate against women are thought to be less likely to discriminate against older workers.

Firm desirability is not strongly correlated with measures of discrimination. However, firms believed to exhibit the most selective hiring are believed to discriminate heavily against Black and younger job seekers. Such firms are also believed to prefer men to women, although the relationship is weaker. While selectivity logically introduces an opportunity for bias, the direction of the biases—favoring, for example, older over younger workers—was not obvious ex-ante. These correlations suggests a belief that elite institutions are biased against less powerful groups that are more broadly disadvantaged in American society.

Finally, firms believed to exhibit greater manager discretion are thought to exhibit less discrimination against Black job seekers, women, and young workers. The strongest correlate of perceived discretion is selectivity: the most selective firms are thought to exhibit the least discretion, perhaps because hiring processes are believed to be more standardized.

The strong negative relationship between perceptions of manager discretion and racial discrimination is surprising in light of recent research indicating that discretion is an important conduit through which bias arises (Hoffman, Kahn and Li, 2018; Kline, Rose and Walters, 2022; Mocanu, 2022). One possibility is that these discretion-discrimination correlations are driven by the strong negative correlation between discretion and selectivity.

To investigate this concern we can convert the correlations reported in Figure 6 to partial correlations. Recall that the partial correlation between two variables conditional on a third variable can be written

$$\rho_{12|3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{23}^2)}},$$

where ρ_{xy} gives the pairwise correlation between x and y.

While the marginal correlation in Likert scores between Discrimination Black (Conduct) and Manager Discretion is -0.82, their partial correlation conditional on Firm Selectivity is 0.11, indicating a positive but weak relationship. On the other hand, the partial correlation in Borda scores conditional on Firm Selectivity is 0.53, a moderately strong relationship. We conclude that the negative marginal correlation between discriminatory beliefs and perceived managerial discretion found with both sorts of scores is largely driven by perceived selectivity. It is difficult to say more about the likely magnitude of the seemingly positive conditional relationship because of the near perfect anti-correlation of the Firm Selectivity and Manager Discretion variables, which makes the denominator of the partial correlation expression extremely small.

5.3 Correlation across subpopulations

We saw earlier that the gender discrimination questions have higher signal variance than the racial discrimination questions. This finding suggests that two randomly selected respondents are more likely to agree in their beliefs about gender discrimination than racial discrimination. However, it is possible that views differ systematically among respondents with different observed characteristics. For instance, one might suspect that Black respondents have more information about racial discrimination than White respondents, in which case the correlation in firm beliefs across racial groups might be attenuated.

To investigate such hypotheses, we compute Likert and Borda scores separately for populations defined by demographic groups and answers to the questions about looking for work and fear of discrimination. Table 4 reports estimates of the signal correlation across these subgroups along with tests of the null hypothesis that their beliefs—as quantified by their Likert and Borda scores—are identical. Panel A contains results for Likert scores, while Panel B studies Borda scores.[2]

The first two columns of the table study beliefs about race discrimination. While we can confidently reject that Likert scores are equal across across Black and White respondents,

[2]Table A3 presents the raw correlations in estimated scores across groups without correction for estimation error.

Table 4: Cross Sample Signal Correlations

	Discrimination Black		Discrimination Female	
	Corr	Wald p-value	Corr	Wald p-value
Panel A: Likert				
Black vs White	0.935	0.001	0.992	0.132
Female vs Male	1.061	0.615	0.977	0.009
Looking for a Job vs Not	1.021	0.620	0.993	0.298
Feared Discrimination vs Not	1.023	0.645	0.994	0.340
Age >= 40 vs < 40	1.026	0.488	0.967	0.013
At Least Some College vs HS Diploma or less	0.898	0.046	0.961	0.002
Convenience vs Probability	0.926	0.074	1.004	0.341
Confident vs Not (Gender)	0.977	0.063	0.996	0.243
Confident vs Not (Race)	0.975	0.010	0.958	0.005
Panel B: Borda				
Black vs White	0.970	0.037	0.992	0.027
Female vs Male	0.980	0.213	0.962	0.003
Looking for a Job vs Not	1.007	0.558	0.985	0.134
Feared Discrimination vs Not	1.026	0.617	0.988	0.188
Age >= 40 vs < 40	0.896	0.000	0.993	0.108
At Least Some College vs HS Diploma or less	1.021	0.383	0.979	0.001
Convenience vs Probability	1.010	0.468	0.975	0.008
Confident vs Not (Gender)	0.991	0.273	0.975	0.025
Confident vs Not (Race)	1.000	0.317	0.960	0.002

Notes: This table reports the correlation in beliefs across sample groups. Columns labeled "Corr" give the estimated signal correlation in firm-level parameters across the two groups. Columns labeled "Wald p-value" evaluate the null hypothesis that the two vectors of firm-level parameters are equal via a Wald test. Panel (a) uses Likert scores, while panel (b) uses Borda scores.

their signal correlation is estimated to be very close to one. This finding suggests that Black and White respondents differ only in the dispersion of their beliefs as quantified by the Likert scores. As discussed above, a group with more dispersed scores can be thought of as exhibiting greater agreement within the group. The correlation across Black and White respondents in Borda scores is similarly high, indicating strong agreement about relative conduct across firms as well.

Beliefs about racial discrimination are strongly correlated between all of the other subgroups considered as well. Men and women, those looking for a job and those not looking, and those who reported putting of a job application due to fear of discrimination and those who did not all seem to exhibit nearly perfect correlation. Moreover, we cannot reject that the firm scores are identical across these groups, which suggests the dispersion of beliefs across these groups is also similar. Reassuringly, we also cannot reject that be-

liefs are equal in the convenience and probability samples, which suggests the information encoded in the Likert and Borda scores is not especially sensitive to the method used to sample respondents.

The largest discrepancy in beliefs emerges between those with at least some college vs. those with a high school degree or less. Interestingly, the correlation in beliefs across these two groups is higher for Borda scores in Panel B, suggesting any discrepancies relate to grading scales rather than beliefs about relative conduct. Comparisons between our two sources of survey participants—the convenience and the probability samples—show similar patterns, with still high but relatively lower correlations in Likert scores and near perfect correlations in Borda scores. The lowest between-group correlation in Borda scores is for respondents over vs. under age 40.

Table 4 repeats this exercise for sex discrimination. Here, the correlations are extremely close to one for all groups despite strong evidence that the scores are not identical across sexes, age groups, or educational attainment. This again signals that dispersion in firm scores can differ meaningfully across groups without meaningfully affecting firm rankings in scores.

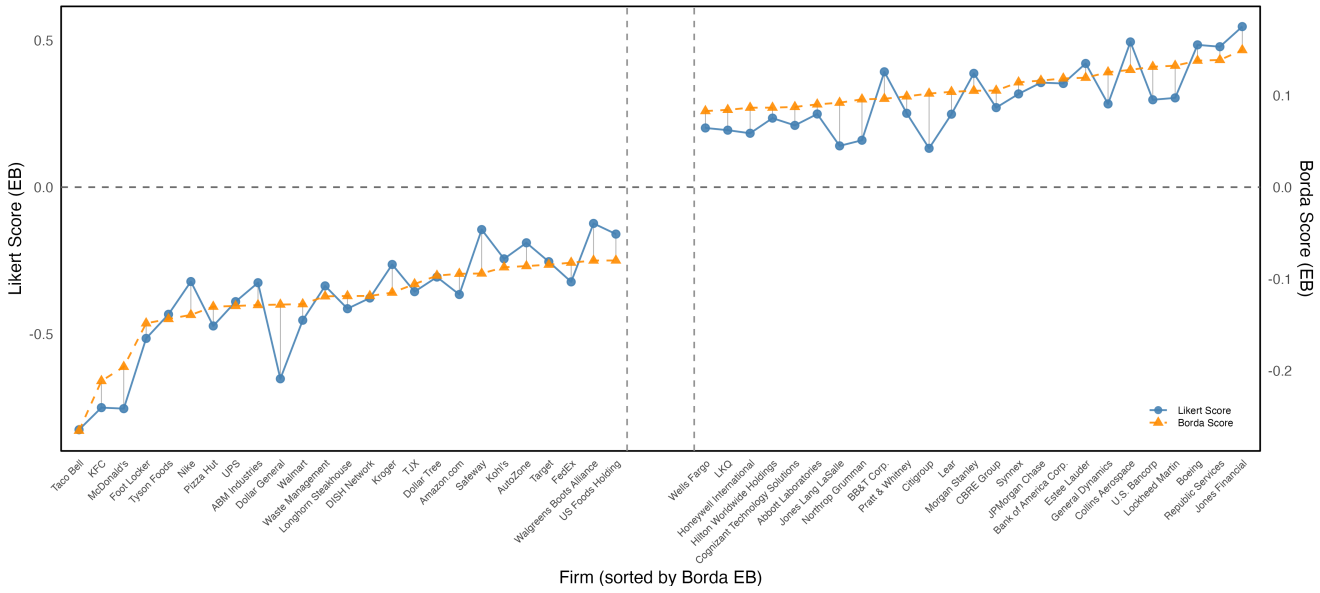
6 Which firms are thought to discriminate?

We now turn to listing some of the firms that are thought to have the strongest preferences over race and sex. In what follows, we report beliefs about firms whose shrunk Borda scores $\{B_j^*\}_{j=1}^J$ are among the top 25 or bottom 25 of those estimated. We then move on to assessing industry average scores.

6.1 Top / bottom 25 firms

Figure 7 reports estimated Likert and Borda scores for racial discrimination obtained by pooling both framings of the conduct and contact questions. The non-monotonicities in the L_j^* series reflect differences in the rankings of the shrunk Likert and Borda scores.

Figure 7: Top/Bottom 25 firms by perceived discrimination against Black job seekers



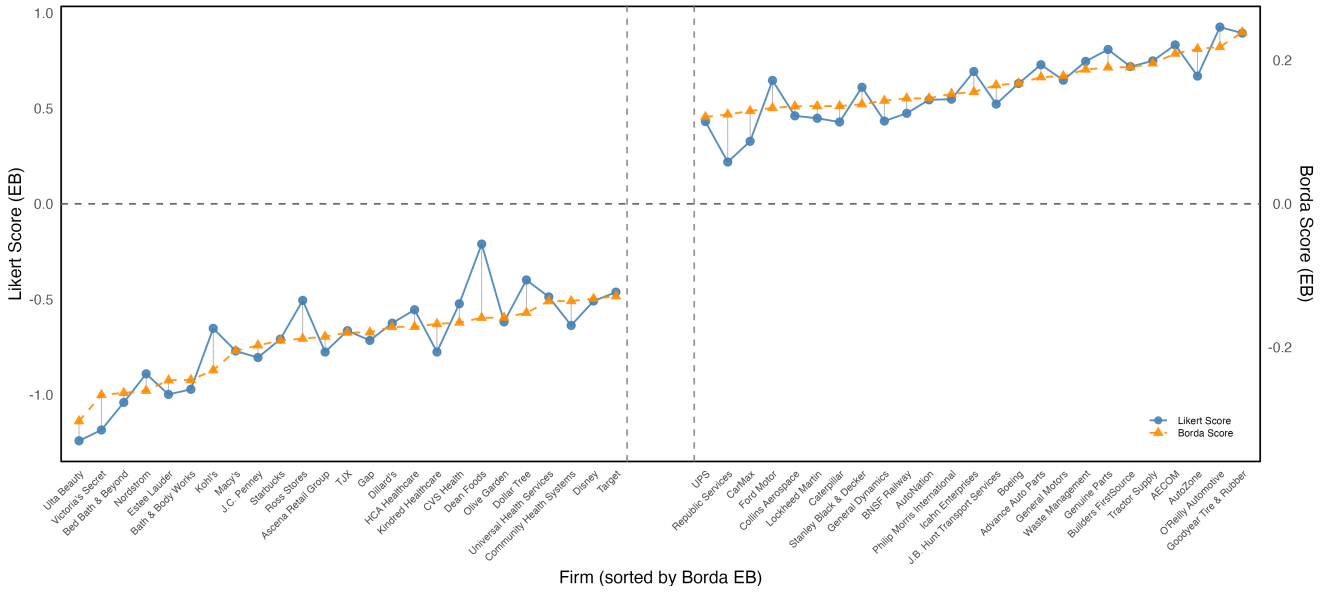
Notes: This figure reports the top and bottom 25 firms ranked by their perceived likelihood of discriminating against Black job seekers according to their shrunken (EB) Borda scores, or B_j^*. The outcome pools responses to both versions of the race question in the conduct arm and both framings of the contact question in the names arm, as in "Discrimination Black (Pooled)" in Table 3. For comparison, the figure also reports shrunken Likert scores, or L_j^*. Firms on the right-hand side of the figure are perceived as more likely to discriminate against Black job seekers.

While such rank disagreements are common, they turn out to be relatively minor in magnitude.

The bending of the Likert and Borda lines on the left of the figure indicates that respondents concur most often about firms believed least likely to discriminate against Black job applicants. The firms thought least likely to discriminate against Black job seekers are Taco Bell, KFC, and McDonald's. In contrast, respondents believe Boeing, Republic Services, and Jones Financial are the most likely to discriminate against Black job seekers. The difference in Likert scores between the top and bottom two firms is slightly more than 1. The difference in Borda scores across the top and bottom firms is roughly 0.5, indicating that top firms are 50 percentage points more likely to be rated by the typical respondent as more discriminatory than a randomly selected firm.

Figure 8 reports corresponding Likert scores and Borda scores for gender discrimination. Interestingly, these schedules of estimated beliefs do not bend in the extremes. Respondents believe that Ulta Beauty, Victoria's Secret, and Bed Bath & Beyond strongly favor

Figure 8: Top/Bottom 25 firms by perceived discrimination against female job seekers



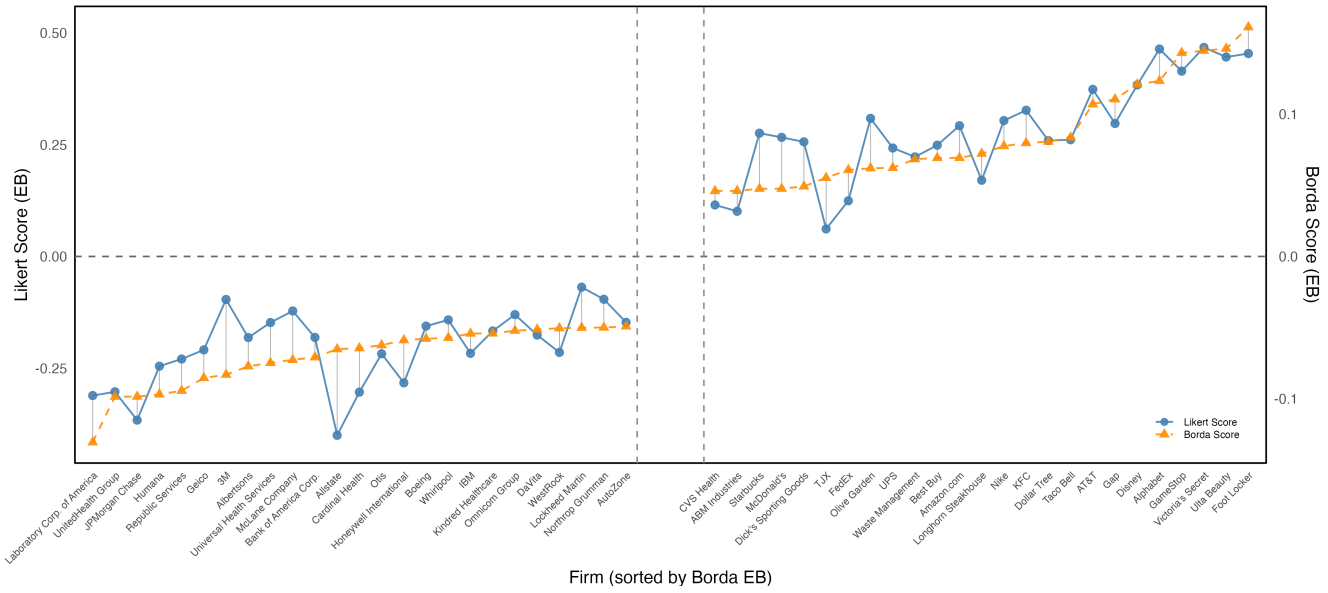
Notes: This figure reports the top and bottom 25 firms ranked by their perceived likelihood of discriminating against female job seekers according to their shrunken (EB) Borda scores, or B_j^*. For comparison, the figure also reports shrunken Likert scores, or L_j^*. The outcome pools responses to both versions of the gender question in the conduct arm and both framings of the contact question in the names arm, as in "Discrimination Female (Pooled)" in Table 3. Firms on the right-hand side of the figure are perceived as more likely to discriminate against female job seekers, while firms on the left-hand side are perceived as more likely to discriminate against male job seekers.

women while AutoZone, O'Reilly Automative, and Goodyear Tire & Rubber strongly favor men. Here, the magnitude of the difference in average Likert scores between these firm pairs is roughly 2 points. The difference in Borda scores is close to 0.6, implying the firms that favor men most are 60 percentage points more likely to be rated by the typical respondent as favoring men more than a randomly selected firm.

Finally, Figure 9 reports perceptions of firms' age preferences. Here, the Likert and Borda estimates exhibit greater disagreement, although in general the correlation in the two metrics is high. The Likert estimates indicate that the two firms that favor young workers most are Victoria's Secret and Alphabet, while the two that favor older workers most are Allstate and JPMorgan Chase. The gap in scores between the top and bottom firms is comparatively small—roughly 0.88.

In contrast to the Likert estimates, the Borda scores suggest that Ulta Beauty and Foot-Locker exhibit the strongest preference for young workers. Conversely, Lab Corp. and

Figure 9: Top/Bottom 25 Firms Sorted by Perceived Age Discrimination



Notes: This figure reports the top and bottom 25 firms ranked by their perceived likelihood of discriminating against older job seekers according to their shrunken (EB) Borda scores, or B_j^*. For comparison, the figure also reports shrunken Likert scores, or L_j^*. The outcome pools both versions of the age question in the conduct arm. Firms on the right-hand side of the figure are perceived as more likely to discriminate against older job seekers.

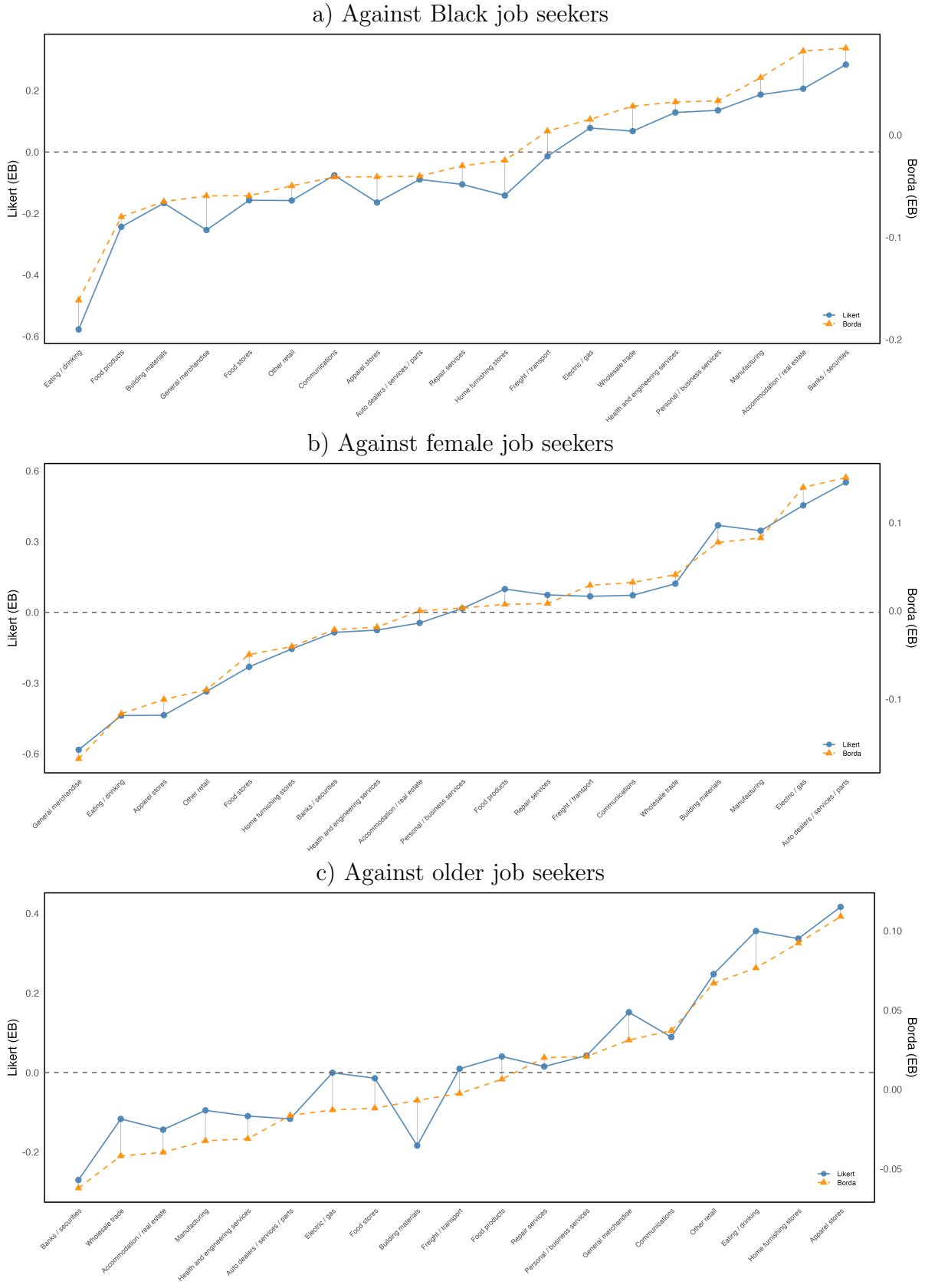
UnitedHealth are thought to favor older workers most strongly. Borda score and Likert scores diverge noticeably for some firms, such as 3M, Allstate, and Starbucks. The distribution of Borda scores is considerably more compressed than for race and sex discrimination, however, consistent with lower overall agreement across respondents about age discrimination.

6.2 Industry-level beliefs

As noted in Section 5.1, a large fraction of the variation in Likert and Borda scores across firms can be captured by variation in industry-average beliefs. Figure 10 examines the structure of beliefs across industries by plotting EB estimates of average Likert and Borda scores for firms in major industry categories. Panel a) studies beliefs about discrimination against Black workers using the same data underlying Figure 7. Panel b) studies discrimination against Female workers using the data from Figure 8. Finally, Panel c) studies age discrimination using the data from Figure 9.

Industry-average beliefs about discrimination against Black workers show wide dispersion.

Figure 10: Distribution of industry-average beliefs



Notes: This figure reports industry-average beliefs ranked by their shrunken (EB) Borda scores, or B_j^*, and shrunken Likert scores, or L_j^*, for race, gender, and age. The outcome is the same as in Figure 7, 8, and 9, but averaged by industry.

The sector where firms are believed least likely to discriminate against Black workers is the eating / drinking sector, which has an average Likert score near a full point below the banks / securities sector, where workers believe firms are most likely to discriminate against Black workers. The difference in Borda scores across the top / bottom industries is also large. Compared to a random firm, a firm in the eating / drinking sector is roughly 17 percentage points less likely to be ranked as more discriminatory than a firm in the banks / securities sector. Ranking industries by Borda scores and Likert scores also roughly agree, although there are some cases of non-monotonicity. Notably, the auto sector, which exhibited the largest contact gaps in Kline, Rose and Walters (2024) receives a middling score here under both the Borda and Likert rankings.

Beliefs about gender discrimination at the industry level show more dispersion than beliefs about race discrimination. The sectors where firms are judged least likely to discriminate against women are general merchandise, eating / drinking, and apparel stores. The eating/drinking establishments ranking is broadly consistent with the matched-pair audit evidence of Neumark, Bank and Van Nort (1996), who find that women are preferred at lower-price restaurants of the sort that are considered in our survey. Manufacturing, electric / gas utilities, and the auto sector are thought to be the most favorable to men. The gap in average Likert scores between the top and bottom firms is more than one point, while the gap in EB scores is roughly 27 percentage points. As they did for race, Borda and Likert scores largely agree on the industry ranking.

Average beliefs about age discrimination across industries show less dispersion than for race and gender but intuitive patterns. Sectors such as apparel, home furnishings, and eating and drinking are thought to favor younger workers, while banks and securities, wholesale trade, and accommodation / real estate are thought to favor older workers.

7 Comparing beliefs to an experiment

Our analysis thus far has established that respondents largely agree which firms discriminate on the basis of gender and, to a lesser extent, race. In this section, we examine

whether those beliefs align with experimental evidence on the conduct of particular firms. For this purpose, we use the large scale correspondence experiment described in Kline, Rose and Walters (2022), which sent up to 1,000 resumes to each of 108 Fortune 500 companies.

We will focus our attention on the 97 of those 108 companies studied in Kline, Rose and Walters (2024), which had enough employer responses to compute reliable measures of proportional contact gaps between applications with different signaled characteristics. For example, the proportional race contact gap measures the difference in log contact rates between applications with distinctively White and distinctively Black names. Kline, Rose and Walters (2024) document that roughly half of the variation in this proportional gaps is between industry.

Denoting the estimated gaps by $\hat{\theta}_j$, we consider linear models of the form

$$\hat{\theta}_j = \beta_0 + \beta_1 S_j + \eta_{s(j)} + \varepsilon_j$$

where S_j denotes either the Likert score L_j or the Borda score B_j, and $\eta_{s(j)}$ is an industry fixed effect for sector $s(j)$. The parameter β_1 can be interpreted as a semi-elasticity giving the percent change in contact gap expected from a unit increase in the relevant score. When industry fixed effects are included, β_1 measures alignment between beliefs and contact gaps within industry.

To estimate β_1, we replace the unknown Likert and Borda scores with their empirical proxies $\hat{L}_j$ and $\hat{B}_j$ and fit errors in variables (EIV) regressions (Fuller, 1987) using the variance estimators described in Section 4.3.1 to estimate the reliability of each proxy $\hat{S}_j$. One can think of these EIV regressions as jackknife IV estimators where a set of firm dummies are the instruments used to address the measurement error in $\hat{S}_j$ (Devereux, 2007). As Chen, Gu and Kwon (2025) emphasize, EIV regression estimates are robust to arbitrary correlation between the standard errors and the underlying signals.

The expert forecasting literature (DellaVigna and Pope, 2018; DellaVigna, Pope and

Vivalt, 2019) often finds that some individuals have a greater ability to forecast experimental outcomes than others. We explore heterogeneity by estimating β_1 in distinct subpopulations of respondents. Our use of the Katz (1961) estimator of variance $\dot{\sigma}_S^2$ ensures the bivariate EIV tests are always computable, even in subpopulations with low signal strength. However, in specifications with industry fixed effects, the EIV estimator cannot be computed when the implied within-industry signal variance is zero or negative. Since this only happens when the simpler bivariate regression is already highly imprecise, we interpret these subsamples as fundamentally lacking the precision to test for a predictive relationship between beliefs and the correspondence experiment.

7.1 Forecasting racial contact gaps

Table 5 shows results for Black-White contact gaps. For reference, the mean racial contact gap in the experiment is 9.5 log points (i.e., 0.095), indicating that distinctive White names were contacted about 10% more often than distinctively Black names. Columns 1 and 3 reveal that beliefs about both the contact and conduct versions of the racial discrimination question are insignificantly related to firm-level racial contact gaps. Pooling responses to the two questions in Column 5 to improve precision yields the same conclusion.

Controlling for industry fixed effects substantially increases the estimated EIV coefficients on both the contact and conduct based beliefs. Columns 2 and 6 reveal a statistically significant relationship in the full sample for both the Likert and Borda based measures of beliefs. By the Frisch-Waugh-Lovell theorem, the difference between the coefficients with industry fixed effects and those omitting them as controls is attributable to the between-industry relationship between beliefs and contact gaps. Evidently, the between-industry structure of beliefs about racial discrimination is inversely correlated with racial contact gaps in the experiment.

The positive within-industry relationship between beliefs about racial discrimination and racial contact gaps is particularly strong in the subsample that reports currently looking

Table 5: Experimental measures of race discrimination vs. beliefs

Sample	(1) Contact	(2) Contact (Industry FE)	(3) Conduct	(4) Conduct (Industry FE)	(5) Pooled	(6) Pooled (Industry FE)
Panel A: Likert						
Full Sample	0.004 (0.046)	0.162 (0.066)	0.120 (0.086)	1.018 (1.198)	0.023 (0.050)	0.195 (0.072)
Black	-0.004 (0.044)	0.122 (0.056)	0.019 (0.054)	0.309 (0.398)	-0.006 (0.043)	0.117 (0.070)
White	0.059 (0.093)	1.045 (1.498)	0.752 (1.011)	NA (NA)	0.150 (0.116)	1.094 (0.871)
Female	0.017 (0.042)	0.180 (0.069)	0.171 (0.100)	1.272 (2.958)	0.039 (0.046)	0.207 (0.082)
Male	-0.040 (0.117)	NA (NA)	0.042 (0.067)	0.392 (0.313)	-0.013 (0.084)	0.235 (0.138)
Looking for a Job	0.055 (0.041)	0.203 (0.091)	0.192 (0.136)	NA (NA)	0.076 (0.046)	0.280 (0.125)
Not Looking for a Job	-0.054 (0.068)	0.160 (0.120)	0.055 (0.104)	0.361 (0.391)	-0.028 (0.069)	0.132 (0.080)
Feared Discrimination	0.057 (0.054)	0.376 (0.235)	0.076 (0.190)	NA (NA)	0.053 (0.055)	0.325 (0.143)
Did Not Fear Discrimination	-0.031 (0.058)	0.140 (0.097)	0.123 (0.105)	1.475 (4.783)	-0.002 (0.061)	0.179 (0.095)
40 Years or Older	-0.009 (0.057)	0.132 (0.070)	0.050 (0.092)	0.222 (0.151)	0.004 (0.061)	0.145 (0.069)
Less than 40 Years Old	0.015 (0.048)	0.226 (0.129)	0.108 (0.070)	NA (NA)	0.047 (0.047)	0.324 (0.174)
Panel B: Borda						
Full Sample	0.028 (0.158)	0.565 (0.254)	-0.024 (0.280)	0.804 (0.741)	0.007 (0.172)	0.555 (0.267)
Black	-0.106 (0.172)	0.355 (0.287)	-0.110 (0.268)	10.903 (129.499)	-0.116 (0.181)	0.390 (0.333)
White	0.294 (0.184)	0.809 (0.302)	0.292 (0.421)	NA (NA)	0.297 (0.212)	1.072 (0.381)
Female	0.052 (0.139)	0.514 (0.228)	-0.011 (0.257)	0.584 (0.702)	0.028 (0.149)	0.538 (0.259)
Male	-0.028 (0.318)	1.860 (2.586)	-0.062 (0.316)	0.361 (1.063)	-0.051 (0.291)	0.647 (0.622)
Looking for a Job	0.231 (0.138)	1.006 (0.399)	0.080 (0.227)	3.620 (8.604)	0.206 (0.147)	1.255 (0.427)
Not Looking for a Job	-0.144 (0.184)	0.260 (0.294)	-0.286 (0.483)	-0.990 (6.404)	-0.159 (0.219)	0.204 (0.319)
Feared Discrimination	0.105 (0.195)	1.198 (0.701)	-0.253 (0.471)	NA (NA)	0.059 (0.235)	1.655 (0.939)
Did Not Fear Discrimination	-0.001 (0.169)	0.561 (0.277)	0.092 (0.244)	0.468 (0.517)	0.007 (0.174)	0.455 (0.257)
40 Years or Older	0.001 (0.167)	0.502 (0.268)	-0.144 (0.281)	-0.048 (0.776)	-0.052 (0.185)	0.367 (0.286)
Less than 40 Years Old	0.093 (0.182)	0.700 (0.377)	0.135 (0.233)	0.778 (0.738)	0.105 (0.189)	0.754 (0.367)

Notes: This table reports Errors-in-Variables (EIV) regressions of experimental log contact gaps (LHS) on Likert Scores (Panel A) and Borda Scores (Panel B). Columns (1)–(2) use responses to the contact question in the names arm recoded so that larger values indicate higher perceived contact for White relative to Black job seekers. Columns (3)–(4) use the question from the conduct arm. We restrict attention the version that asks about discrimination against Black job seekers because the opposite formulation exhibited low signal variance. Columns (1) and (3) include no additional controls; Columns (2) and (4) add two-digit industry fixed effects. The EIV treats the regressor as measured with error; the measurement-error variance supplied to the estimator equals the sample variance of the observed regressor minus the Katz (1961)-corrected estimate $\dot{\sigma}_S^2$ of its signal variance. Entries shown as NA indicate the EIV estimator could not be computed because the noise-corrected design matrix was ill-conditioned (nearly singular).

for a job. For example, when working with the contact version of beliefs, the coefficient $\hat{\beta}_1$ suggests that a 10 percentage point increase in the Borda win rate is associated with a 10% larger Black-White contact gap.

Focusing on column 6, which pools the conduct and contact questions, workers who are looking for a job, workers less 40 years old, and workers who have put off applying to a job out of fear of discrimination, all have beliefs more closely aligned with the audit study evidence. The elevated forecasting ability of this last subgroup is especially interesting, suggesting that those who fear discrimination may, in fact, have a better understanding of which employers discriminate.

7.2 Forecasting gender contact gaps

Table 6 reports the results for gender discrimination. For reference, the overall gender contact gap is -0.6 log points, which is statistically indistinguishable from zero. In contrast to the earlier findings for racial discrimination, the estimated value of β_1 for gender discrimination is distinguishable from zero in the full sample with both the Likert and Borda measures.

Table 6: Experimental measures of gender discrimination vs. beliefs

Sample	(1) Contact	(2) Contact (Industry FE)	(3) Conduct	(4) Conduct (Industry FE)	(5) Pooled	(6) Pooled (Industry FE)
Panel A: Likert						
Full Sample	0.111 (0.043)	0.108 (0.048)	0.113 (0.049)	0.082 (0.057)	0.116 (0.046)	0.102 (0.048)
Black	0.113 (0.058)	0.119 (0.066)	0.120 (0.055)	0.120 (0.078)	0.120 (0.055)	0.119 (0.060)
White	0.108 (0.034)	0.100 (0.046)	0.117 (0.050)	0.076 (0.073)	0.115 (0.040)	0.093 (0.048)
Female	0.097 (0.038)	0.096 (0.045)	0.103 (0.048)	0.105 (0.058)	0.105 (0.043)	0.106 (0.048)
Male	0.151 (0.064)	0.159 (0.099)	0.131 (0.052)	0.060 (0.093)	0.147 (0.054)	0.106 (0.067)
Looking for a Job	0.102 (0.046)	0.107 (0.053)	0.097 (0.042)	0.078 (0.069)	0.110 (0.044)	0.103 (0.053)
Not Looking for a Job	0.116 (0.043)	0.120 (0.057)	0.125 (0.060)	0.096 (0.067)	0.123 (0.049)	0.108 (0.052)
Feared Discrimination	0.098 (0.042)	0.086 (0.055)	0.071 (0.047)	0.002 (0.056)	0.089 (0.043)	0.043 (0.047)
Did Not Fear Discrimination	0.122 (0.046)	0.136 (0.057)	0.142 (0.056)	0.141 (0.079)	0.134 (0.050)	0.141 (0.059)
40 Years or Older	0.096 (0.041)	0.089 (0.047)	0.127 (0.050)	0.103 (0.059)	0.111 (0.044)	0.100 (0.044)
Less than 40 Years Old	0.136 (0.051)	0.154 (0.066)	0.088 (0.047)	0.060 (0.063)	0.120 (0.049)	0.113 (0.064)
Panel B: Borda						
Full Sample	0.435 (0.171)	0.564 (0.216)	0.381 (0.195)	0.350 (0.210)	0.411 (0.179)	0.467 (0.198)
Black	0.583 (0.240)	0.860 (0.326)	0.372 (0.207)	0.442 (0.245)	0.490 (0.216)	0.652 (0.250)
White	0.343 (0.130)	0.407 (0.200)	0.399 (0.194)	0.312 (0.243)	0.353 (0.153)	0.343 (0.188)
Female	0.406 (0.151)	0.596 (0.210)	0.366 (0.195)	0.363 (0.238)	0.394 (0.165)	0.496 (0.200)
Male	0.488 (0.246)	0.444 (0.370)	0.416 (0.214)	0.476 (0.285)	0.439 (0.221)	0.417 (0.256)
Looking for a Job	0.425 (0.177)	0.593 (0.234)	0.366 (0.176)	0.301 (0.272)	0.415 (0.177)	0.472 (0.219)
Not Looking for a Job	0.423 (0.165)	0.567 (0.233)	0.375 (0.218)	0.381 (0.237)	0.405 (0.183)	0.470 (0.205)
Feared Discrimination	0.507 (0.197)	0.740 (0.308)	0.290 (0.209)	0.232 (0.266)	0.419 (0.197)	0.453 (0.240)
Did Not Fear Discrimination	0.409 (0.168)	0.560 (0.232)	0.395 (0.187)	0.324 (0.211)	0.401 (0.173)	0.471 (0.199)
40 Years or Older	0.344 (0.140)	0.481 (0.184)	0.406 (0.201)	0.434 (0.213)	0.361 (0.156)	0.444 (0.172)
Less than 40 Years Old	0.574 (0.239)	0.816 (0.378)	0.320 (0.200)	0.199 (0.268)	0.460 (0.218)	0.499 (0.292)

Notes: This table reports Errors-in-Variables (EIV) regressions of experimental log contact gaps (LHS) on Likert Scores (Panel A) and Borda Scores (Panel B). Columns (1)–(2) use responses to the contact question in the names arm recoded so that larger values indicate higher perceived contact for male relative to female. Columns (3)–(4) use the question from the conduct arm, pooling both the against male and against female formulations after recoding appropriately. Columns (1) and (3) include no additional controls; Columns (2) and (4) add two-digit industry fixed effects. The EIV treats the regressor as measured with error; the measurement-error variance supplied to the estimator equals the sample variance of the observed regressor minus the Katz (1961)-corrected estimate $\dot{\sigma}_S^2$ of its signal variance. Standard errors are from bootstrapping the full procedure. Entries shown as NA indicate the EIV estimator could not be computed because the noise-corrected design matrix was ill-conditioned (nearly singular).

The estimated values of β_1 are remarkably insensitive to the inclusion of fixed effects. The general insensitivity of the EIV point estimates to industry fixed effects suggests that respondent beliefs predict both industry average gender contact gaps and firm-specific variation in contact gaps within industries.

No clear heterogeneity patterns in the alignment of beliefs with experimental sex gaps emerge across the demographic subgroups. With Borda scores, the beliefs of Black re-

spondents appear to have stronger correlation with the experimental sex gaps than the full sample; however, this pattern is not present in the Likert scores.

7.3 Perceived discretion, selectivity, and contact gaps

Section 5.2 established that perceptions of discrimination against Black and, to a lesser extent, female job seekers are strongly correlated with perceptions of employer selectivity and discretion. We now study the extent to which experimental contact gaps track these beliefs.

Table 7 shows that both perceived selectivity and perceived discretion seem to predict racial contact gaps but in directions opposite to those predicted by respondent beliefs about racial discrimination. Firms thought to be more selective have smaller racial contact gaps, while firms thought to exhibit the most hiring discretion exhibit substantially larger racial contact gaps. Controlling for industry fixed effects weakens the magnitude of the estimated relationships but leaves the sign of the estimated relationships unaltered. Evidently, a substantial amount of racial discrimination takes place at firms believed to be unselective.

Gender contact gaps are insignificantly related to beliefs about selectivity and discretion. The same is true conditional on industry fixed effects. This finding may be the result of the bi-directional nature of gender discrimination. The final two columns of Table 7 test this idea by studying the relationship between beliefs and squared experimental contact gaps (to capture departures from equal treatment in either direction). The results show no significant relationship for race, but stronger relationships for gender. Although the estimates remain insignificant, beliefs about selectivity are positively correlated with the magnitude of gender discrimination while beliefs about discretion are negatively correlated.

Table 7: Contact gaps, perceived discretion, and perceived selectivity

	Separate Models	Separate Models (Industry FEs)	Separate Models (Squared)	Separate Models (Squared, Ind FEs)
Panel A: Likert				
Race				
Selectivity	-0.049 (0.025)	-0.021 (0.038)		
Discretion	0.087 (0.031)	0.042 (0.102)		
Gender				
Selectivity	-0.015 (0.042)	-0.002 (0.061)	0.034 (0.019)	0.026 (0.038)
Discretion	-0.015 (0.056)	-0.156 (0.109)	-0.041 (0.032)	-0.032 (0.054)
Panel B: Borda				
Race				
Selectivity	-0.180 (0.097)	-0.074 (0.157)		
Discretion	0.329 (0.125)	0.143 (0.405)		
Gender				
Selectivity	-0.008 (0.175)	0.115 (0.276)	0.159 (0.088)	0.132 (0.187)
Discretion	-0.041 (0.239)	-0.498 (0.526)	-0.181 (0.118)	-0.155 (0.290)

Notes: This table reports Errors-in-Variables (EIV) regressions of contact gaps on Likert scores (Panel A) and firm-level Borda Scores (Panel B). The EIV treats the regressor as measured with error; the measurement-error variance supplied to the estimator equals the sample variance of the observed regressor minus the Katz (1961)-corrected estimate $\dot{\sigma}_S^2$ of its signal variance. Each entry in the table is the coefficient on either selectivity or discretion beliefs from separate bivariate regressions of either race or gender contact gaps on beliefs. The first two columns use log contact gaps as the outcome, while the final two columns use squared log contact gaps minus their squared standard errors as the outcome. Columns 2 and 4 include industry fixed effects.

8 Conclusion

Though firms devote considerable effort to cultivating favorable reputations, our analysis suggests the public believes many large employers engage in biased (and potentially illegal) recruiting practices. These beliefs are widely held but diffuse: average beliefs are similar across demographic groups, yet the expected agreement between two randomly selected respondents from the same demographic group is modest.

Average beliefs about gender discrimination align closely with gender contact gaps in a large-scale correspondence experiment. In this sense, gender discrimination appears to be an open secret. Beliefs about racial discrimination are more weakly related to racial contact gaps, approaching statistical significance only within industry. While recent research points to manager discretion as an important mediator of sex and race discrimination,

perceived discrimination is inversely related to average beliefs about manager discretion. Instead, respondents tend to believe the most selective firms are the most biased.

A large share of respondents report having put off applying for a job out of fear of discrimination. The average beliefs of these respondents are especially predictive of experimental racial contact gaps. However, the averages mask substantial dispersion in beliefs, which raises the possibility that views about bias are inaccurate in a substantial share of individual cases. Little is known about the psychic and economic costs of mistakenly inferring bias where it does not exist, or of failing to recognize bias where it does exist. Quantifying these costs is an important priority for future research.

References

Abdulkadiroğlu, Atila, Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters. 2020. "Do parents value school effectiveness?" *American Economic Review*, 110(5): 1502–1539.

Alesina, Alberto, Matteo F Ferroni, and Stefanie Stantcheva. 2021. "Perceptions of racial gaps, their causes, and ways to reduce them." National Bureau of Economic Research.

Allison, Paul D, and Nicholas A Christakis. 1994. "Logit models for sets of ranked items." *Sociological methodology*, 199–228.

Ballinger, T Parker, and Nathaniel T Wilcox. 1997. "Decisions, error and heterogeneity." *The Economic Journal*, 107(443): 1090–1105.

Becker, Gary S. 1957. *The Economics of Discrimination.* University of Chicago Press.

Beggs, Steven, Scott Cardell, and Jerry Hausman. 1981. "Assessing the potential demand for electric cars." *Journal of econometrics*, 17(1): 1–19.

Black, Dan A. 1995. "Discrimination in an equilibrium search model." *Journal of labor Economics*, 13(2): 309–334.

Borda, JC de. 1784. "Mémoire sur les élections au scrutin." *Histoire de l'Academie Royale des Sciences pour 1781 (Paris, 1784).*

Bowlus, Audra J, and Zvi Eckstein. 2002. "Discrimination and skill differences in an equilibrium search model." *International Economic Review*, 43(4): 1309–1345.

Caldwell, Sydnee, Ingrid Haegele, and Jörg Heining. 2025. "Bargaining and inequality in the labor market." National Bureau of Economic Research.

Chen, Jiafeng, Jiaying Gu, and Soonwoo Kwon. 2025. "Empirical Bayes shrinkage (mostly) does not correct the measurement error in regression." *arXiv preprint arXiv:2503.19095.*

Davidson, Donald, and Jacob Marschak. 1959. "Experimental tests of a stochastic decision theory." *Measurement: Definitions and theories*, 17(2).

DellaVigna, Stefano, and Devin Pope. 2018. "Predicting experimental results: who knows what?" *Journal of Political Economy*, 126(6): 2410–2456.

DellaVigna, Stefano, Devin Pope, and Eva Vivalt. 2019. "Predict science to improve science." *Science*, 366(6464): 428–429.

Devereux, Paul J. 2007. "Improved errors-in-variables estimators for grouped data." *Journal of Business & Economic Statistics*, 25(3): 278–287.

Fuller, Wayne A. 1987. "Measurement Error Models." *Wiley Series in Probability and Statistics*.

Haaland, Ingar, and Christopher Roth. 2023. "Beliefs about racial discrimination and support for pro-black policies." *Review of Economics and Statistics*, 105(1): 40–53.

Hausman, Jerry A., and Paul A. Ruud. 1987. "Specifying and testing econometric models for rank-ordered data." *Journal of Econometrics*, 34(1): 83–104.

Hoffman, Mitchell, Lisa B Kahn, and Danielle Li. 2018. "Discretion in hiring." *The Quarterly Journal of Economics*, 133(2): 765–800.

Katz, Morris W. 1961. "Admissible and minimax estimates of parameters in truncated spaces." *The Annals of Mathematical Statistics*, 32(1): 136–142.

Kline, Patrick, Evan K Rose, and Christopher R Walters. 2022. "Systemic discrimination among large US employers." *The Quarterly Journal of Economics*, 137(4): 1963–2036.

Kline, Patrick, Evan K Rose, and Christopher R Walters. 2024. "A Discrimination Report Card." *American Economic Review*, 114(8): 2472–2525.

Korba, Anna, Stephan Clémençon, and Eric Sibony. 2017. "A learning theory of ranking aggregation." 1001–1010, PMLR.

Likert, Rensis. 1932. "A technique for the measurement of attitudes." *Archives of Psychology*, 22(140): 5–55.

Miller, Conrad. 2017. "The persistent effect of temporary affirmative action." *American Economic Journal: Applied Economics*, 9(3): 152–90.

Mocanu, Tatiana. 2022. "Designing gender equity: Evidence from hiring practices and committees." Working paper.

Neumark, David, Roy J Bank, and Kyle D Van Nort. 1996. "Sex discrimination in restaurant hiring: An audit study." *The Quarterly journal of economics*, 111(3): 915–941.

Portnoy, Stephen. 1982. "Maximizing the probability of correctly ordering random variables using linear predictors." *Journal of Multivariate Analysis*, 12(2): 256–269.

Shah, Nihar B, and Martin J Wainwright. 2018. "Simple, robust and optimal ranking from pairwise comparisons." *Journal of machine learning research*, 18(199): 1–38.

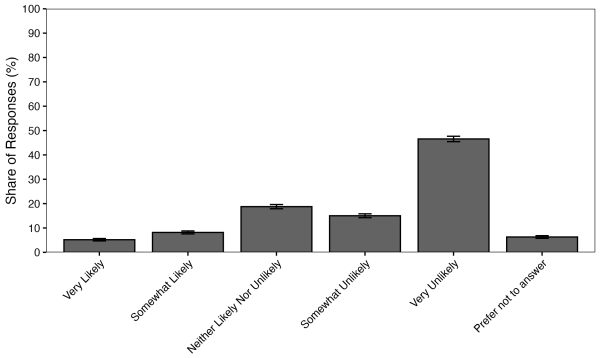
U.S. Census Bureau. 2025. "PINC-11. Income Distribution to $250,000 or More for Males and Females: 2024." *Current Population Survey, 2025 Annual Social and Economic (ASEC) Supplement*, Table PINC-11 (Person income in 2024, CPS 2025 ASEC).

Walters, Christopher. 2024. "Empirical Bayes methods in labor economics." In *Handbook of Labor Economics*. Vol. 5, 183–260. Elsevier.

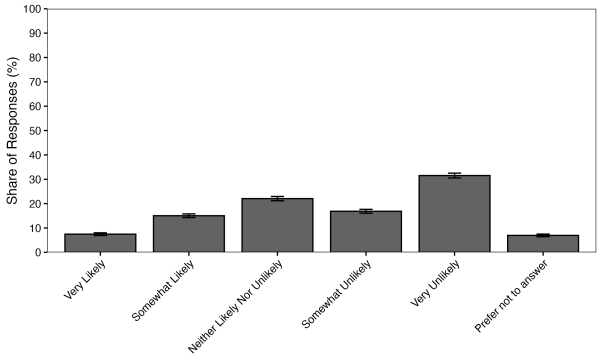
Appendix A: Additional figures

Figure A1: Distribution of perceived gender discrimination likelihood

(a) Against Male



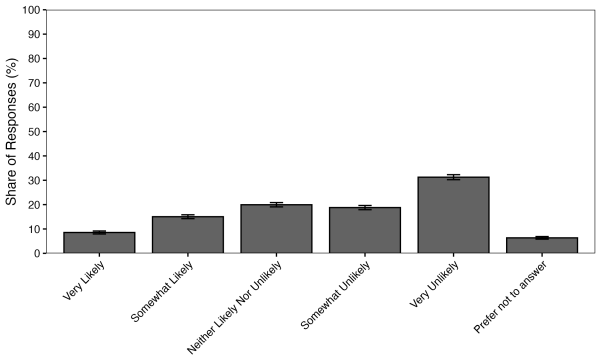
(b) Against Female



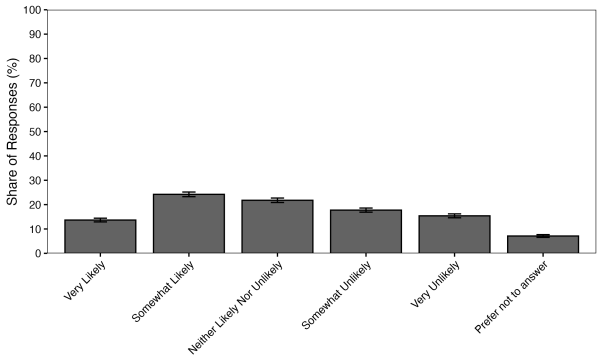
Notes: This figure reports the distribution of companies' perceived likelihood of discriminating against male job seekers (panel A) and against female job seekers (panel B). Only respondents in the conduct arm were shown this question, with the against female vs. against male framing chosen at random. Question text: Please indicate how likely you think it is that each company below would discriminate against [male / female] job-seekers.

Figure A2: Distribution of perceived age discrimination likelihood

(a) Against Younger



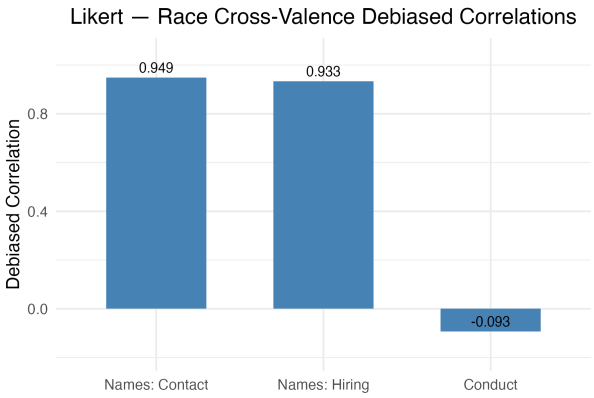
(b) Against Older



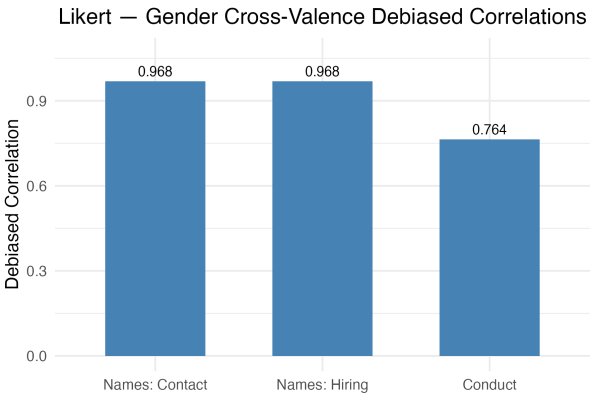
Notes: This figure reports the distribution of companies' perceived likelihood of discriminating against younger job seekers (panel A) and against older job seekers (panel B). Only respondents in the conduct arm were shown this question, with the against younger vs. against older framing chosen at random. Question text: Please indicate how likely you think it is that each company below would discriminate against [younger / older] job-seekers.

Figure A3: Valence Correlations

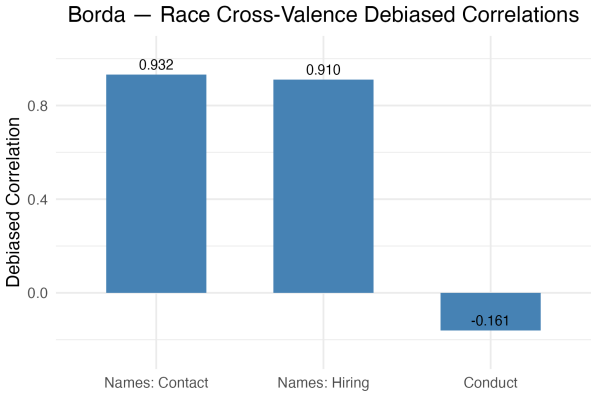
(a) Likert, Race



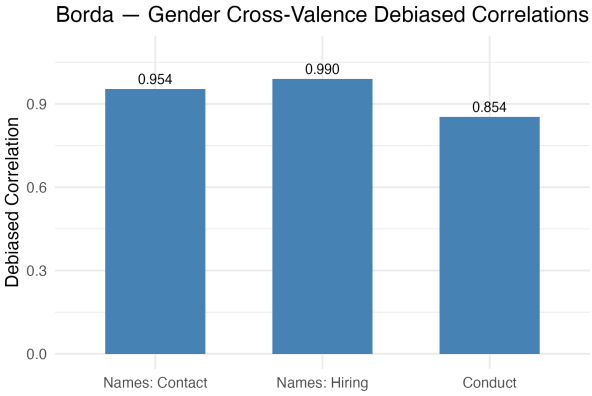
(b) Likert, Gender



(c) Borda, Race



(d) Borda, Gender



Notes: These figures report debiased correlation in Likert (Panels A and b) and Borda (Panels C and D) scores across randomized framings of the contact, hiring, and conduct questions. Correlations are computed by estimating firm-level scores for each framing separately and dividing their sample correlation by the square root of the product of debiased variances. Scores are reversed in one framing, so that a correlation of 1 indicates prefect alignment of firm-level beliefs across the two framings.

Appendix B: Additional tables

Table A1: Alternative Framings Signal Strength

	Likert			Borda			
Outcome	Std Dev	Signal Std Dev	T-stat no signal	Std Dev	Signal Std Dev	T-stat no signal	Normed Variance
Discrimination Black (Conduct)	0.312	0.233	5.829	0.080	0.064	6.968	0.048
Discrimination Black (Contact)	0.336	0.296	10.643	0.102	0.093	13.158	0.102
Discrimination Black (Hire)	0.345	0.304	10.573	0.101	0.092	12.646	0.100
Discrimination Black - Black Wording (Conduct)	0.437	0.362	8.740	0.119	0.102	10.211	0.124
Discrimination Black - Black Wording (Contact)	0.428	0.367	9.962	0.122	0.107	11.434	0.137
Discrimination Black - Black Wording (Hire)	0.454	0.390	10.326	0.124	0.109	11.655	0.141
Discrimination Black - White Wording (Conduct)	0.347	0.101	0.493	0.086	0.030	1.442	0.011
Discrimination Black - White Wording (Contact)	0.310	0.219	5.668	0.099	0.078	7.719	0.073
Discrimination Black - White Wording (Hire)	0.307	0.218	5.329	0.098	0.077	7.325	0.071
Discrimination Female (Conduct)	0.486	0.449	14.251	0.124	0.117	17.009	0.162
Discrimination Female (Contact)	0.483	0.460	16.863	0.127	0.121	18.496	0.172
Discrimination Female (Hire)	0.512	0.490	18.321	0.131	0.125	19.109	0.185
Discrimination Female - Female Wording (Conduct)	0.606	0.563	14.571	0.144	0.133	15.080	0.209
Discrimination Female - Female Wording (Contact)	0.544	0.502	13.116	0.140	0.128	14.496	0.194
Discrimination Female - Female Wording (Hire)	0.566	0.524	14.083	0.143	0.132	14.846	0.207
Discrimination Female - Male Wording (Conduct)	0.450	0.359	7.921	0.123	0.105	10.043	0.131
Discrimination Female - Male Wording (Contact)	0.466	0.424	11.876	0.129	0.117	12.694	0.162
Discrimination Female - Male Wording (Hire)	0.506	0.465	12.640	0.131	0.119	12.684	0.167
Discrimination Older (Conduct)	0.277	0.219	6.859	0.075	0.061	8.206	0.044
Discrimination Older - Older Wording (Conduct)	0.374	0.298	7.940	0.099	0.081	9.008	0.077
Discrimination Older - Younger Wording (Conduct)	0.487	0.422	9.809	0.132	0.118	12.784	0.166

Notes: This table reports estimates of the dispersion of Likert and Borda scores for each framing of key questions in the names and conduct arms. "Black wording" indicates that Black workers were the disfavored group in the question, and analogously for "White wording," "Female wording," etc. Columns labeled "Std Dev" report the square root of the sample variance of estimates. "Signal Std Dev" reports the square root of asymptotically unbiased estimates of the variances of scores, or $\dot{\sigma}_L$ and $\dot{\sigma}_B$, respectively. "T-stat no signal" reports $\hat{\sigma}^2/\mathbb{V}[\hat{\sigma}^2]$, which allows a t-test of the null hypothesis that the signal standard deviation is zero. "Normed variance" reports the ratio of the Borda signal variance to its theoretical maximum, which occurs when all respondents share a common strict preference ordering.

Table A2: Within- vs. between industry Variance of firm-level beliefs

	Likert			Borda			
Outcome	Std Dev	Signal Std Dev	T-stat no signal	Std Dev	Signal Std Dev	T-stat no signal	Normed Variance
Panel A: Within-industry							
Discrimination Black (Contact)	0.239	0.186	6.588	0.073	0.061	8.404	0.044
Discrimination Black (Hire)	0.251	0.198	6.692	0.074	0.062	8.452	0.046
Discrimination Black (Conduct)	0.246	0.150	3.487	0.061	0.040	4.013	0.018
Discrimination Female (Contact)	0.368	0.340	13.095	0.096	0.089	13.861	0.093
Discrimination Female (Hire)	0.379	0.351	13.362	0.099	0.091	14.025	0.099
Discrimination Female (Conduct)	0.376	0.333	10.861	0.094	0.085	12.507	0.085
Discrimination Older (Conduct)	0.202	0.124	3.619	0.057	0.040	5.204	0.019
Manager Discretion	0.247	0.222	10.723	0.061	0.055	11.213	0.035
Firm Selectivity	0.350	0.338	20.903	0.083	0.079	19.658	0.075
Firm Desirability	0.152	0.109	5.043	0.043	0.033	6.303	0.013
Discrimination Black (Pooled)	0.210	0.172	7.638	0.061	0.053	9.531	0.033
Discrimination Female (Pooled)	0.351	0.333	17.116	0.090	0.086	18.431	0.087
Panel B: Between-industry							
Discrimination Black (Contact)	0.237	0.235	1.669	0.071	0.071	1.827	0.059
Discrimination Black (Hire)	0.237	0.238	1.503	0.069	0.070	1.563	0.058
Discrimination Black (Conduct)	0.192	0.218	0.706	0.052	0.058	0.849	0.040
Discrimination Female (Contact)	0.313	0.309	2.226	0.083	0.082	2.153	0.080
Discrimination Female (Hire)	0.345	0.342	2.360	0.087	0.086	2.158	0.087
Discrimination Female (Conduct)	0.307	0.311	1.542	0.082	0.081	1.883	0.078
Discrimination Older (Conduct)	0.190	0.210	0.851	0.048	0.053	0.883	0.033
Manager Discretion	0.399	0.397	3.327	0.096	0.095	3.252	0.107
Firm Selectivity	0.532	0.531	5.799	0.135	0.135	5.782	0.215
Firm Desirability	0.101	0.131	0.483	0.028	0.036	0.480	0.016
Discrimination Black (Pooled)	0.214	0.213	1.763	0.062	0.062	1.951	0.045
Discrimination Female (Pooled)	0.308	0.305	2.643	0.082	0.081	2.858	0.078

Notes: This table reports estimates of the dispersion of Likert and Borda scores within- and between-industry. The outcomes are defined as in Table 3. Panel A presents within-industry variation computed by subtracting industry-average Likert or Borda scores from each firm's score, while Panel B studies the dispersion in industry averages. The columns are the same as in Table 3.

Table A3: Cross Sample Raw Correlations

	Discrimination Black	Discrimination Female
Panel A: Likert (Raw Correlation)		
Black vs White	0.615	0.874
Female vs Male	0.659	0.828
Looking for a Job vs Not	0.704	0.874
Feared Discrimination vs Not	0.688	0.870
Age >= 40 vs < 40	0.692	0.852
At Least Some College vs HS Diploma or less	0.587	0.799
Convenience vs Probability	0.636	0.884
Confident vs Not (Gender)	0.684	0.873
Confident vs Not (Race)	0.681	0.842
Panel B: Borda (Raw Correlation)		
Black vs White	0.721	0.885
Female vs Male	0.708	0.836
Looking for a Job vs Not	0.766	0.880
Feared Discrimination vs Not	0.757	0.877
Age >= 40 vs < 40	0.671	0.882
At Least Some College vs HS Diploma or less	0.698	0.817
Convenience vs Probability	0.766	0.872
Confident vs Not (Gender)	0.759	0.872
Confident vs Not (Race)	0.764	0.858

Notes: This table reports the raw correlation in estimated beliefs across sample groups without adjustment for over-dispersion due to sampling error. Columns labeled "Corr" give the estimated raw correlation in firm-level parameters across the two groups. Columns labeled "p-value" evaluate the null hypothesis that the two vectors of firm-level parameters are equal via a likelihood ratio test. Panel (a) uses Likert scores, while panel (b) uses Borda scores.

Appendix C: Context-dependent preferences and SST

A simple way to represent non-IIA preferences is to work with a context-dependent rank order logit model in which Likert scores are allowed to depend on the choice set C_i. This yields choice probabilities of the form:

$$\Pr(j \succ_i k \mid C_i) = \frac{\exp(L_{j,C_i})}{\exp(L_{j,C_i}) + \exp(L_{k,C_i})}.$$

To obtain a tractable class of such models, suppose the menu-dependent worths admit additive item-specific biases,

$$L_{j,C} = L_j + b_j(C),$$

where L_j is the intrinsic worth and $b_j(C)$ captures context effects. Define the menu-specific worth gap

$$\Delta_{jk}(C) = L_{j,C} - L_{k,C}.$$

The following proposition establishes a sufficient condition for this model to retain the transitivity properties of the standard Plackett-Luce model despite the violation of IIA.

Proposition 2 (Sufficient Condition for SST)**.** Let C_{xy} denote any menu containing x and y. The context-dependent PL model satisfies Strong Stochastic Transitivity if, for every triple of firms (j, k, l) and every collection of menus C_{jk}, C_{kl}, C_{jl}, the following implication holds:

$$\Delta_{jk}(C_{jk}) \geq 0 \quad \text{and} \quad \Delta_{kl}(C_{kl}) \geq 0 \quad \Longrightarrow \quad \Delta_{jl}(C_{jl}) \;\geq\; \max\big\{\Delta_{jk}(C_{jk}),\; \Delta_{kl}(C_{kl})\big\}.$$

Proof. The choice probability $\Pr(j \succ_i k \mid C_i)$ is a strictly increasing function of the worth gap $\Delta_{jk}(C_i)$. Consequently, the probabilistic inequalities defining SST map directly to the algebraic inequalities on the utility gaps stated above. □

In the standard rank-ordered logit model, where $b_j(C) = 0$, this condition holds automatically because $\Delta_{jl} = \Delta_{jk} + \Delta_{kl}$, and the sum of two non-negative numbers exceeds their maximum. Thus, in this context-dependent model, SST holds provided the context effects $b_j(C)$ are not sufficiently heterogeneous across menus to disrupt the additivity of the utility gaps.

Appendix D: Proof of Proposition 1

We analyze the properties of the *population* Borda score B_j under the "Common-Opponent" definition provided in the text. Recall that $\mathcal{N}$ denotes the set of non-anchor firms and the score is defined as:

$$B_j = \frac{1}{|\mathcal{N}|}\left[\tfrac{1}{2}\,\mathbf{1}\{j \in \mathcal{N}\} + \sum_{\ell \in \mathcal{N}\setminus\{j\}} w_{j\ell}\right],$$

where $w_{j\ell} = p_{j\ell} + \frac{1}{2}t_{j\ell}$ is the expected pairwise score. For anchors, the $\frac{1}{2}$ term vanishes and the score reduces to the average pairwise score against all non-anchors. For non-anchors, the additional $\frac{1}{2}$ can be interpreted as a phantom self-match scored as a tie, ensuring that all firms' scores are averages over exactly $|\mathcal{N}|$ terms.

Proof. Let $\succ^*$ denote the strict total order implied by SST, such that $j \succ^* k \iff p_{jk} > 1/2$. We show that $B_j > B_k$ for any distinct pair j, k, regardless of their class assignment.

Recall that under SST, if $j \succ^* k$, then for any third firm ℓ:

- If $\ell \succ^* j \succ^* k$, then $p_{\ell k} \geq p_{\ell j} \implies w_{j\ell} \geq w_{k\ell}$.
- If $j \succ^* k \succ^* \ell$, then $p_{j\ell} \geq p_{k\ell} \implies w_{j\ell} \geq w_{k\ell}$.
- If $j \succ^* \ell \succ^* k$, then $w_{j\ell} > 0.5 > w_{k\ell}$.

Thus, $j \succ^* k$ implies $w_{j\ell} \geq w_{k\ell}$ for all $\ell \neq j, k$.

Since all scores share the common denominator $|\mathcal{N}|$, it suffices to compare numerators. We consider the three possible class configurations for the pair j, k.

Case 1a: Both are Anchors ($j, k \in \mathcal{A}$). The phantom terms vanish for both firms, so

$$B_j - B_k = \frac{1}{|\mathcal{N}|}\sum_{\ell \in \mathcal{N}}(w_{j\ell} - w_{k\ell}).$$

Since $j \succ^* k$, SST implies $w_{j\ell} \geq w_{k\ell}$ for all $\ell \in \mathcal{N}$. The no-knife-edge assumption ensures

that the inequality is strict for at least one ℓ. Thus $B_j > B_k$.

Case 1b: Both are Non-Anchors ($j, k \in \mathcal{N}$). The phantom $\frac{1}{2}$ terms cancel in the difference:

$$B_j - B_k = \frac{1}{|\mathcal{N}|}\left[(w_{jk} - w_{kj}) + \sum_{\ell \in \mathcal{N}\setminus\{j,k\}} (w_{j\ell} - w_{k\ell})\right].$$

The direct term $w_{jk} - w_{kj} = p_{jk} - p_{kj}$ is strictly positive because $j \succ^* k$. The indirect terms are non-negative by SST. Thus $B_j > B_k$.

Case 2: j is an Anchor, k is a Non-Anchor ($j \in \mathcal{A}, k \in \mathcal{N}$). The phantom term is present only for k:

$$B_j - B_k = \frac{1}{|\mathcal{N}|}\left[\left(w_{jk} - \tfrac{1}{2}\right) + \sum_{\ell \in \mathcal{N}\setminus\{k\}} (w_{j\ell} - w_{k\ell})\right].$$

Since $j \succ^* k$, we have $w_{jk} > 1/2$, so the first term is strictly positive. Every term in the sum is non-negative by SST. Thus $B_j > B_k$.

Case 3: j is a Non-Anchor, k is an Anchor ($j \in \mathcal{N}, k \in \mathcal{A}$). The phantom term is present only for j:

$$B_j - B_k = \frac{1}{|\mathcal{N}|}\left[\left(\tfrac{1}{2} - w_{kj}\right) + \sum_{\ell \in \mathcal{N}\setminus\{j\}} (w_{j\ell} - w_{k\ell})\right].$$

Since $j \succ^* k$, we have $w_{kj} < 1/2$, so the first term is strictly positive. Every term in the sum is non-negative by SST. Thus $B_j > B_k$.

In all cases, $j \succ^* k \implies B_j > B_k$. □

Appendix E: Estimating $\mathbb{V}[\hat{\sigma}_L^2]$ and $\mathbb{V}[\hat{\sigma}_B^2]$

Let $\hat{L} = \left(\hat{L}_1 - \bar{L}, \ldots, \hat{L}_J - \bar{L}\right)'$ and $\hat{B} = \left(\hat{B}_1 - \bar{B}, \ldots, \hat{B}_J - \bar{B}\right)'$. Letting $\mathbf{1}$ denote a $J \times 1$ vector of ones, these definitions ensure $\mathbf{1}'\hat{L} = \mathbf{1}'\hat{B} = 0$. Denote the associated variance matrices of these vectors by $\Sigma_L = \mathbb{V}[\hat{L}]$ and $\Sigma_B = \mathbb{V}[\hat{B}]$. We estimate these matrices analytically and will treat them as known.

Under the maintained assumption that the Likert and Borda estimates have reached their asymptotic normal limit, we have

$$\hat{L} \sim N(L, \Sigma_L), \quad \hat{B} \sim N(B, \Sigma_B),$$

where $L = (L_1 - \mu_L, \ldots, L_J - \mu_L)'$ and $B = (B_1 - \mu_B, \ldots, B_J - \mu_B)'$.

In this notation, we can write the variance component estimators

$$\hat{\sigma}_L^2 = \frac{\hat{L}'\hat{L} - \operatorname{tr}(\Sigma_L)}{J}, \quad \hat{\sigma}_B^2 = \frac{\hat{B}'\hat{B} - \operatorname{tr}(\Sigma_B)}{J}.$$

Recall that if $X \sim N(\mu, \Sigma)$ then $\mathbb{V}[X'X] = 2\operatorname{tr}(\Sigma^2) + 4\mu'\Sigma\mu$. We therefore use the following unbiased estimators of $\mathbb{V}[\hat{\sigma}_L^2]$ and $\mathbb{V}[\hat{\sigma}_B^2]$:

$$\widehat{\mathbb{V}}[\hat{\sigma}_L^2] = \frac{2\operatorname{tr}(\Sigma_L^2) + 4\left(\hat{L}'\Sigma_L\hat{L} - \operatorname{tr}(\Sigma_L^2)\right)}{J^2} = \frac{4\hat{L}'\Sigma_L\hat{L} - 2\operatorname{tr}(\Sigma_L^2)}{J^2}$$
$$\widehat{\mathbb{V}}[\hat{\sigma}_B^2] = \frac{2\operatorname{tr}(\Sigma_B^2) + 4\left(\hat{B}'\Sigma_B\hat{B} - \operatorname{tr}(\Sigma_B^2)\right)}{J^2} = \frac{4\hat{B}'\Sigma_B\hat{B} - 2\operatorname{tr}(\Sigma_B^2)}{J^2}.$$