SOCIAL SCIENCES

Gender composition predicts gender bias: A meta-reanalysis of hiring discrimination audit experiments

Diana Roxana Galos[1]*† and Alexander Coppock[2]†

Since 1983, more than 70 employment audit experiments, carried out in more than 26 countries across five continents, have randomized the gender of fictitious applicants to measure the extent of hiring discrimination on the basis of gender. The results are mixed: Some studies find discrimination against men, and others find discrimination against women. We reconcile these heterogeneous findings through a "meta-reanalysis" of the average effects of being described as a woman (versus a man), conditional on occupation. We find a strongly positive gender gradient. In (relatively better paying) occupations dominated by men, the effect of being a woman is negative, while in the (relatively lower paying) occupations dominated by women, the effect is positive. In this way, heterogeneous employment discrimination on the basis of gender preserves status quo gender distributions and earnings gaps. These patterns hold among both minority and majority status applicants.

INTRODUCTION

The audit experiment (or correspondence study) has been used to study many forms of discrimination, including on the basis of race and ethnicity (*1*), religion (*2*), sexual orientation (*3*), and social class (*4*), across several domains, including housing (*5*), education (*6*), and employment (*7*). As applied in employment settings, the main virtue of audit experiments (relative to other study designs) is that they allow us to study the effects of applicant characteristics on hiring free of the selection concerns that different sorts of people apply to different jobs at different rates. Because audit experiments randomly assign candidate characteristics, they provide design-based assurances of unbiased discrimination estimates.

The first study to deploy the audit experiment design for the study of discrimination on the basis of gender appears to be (*8*), who sent 1982 resumes of fictitious men and women applicants to 991 job openings in Victoria, Australia in 1986 and found a −2.4 percentage point effect of being a woman on receiving an invitation to interview (cluster-robust SE: 1.2 points). A second early study in this line (*9*) randomized the gender of 130 applicants for 65 restaurant jobs in New York City in 1994, finding a large and negative but nonsignificant effect (−12.3 points, SE: 8.4 points) of being a woman. In the following decades, more than 70 total employment audit studies that randomize applicant gender have been conducted. We have obtained or reconstructed the data for a total of 57 such studies, reported in 51 papers. The goal of this article is to synthesize what we have learned from the repeated application of the audit experiment research design about employment discrimination on the basis of gender.

The average treatment effect estimates yielded by these studies are mixed. In contrast to the early results, two-thirds (37 of 57) show a positive effect of being a woman, and the remainder show a negative effect. A standard random-effects meta-analysis of the average treatment effect estimates of these studies yields an estimate of 1.2 percentage points with an SE of 0.4 points. In fig. S1.1, we formally present this analysis, which is in line with previous meta-analyses of gender discrimination audit experiments (*10*, *11*). In our view, this standard meta-analytic estimate holds only marginal scientific interest because study-level average treatment effects are sometimes positive and sometimes negative. The standard approach simply averages over this heterogeneity. Which side of zero the meta-analytic average ends up on therefore depends uncomfortably on the set of contexts social scientists happened to select when studying employment discrimination on the basis of gender.

A theoretically key contextual variable is gender composition, parameterized here as the fraction of employees in an occupational setting who are women. Gender composition is mentioned in most theoretical treatments of how effects ought to vary [including in the very first such study (*8*) and in many subsequent efforts, e.g., (*12*–*14*)] and is incorporated into the study design of many of the experiments that we collected. Unfortunately, the statistical power of any one study to distinguish the effects in men-dominated contexts from those in women-dominated contexts is usually quite low. Some studies find evidence of heterogeneity by context (*13*–*15*), while others do not (*16*). As a result of these within-study sample size limitations, the broadly-held intuition that gender composition ought to predict gender bias has not received the systematic, synthetic investigation that it deserves.

Here, we offer conclusive confirmation of this intuition using a research design that we call a "meta-reanalysis." Our approach involves three steps: First, we reanalyze each experiment to yield conditional average treatment effect (CATE) estimates by occupation. Next, we obtain occupation-level gender composition data from the International Labor Organization (ILO). Third, we meta-analyze the occupation-level CATEs according to gender composition. Our meta-reanalysis reveals a clear mirroring of the status quo. In women-dominated occupations, the effects are positive. In men-dominated occupations, the effects are negative. Seen in one light, discrimination is even-handed: Yes, men are advantaged in men-dominated settings, but women are advantaged in women-dominated settings. However, this result does not imply that women are on equal employment footing with men, as average salaries

[1]Centre for the Experimental-Philosophical Study of Discrimination (CEPDISC), Department of Political Science, Aarhus University, Aarhus, Denmark. [2]Department of Political Science, Yale University, New Haven, CT, USA.
*Corresponding author. Email: diana.galos@ps.au.dk
†These authors contributed equally to this work.

are lower in women-dominated industries compared with men-dominated-industries (*17*–*19*).

Theory

Like others before us, we conceive of employment discrimination based on gender as the causal effect of learning that an applicant is a woman versus learning that an applicant is a man on hiring decisions. In potential outcomes notation, $Y_i(1)$ is the decision rendered by a hiring process for job opening i after learning that an applicant is a woman and $Y_i(0)$ is the decision after learning that the applicant is a man. Pro-woman discrimination occurs if $Y_i(1) = 1$ and $Y_i(0) = 0$ and pro-man discrimination occurs if $Y_i(1) = 0$ and $Y_i(0) = 1$. If $Y_i(1)$ and $Y_i(0)$ are equal, then no discrimination occurs.

Audit experiments provide estimates of net gender bias in a sample of N job openings: $\frac{\sum_i^N Y_i(1) - Y_i(0)}{N}$. Gender bias equals the rate of pro-woman discrimination minus the rate of pro-man discrimination.

With others, we theorize that the sign of this average causal effect will depend on context (*8*, *12*–*14*). To distinguish among contexts, we seek a measure that will allow us to predict when the CATE of being a woman (versus a man) on hiring will be positive and when it will be negative. We will use the term gender composition to refer to the status quo share of workers in a context who are women. In our empirical section, we will operationalize "context" as an occupation in a country and year. Among occupations that are fully men dominated, gender composition is 0, and among occupations that are fully women-dominated, gender composition is 1.

We expect gender composition to predict the sign and magnitude of gender bias, precisely because the existing gender composition is a product of the unobserved forces that lead to discrimination. To be clear, we do not assign a causal role to gender composition per se. Instead, we conceive of the existing fraction of women employed in an occupation as being the result of the same causal forces that act on applicants. Pro–status quo forces reproduce and maintain the status quo by discriminating against women in men-dominated contexts and by discriminating against men in women-dominated contexts. We therefore think of gender composition as a correlate (but not necessarily a cause) of the conditional effects of being a woman on hiring decisions. We refer to this correlation as the gender gradient. Like others (*8*, *12*–*14*), we expect the gender gradient to be positive: positive CATEs in women-dominated contexts and negative CATEs in men-dominated contexts. We grant that gender composition could also, in principle, exert its own independent causal effect on gender bias, although such an effect would be challenging to estimate as gender composition and gender bias likely share common unobserved causes.

But what about attempts to correct gender imbalances? Some employers self-consciously adopt hiring policies and practices that aim to bring the gender composition of a firm (or subdivision of a firm) closer to 50%. These anti–status quo forces could include affirmative action policies; gender quotas; diversity, equity, and inclusion initiatives; or empowerment-based confrontation programs (*20*). Some experimental studies show that such attempts can be successful in decreasing bias in men-dominated fields (*21*, *22*) and improving the representation of women in elected offices (*23*). If anti–status quo forces were to dominate at every level of gender composition, we would expect the gender gradient to be negative. Despite the increased prevalence of these initiatives, however, descriptive studies show that they are still a relatively long way from achieving their goals (*24*–*26*).

Furthermore, attempts to correct gender imbalances have (often appropriately) been asymmetric, focusing more on increasing women's representation in men-dominated fields than on increasing men's representation in women-dominated fields (*27*, *28*). This is not to say that men are disadvantaged upon entering women-dominated occupations. Observational studies suggest that once entering women-dominated occupations, men do not seem to be disadvantaged in earnings (*29*, *30*), and they are rewarded for spells in women-dominated occupations when subsequently transitioning into men-dominated occupations (*30*).

Figure 1 displays two ways the CATEs of being a woman (versus a man) on hiring decisions might depend on gender composition. In the left panel, we see that if pro–status quo forces dominate, the gender gradient will be positive: pro-woman bias in women-dominated occupations and pro-man bias in men-dominated

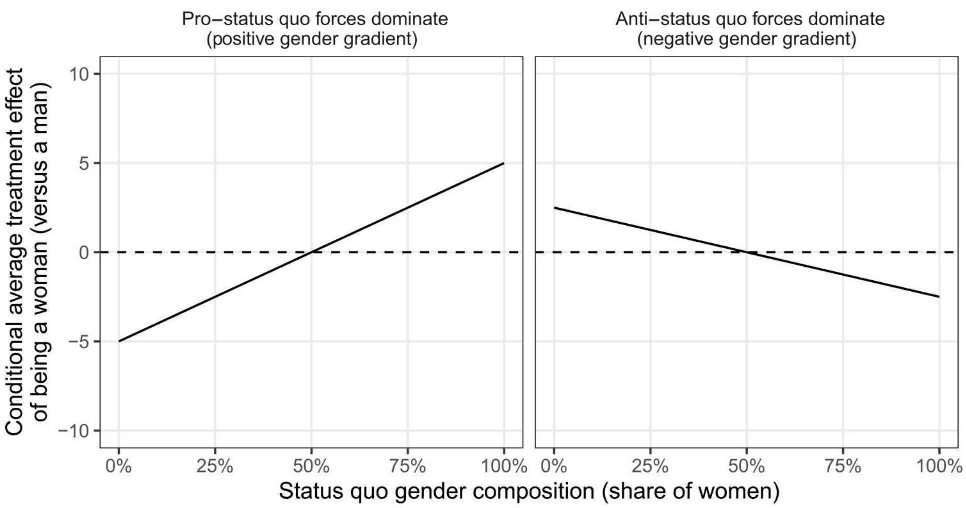

Fig. 1. Gender gradient (how the CATEs of being a woman depend on gender composition), depending on whether pro– or anti–status quo forces dominate.

occupations. In the right panel, we see that if anti–status quo forces dominate, the gender gradient will be negative: Hiring processes will work to correct the status quo gender composition by preferentially hiring more women in men-dominated occupations and more men in women-dominated occupations. We note that these are just two possible gender gradients; others, including nonlinear gradients, are possible as well.

Which scenario, a positive or a negative gender gradient, is more likely? All previous theoretical predictions have pointed to a positive gender gradient. As the foregoing review suggests, the attempts to address workplace gender imbalances have been uneven and only partially applied, suggesting that anti–status quo forces are typically overpowered by pro–status quo forces. Our ex ante prediction is therefore that the gender gradient will be positive. We do note, however, that some popular discourse emphasizes a view that the gradient is negative. For example, a New York Times headline (*31*) reads "Push for Gender Equality in Tech? Some Men Say It's Gone Too Far." A 2019 YouGov poll (*32*) finds that 37% of men "believe workplace gender diversity efforts ultimately disadvantage males." A 2020 Pew Research poll (*33*) arrives at a similar figure, with 28% of men respondents saying that "women's gains toward equality have come at their expense." Stated in our terms, these perspectives claim that the anti–status quo forces dominate, yielding a negative gender gradient (at least across the 0 to 50% gender composition range).

Thus far, we have focused on the sign of the gender gradient, but we now turn to how the magnitude of the gender gradient might differ depending on other applicant features. In particular, we consider applicants' membership in minority or majority social groups.

Studies of intersectionality suggest that groups that are disadvantaged on multiple dimensions (e.g., both gender and majority/minority status) might face larger disadvantages than groups that are disadvantaged on fewer dimensions (*34*–*36*). When pro–status quo forces dominate, then, one can expect that the positive gender gradient should be steeper for disadvantaged groups and shallower for advantaged groups. That is, we expect pro–status quo forces to operate more strongly on disadvantaged than advantaged groups. If, however, anti–status quo forces dominate, then the negative gender gradient should also be steeper for disadvantaged groups and shallower for advantaged groups. Under an anti–status quo mindset, hiring managers would take advantage of any opportunity to correct multiple imbalances at the same time.

These possibilities are summarized in Fig. 2. The left facet shows how, if pro–status quo forces dominate, the gender gradient is more positive for disadvantaged groups than for advantaged groups. The right facet shows how if anti–status quo forces dominate, the gender gradient will be more negative for disadvantaged groups than for advantaged groups. In the empirical section, we will measure the gender gradients separately for majority and minority social groups.

Mechanisms of gender discrimination

Before turning to the design and results of our empirical meta-reanalysis, we briefly pause to offer a "theoretical" meta-analysis or a synthesis of the theoretical frameworks used by our study authors to explain the mechanisms through which applicant gender affects hiring decisions. We manually coded the mechanisms described in 51 papers (encompassing 57 studies) into five broad categories, described in brief below. We include a sixth category for studies whose theory sections were not about the mechanisms through which applicant gender results in hiring decisions. Figure 3 shows the results of this exercise.

Most commonly, authors draw the distinction between taste-based and statistical discrimination (*37*–*40*). Statistical discrimination is the term used to describe the heuristic inferences employers draw about applicants' productivity on the basis of the sparse information available to them (*41*–*43*). Employers' inferences can be based on accurate information about group productivity or they can be based on inaccurate stereotypes (*44*, *45*). Taste-based discrimination, by contrast, refers to the notion that employers have an inherent preference ("taste") for individuals who belong to certain social groups (*46*). Sexist, misogynist, or misandrist employers may actively discriminate against members of different gender groups because of their gender per se and not because of the productivity characteristics that may be stereotypically correlated with their gender in a given setting.

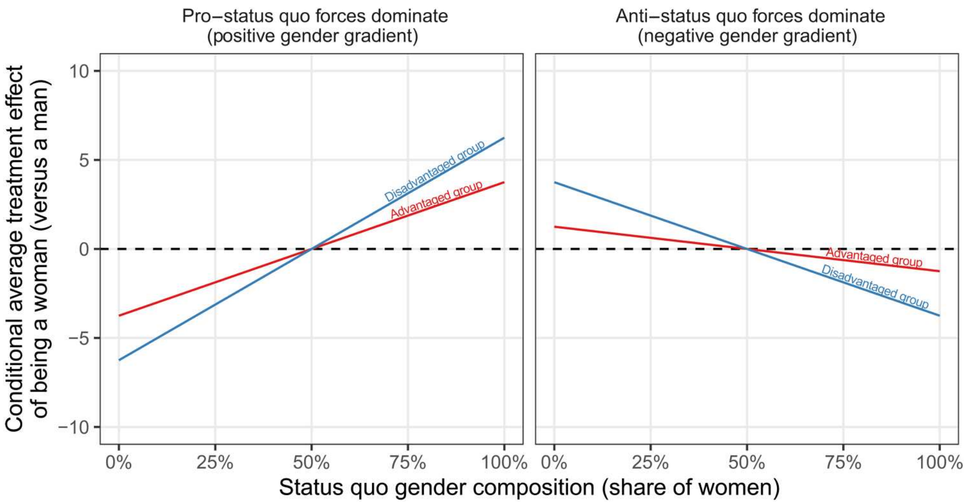

Fig. 2. Gender gradient (how the CATEs of being a woman depend on gender composition), depending on whether pro– or anti–status quo forces dominate and on membership in an advantaged or disadvantaged group.

The second most common framework for describing the mechanisms through which gender discrimination occurs is gender stereotypes, broadly conceived. These stereotypes may be either descriptive or prescriptive (*47*, *48*). The stereotype content model (*49–51*) describes how stereotypes for men (e.g., competence) and women (e.g., warmth) may result in employment discrimination. In a similar vein, role congruity theory predicts that gender discrimination derives from the congruence of gender roles with occupational roles (*47*). Correspondingly, these stereotypes might fuel the perception that women exhibit a "lack of fit" in men-dominated occupations and vice versa (*48*, *52*, *53*).

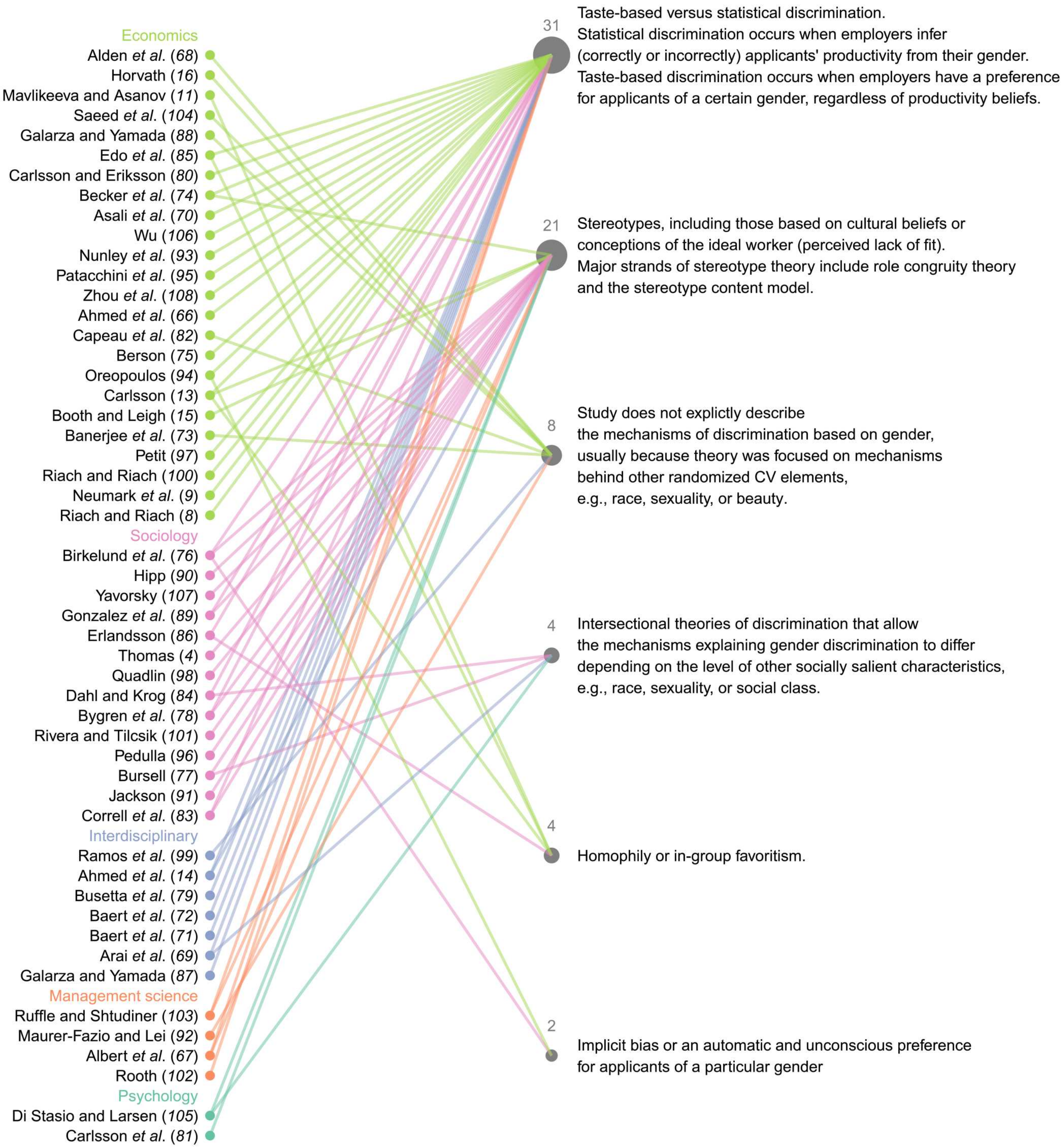

Fig. 3. Overview of the discrimination mechanisms discussed in the studies included in the meta-analysis.

Beyond these two largest categories, authors have also turned to theories of implicit bias, homophily, and intersectionality to explain the mechanisms of discrimination. Theories of implicit bias emphasize that biased employment decisions may not be rooted in employers' conscious beliefs about applicants' productivity but rather in their automatic and unconscious reactions to applicants' gender and exposure to pervasive stereotypes (*54*–*57*). In the same vein, Moss-Racusin *et al.* (*58*) report an experiment that shows that science faculty (regardless of their own gender) from research-intensive universities rated men as more hireable and competent than equally qualified women. Furthermore, in a laboratory study, implicit attitudes about race (as measured by an implicit attitudes test) were found to correlate with evaluations of hypothetical resumes (*59*).

Homophily, in general, refers to the tendency of group members to prefer in-group members to out-group members (*60*). Discrimination based on homophily would occur if hiring managers prefer applicants of their own gender to those of a different gender. Last, theories of intersectionality (*61*, *62*) emphasize that whatever the precise mechanisms of gender discrimination may be, they likely differ for members of different socially salient groups.

Most of our included studies were published in either economics (*24*) or sociology (*14*) journals, with a handful of studies published in interdisciplinary, psychology, or management science journals (*13*). All five groups of theories are represented in multiple disciplines. That said, we see in Fig. 3 that the taste-based versus statistical discrimination framework is more common in economics and the stereotype explanation is more common in sociology. We view these explanations as more complimentary than competing. Employers who engage in statistical discrimination use stereotypes to infer applicant productivity from their gender; homophily may represent one flavor of taste-based discrimination among others.

Audit experiments do not provide information that would allow researchers to determine which mechanisms are at play in a given setting, as they randomize the treatment (the gender of the applicant) through names, photos, or direct disclosure and measure the posttreatment outcome (hiring decisions) but do not measure any intervening variables. An early hope for audit experiments was that they would yield evidence of taste-based discrimination by directly providing productivity information in CVs. However, this hope was soon dashed because even conditional on the observed features of the CV, employers may draw heuristic inferences about unobserved characteristics. From audit experiments, we learn about the level of gender bias in a setting, but we do not learn about the mechanisms that explain gender bias. Bohren *et al.* (*63*) raise a related concern that while audit experiments can measure "direct" discrimination, they cannot measure the "systemic discrimination" that yields group-based inequalities in signaling technologies. Our goal in the following empirical section is to use the existing experimental record to learn how the level of gender bias covaries with the occupation gender composition, not to distinguish among potential mechanisms through which discrimination may occur.

Design

Using the terms given in (*64*), we now describe our data strategy (how we collected and prepared our data) and our answer strategy (how we summarized the resulting information).

Data strategy

We attempted to collect all published employment audit experiments that randomized applicant gender, regardless of whether the main focus of the study was applicant gender or some other characteristic. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method (*65*) for conducting systematic reviews. We searched for studies on Google Scholar and Web of Science using the keywords "field experiment gender," "field experiment labor market," "experiment gender," "experiment labor market," "gender discrimination," "discrimination testing," "correspondence testing," and "hiring discrimination." The PRISMA search process yielded over 70 papers that met our inclusion criteria. Figure S6.66 is a "PRISMA flow diagram" that illustrates the search strategy and reports the number of records retained at each stage of the process. We have obtained or reconstructed the microdata for 51 such papers, for a total of 57 separate audit studies (*4*, *8*, *9*, *11*, *13*–*16*, *66*–*108*). Figure S.2.2 shows an "evidence map" of where and when these studies were conducted. The explosion in audit experimentation began around 2005, mainly in Europe and North America.

All included studies met two criteria. First, the gender of the applicants must have been randomized; studies that did not randomize gender were excluded. Some studies included applications from both men and women, but they were sent systematically to different kinds of job openings [e.g., (*7*, *109*)]. We note that the effect of applicant gender was not always the primary inferential target of the studies we do include; some were designed to study other things like the effects of race or beauty. Second, the outcome variable must have been employers' binary hiring decision. We took an expansive view of "hiring decision." In most studies, the hiring decision was whether the applicant received a callback. In some, the decision was whether the applicant was asked to interview. In no cases was the hiring decisions the actual hiring of the applicant because the applications were all fictitious, as required by the design.

One drawback of this measurement strategy is that we can only learn about gender discrimination at the early stages of the hiring process, but discrimination could be different at later stages. For example, suppose at the interview stage, men-dominated firms use an anti–status quo policy like a gender quota, aiming to invite equal numbers of men and women to interview. If women and men are nevertheless evaluated differently at the interview stage, the audit experiment design would fail to measure such discrimination.

We obtained original datasets where possible and reconstructed datasets from reported tables and figures where necessary. From each study, we collected as fine-grained information as was possible on the country, year, and occupation for each job opening, as our goal is to estimate CATEs at the country-year-occupation level. Some studies were conducted in multiple countries, but the reported data were not disaggregated by study site, so we average over country in such cases. Some studies did not disaggregate by occupation, so we average over occupation in those cases. A description of all included studies is offered in the study manifest reported in the table S2.1.

We aimed to measure the gender composition of each occupation in the relevant country and year. To do so, we associated each job opening with an International Standard Classification of Occupations (ISCO) code (ISCO-08, or the two-digit, submajor group level code). When the original study authors provided the ISCO code, we followed their coding. Otherwise, we associated each occupation in each study with an ISCO code, following documentation from the International Labour Organization website (*110*). That same organization compiles gender composition information at the ISCO-08 level for each country and year. Where possible, we obtain the fraction of employees who are women at the country, year, and occupation level. For some countries, annual data were not available, so we rely on the fraction of women measured at the country and occupation level. This substitution is imprecise, but we do observe that in countries for which complete data were available, the fraction of women in an occupation does not typically vary much over time. Last, in a handful of cases, country-occupation level data were not available, and in those cases, we substituted in the average occupation-level estimate. Although the resulting estimates of gender composition are undoubtedly measured with some error, we take comfort that it correlates quite well with the coarse categorizations given by the study authors who characterized occupations as being men-dominated, women-dominated, or gender-balanced (see Fig. 4). A further source of measurement error derives from the heterogeneity within occupations in gender composition (*111*). Because occupation is the lowest level of aggregation at which we can summarize the experimental evidence, the measurement error associated with averaging over the within-occupation heterogeneity unfortunately cannot be avoided. However, we note that to the extent that the measurement error in gender composition is orthogonal to gender bias, our estimates of the gender gradient will be attenuated toward zero.

We harmonized measurements of applicants′ membership in a majority or a minority social group. This feature was chosen by the original authors using setting-specific definitions. For example, in the United States, the majority category usually refers to white people and the minority category usually refers to Black people (*93*). In a study conducted in the Netherlands, the minority category is applicants of Moroccan descent and the majority category is applicants of Dutch descent (*99*). We fully admit that the majority and minority labels mean different things in different contexts, but we proceed with this coarse labeling in an effort to evaluate the differential gender gradient theoretical predictions, which do not depend on the otherwise important distinctions across these categories. We were only able to reanalyze studies along this dimension if majority-minority status was randomized and distinguished in the replication or reconstructed dataset.

We also attempted to collect all survey experiments that randomized the gender of hypothetical candidates for employment, although we do not claim that our search for such studies was exhaustive. Accordingly, we report provisional meta-analytic results of these survey experimental estimates of gender discrimination in section S3.

Answer strategy

We describe our study as a meta-reanalysis because we first reanalyze each study and then meta-analyze our reanalyses. The reanalysis step proceeds as follows. First, we estimated the CATEs of being a woman (versus a man) on the binary hiring decision by subtracting the callback rate among men from the callback rate among women in a given study and occupation. The resulting treatment effect estimates are in percentage points.

Where multiple applications were submitted for the same position, we clustered our standard errors by job opening. In some of our reconstructed datasets, we were unable to reconstruct the job opening clusters, so we report unclustered (HC2) SE estimates in those cases. When we compare the clustered (CR2) and unclustered (HC2) SE estimates for the cases in which we can cluster, we find that the clustered SE is sometimes larger, sometimes smaller, never by much. Because random assignment of gender is carried out within cluster (as in a blocked experiment), clustering the SEs is not strictly necessary when characterizing within-sample uncertainty. That said, whenever possible, we follow (*112*) to cluster at the level of sampling or assignment, whichever is higher. In five

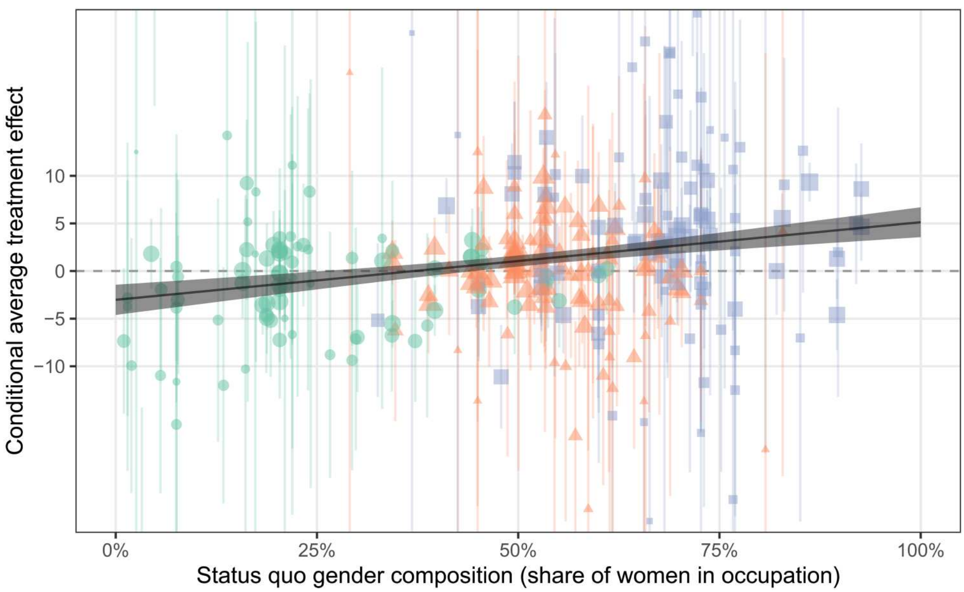

Fig. 4. Study-occupation level CATE estimates of signaling applicant is a woman versus a man on callbacks. Point size is proportional to meta-analytic weight. The regression line is derived from the model reported in column 1 of Table 1. The gender composition categories (as given by study authors where available and by us otherwise) are distinguished by color and shape (green circles, men-dominated; orange triangles, gender-balanced; blue squares, women-dominated).

occupation settings, no applicant (man or woman) received a callback, so in those cases, we used the adjustment described in (*113*) to estimate an SE. We present CATE estimates for all 57 studies in section S4.

The meta-analysis step conducts random effects meta-regression of the CATE estimates on the gender composition variable. The random effects model is given by $y_i \sim \mathcal{N}(\alpha + \beta x_i, \tau^2 + \xi^2)$, where y_i is the true CATE for study-occupation i, assumed to be drawn from a normal distribution with a mean of α plus β times the gender composition covariate x_i and with a variance that is the sum of two variance terms, the between-study variance τ^2 and within-study sampling variance ξ^2. We use maximum likelihood to estimate this model. Our main interest is the estimate of β, which we refer to as the gender gradient. From a theoretical perspective, we do not assume that the gender gradient must be linear, so we interpret β as a linear approximation of the possibly quite flexible conditional expectation function.

RESULTS

We present our results in two sections, our main analysis of the gender gradient and then our secondary analysis of how the gender gradient depends on majority versus minority status. Figure 4 plots the gender composition of each study-occupation on the horizontal axis and the associated CATE estimate on the vertical axis. Each point is sized proportionally to the weight that it receives in the meta-regression, with estimates that are more precise receiving more weight than estimates that are less precise. The meta-regression fit and 95% confidence region are overlaid on the data.

The meta-regression estimate itself is shown in column 1 of Table 1. We find that gender gradient is 8.15 with an SE of 1.50. In an occupation with 0% women, we would expect the average CATE to be −3.02 percentage points and in an occupation with 100% women, we would expect the average CATE to be −3.02 + 8.15 = +5.13 points. The magnitude of these biases is higher in women-dominated occupations than in men-dominated occupations, by a statistically significant 2.10 points (bootstrapped SE: 0.58), perhaps suggesting that pro–status quo forces operate differently across contexts. In columns 2 through 5 of Table 1, we present a series of fixed-effect models that variously include indicators for country, year, country and year, and study. Across these specifications, the estimated gender gradient remains quite stable, indicating that the overall pattern that we observe in Fig. 4 also holds within subsets of the data.

Next, we investigate how the gender gradient may differ for disadvantaged groups versus advantaged groups. Figure 5 shows the gender gradient separately for minority (in blue triangles) and majority subjects (in red circles). Contrary to expectations, the gender gradient is quite similar for these two groups. The third column of Table 2 shows that while it is true that the gender gradient is estimated to be steeper for minority applicants than for majority applicants, the uncertainty attending to the estimates is too great to declare the difference between them statistically significant.

On the whole, our results do not provide a clear answer to the question of whether the gender gradient is steeper for disadvantaged groups. Our best guess from the available evidence is that it is, but this guess is quite uncertain.

DISCUSSION

We set out to synthesize what has been learned over the past four decades from the repeated application of a single research design (the employment audit experiment) to a difficult measurement problem (the presence of gender-based employment discrimination).

We might describe this literature as "mixed," because some studies find bias against women, some find bias against men, and still others uncover no evidence of bias either way. Alternatively, we might characterize this whole literature with a single number summary like the meta-analytic average treatment effect, which turns out to show a weak pro-woman bias overall. Both of these characterizations would miss the point. To synthesize this research literature, we need to describe those contexts in which effects will be positive and those in which they will be negative.

Our meta-reanalysis breaks each study down by occupation, re-estimates gender bias within occupation, and then arranges the estimates according to the gender composition of each occupation. We learn from this procedure that gender composition predicts gender bias. The gender gradient is positive: Women are discriminated against where there are few women, and men are discriminated against where there are few men. Although this positive gender gradient hypothesis has been posited since the earliest audit experiment and likely before, the typical audit experiment is underpowered to demonstrate it. As described in section S5, we can obtain study-specific gender gradient estimates for 37 of our studies, only 7 of which reach statistical significance. Informally speaking, this analysis puts the statistical power of the average gender employment audit study to detect the gender gradient at 7/37 ≈ 19%.

Table 1. Meta-regression estimates of the gender gradient. *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$.

	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	−3.02*** (0.81)	−2.48 (1.58)	−5.43** (1.91)	−5.20** (1.82)	1.05 (1.96)
Gender gradient	8.15*** (1.50)	8.33*** (1.58)	8.69*** (1.56)	7.97*** (1.57)	7.51*** (1.54)
Fixed effects	None	Country	Year	Country and year	Study
Tau.squared	12.57	11.71	12.56	10.22	8.89
Nobs	298	298	298	298	298

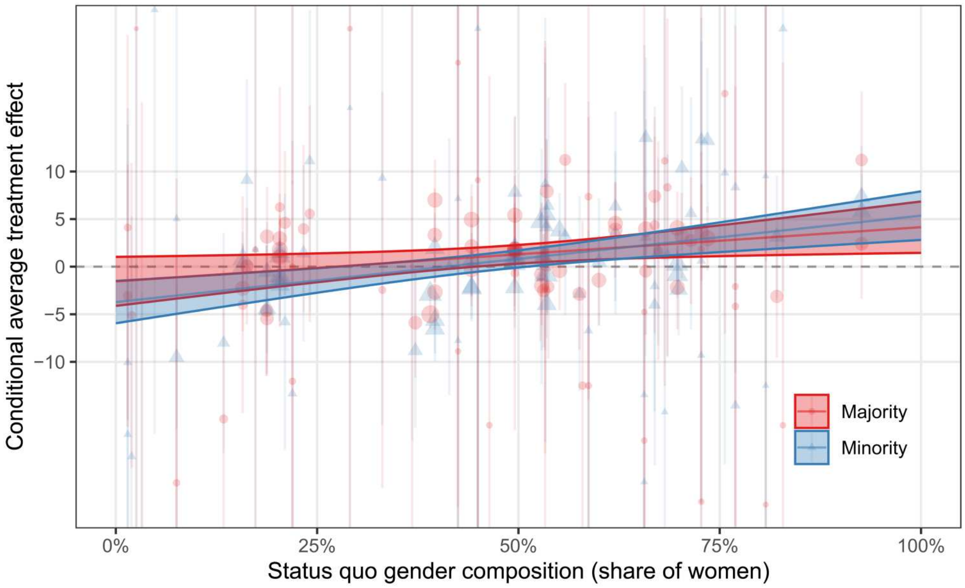

Fig. 5. Study-occupation level CATE estimates of signaling applicant is a woman versus a man on callbacks, conditional on majority/minority status. Point size is proportional to meta-analytic weight. Regression lines derived from models reported in columns 1 and 2 of Table 2. Estimates for majority group applicants plotted as red circles and as blue triangles for minority group applicants.

Learning about the gender gradient across many contexts requires the systematic aggregation of many dozens of experiments conducted on many tens of thousands of employers.

It might seem like our main finding reflects "equal opportunity bias" because bias against women is mirrored by bias against men, but this interpretation is misleading. Wages are not equivalent across industries: Men are advantaged precisely in those industries that are higher-paying, and women are advantaged in lower-paying settings (*17*–*19*). We attempted to confirm a negative gender gradient with respect to wages, but unfortunately, the requisite occupation-level wage information was not available.

In our theory section, we described how a positive gender gradient would occur if pro–status quo forces dominate and how a negative gender gradient would occur if anti–status quo forces dominate. We find clear evidence of status quo bias in this body of evidence. Hiring practices do not appear to be tuned to correct imbalances in gender composition; instead, they seem to reinforce them. This finding accords with what some describe as a "stalled" or "slowed gender revolution" (*114*, *115*), where progress toward gender employment equality in occupations has slowed.

Audit experiments are rightly praised for measuring discrimination in real-world contexts, and hypothetical vignette experiments conducted with survey respondents are rightly criticized for their artificiality. For example, Incerti (*116*) compares survey and field experimental estimates of the effects of corruption information on support for politicians, finding that the survey experimental estimates overstate the field experimental estimates by an order of magnitude. With that backdrop, it is all the more unexpected that our provisional meta-analysis of the survey experimental record on gender bias in hiring also recovers a positive gender gradient (see fig. S3.3).

Table 2. Meta-regression estimates of the gender gradient by majority/minority status. $^*P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$.

	Model 1	Model 2	Model 3
Intercept	−1.55 (1.31)	−3.72 (1.13)	−3.73** (1.14)
Gender gradient	5.70* (2.50)	9.08*** (2.26)	9.11*** (2.28)
Majority status			2.19 (1.74)
Majority × gradient			−3.42 (3.37)
Sample	Majority status group	Minority status group	Both groups
Tau.squared	6.60	6.28	6.44
Nobs	91	96	187

One important limitation of audit experiments is their reliance on fictitious resumes. The resumes must be fictitious because we cannot otherwise randomize applicant gender. As a result, we are only able to measure gender bias when employers do not already know the applicants, as in entry-level positions. Audit experiments are therefore not well suited to study employment discrimination at the promotion stage [for a clever exception see (*71*), which found that women, relative to men, receive fewer callbacks when applying for jobs that are at a higher level compared to their current job]. Although we cannot demonstrate this claim with experimental evidence, we would extrapolate from our meta-analysis that the gender gradient that we see across occupations would also occur within occupations. To the extent that higher ranks within an occupation are dominated by men, we would expect bias against women for promotion, i.e., a glass ceiling (*117*, *118*).

The audit experimental record on gender bias in hiring currently stands at over 70 studies and by synthesizing 57 of them, we learn that gender composition predicts gender bias. What is there left to learn about gender bias from audit experiments?

First, we have have very large gaps in our evidence base. Figure S2.2 shows where and when the existing studies were conducted. Most were conducted in North America and Europe since 2005; only a handful of studies have been conducted in South America, Asia, or Africa. That said, our expectation is that if we were to conduct audit experiments in those places, we would again recover a positive gender gradient.

Second, we might wish to conduct future experiments in exactly those places where we expect a deviation from this pattern, i.e., where we expect positive effects in men-dominated settings or negative effects in women-dominated settings. Such a deviation would require that anti–status quo forces (affirmative action policies or similar) dominate the forces that maintain the status quo. Future audit experiments should deliberately choose research sites undergoing change to show the reductions in bias that accompany affirmative action.

Last, future audit experimenters should heed the advice of Butler and Crabtree (*119*) to "move beyond measurement," which is to say we should marry the inferential powers of audit experiments for measuring gender bias with randomized interventions for reducing gender bias. For example, Fang *et al.* (*120*) test the effectiveness of a policy to reduce housing discrimination on the basis of race in New York City by randomizing the policy and then measuring housing discrimination with an audit design. Previous literature suggests some routes forward for reducing gender bias, e.g., (*23*, *121*), that could be profitably crossed with future audit experiments.

Supplementary Materials

This PDF file includes:
Sections S1 to S6
Figs. S1.1 to S6.63
Table S2.1
References

REFERENCES AND NOTES

1. E. Zschirnt, D. Ruedin, Ethnic discrimination in hiring decisions: A meta-analysis of correspondence tests 1990–2015. *J. Ethn. Migr. Stud.* **42**, 1115–1134 (2016).
2. M.-A. Valfort, Anti-muslim discrimination in france: Evidence from a field experiment. *World Dev.* **135**, 105022 (2020).
3. N. Drydakis, Sexual orientation discrimination in the united kingdom's labour market: A field experiment. *Hum. Relat.* **68**, 1769–1796 (2015).
4. K. Thomas, The labor market value of taste: An experimental study of class bias in us employment. *Sociol. Sci.* **5**, 562–595 (2018).
5. K. Auspurg, T. Hinz, L. Schmid, Contexts and conditions of ethnic discrimination: Evidence from a field experiment in a german housing market. *J. Hous. Econ.* **35**, 26–36 (2017).
6. D. M. de Lafuente, Cultural assimilation and ethnic discrimination: An audit study with schools. *Labour Econ.* **72**, 102058 (2021).
7. M. Bertrand, S. Mullainathan, Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am. Econ. Rev.* **94**, 991–1013 (2004).
8. P. A. Riach, J. Rich, Testing for sexual discrimination in the labour market. *Aust. Econ. Pap.* **26**, 165–178 (1987).
9. D. Neumark; R. J. Bank, K. D. Van Nort, Sex discrimination in restaurant hiring: An audit study. *Q. J. Econ.* **111**, 915–941 (1996).
10. L. Lippens, S. Vermeiren, S. Baert, The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments. *Eur. Econ. Rev.* **151**, 104315 (2022).
11. M. Mavlikeeva, I. Asanov, *Can Group Identity Explain Gender Gap in Recruitment Process?* (ZBW - Leibniz Information Centre for Economics, 2020).
12. G. Azmat, B. Petrongolo, Gender and the labor market: What have we learned from field and lab experiments? *Labour Econ.* **30**, 32–40 (2014).
13. M. Carlsson, Does hiring discrimination cause gender segregation in the Swedish labor market? *Fem. Econ.* **17**, 71–102 (2011).
14. A. Ahmed, M. Granberg, S. Khanna, Gender discrimination in hiring: An experimental reexamination of the swedish case. *PLOS ONE* **16**, e0245513 (2021).
15. A. Booth, A. Leigh, Do employers discriminate by gender? A field experiment in female-dominated occupations. *Econ. Lett.* **107**, 236–238 (2010).
16. G. Horváth, The impact of marital status on job finding: A field experiment in the chinese labor market. *BE J. Econ. Anal. Policy* **20**, 20190364 (2020).
17. A. Levanon, P. England, P. Allison, Occupational feminization and pay: Assessing causal dynamics using 1950–2000 U.S. census data. *Soc. Forces* **88**, 865–891 (2009).
18. P. N. Cohen, M. L. Huffman, Individuals, jobs, and labor markets: The devaluation of women's work. *Am. Sociol. Rev.* **68**, 443–463 (2003).
19. C. Kim, C. R. Tamborini, A. Sakamoto, Field of study in college and lifetime earnings in the united states. *Sociol. Educ.* **88**, 320–339 (2015).
20. W. T. Cox, P. G. Devine, in *Confronting Prejudice and Discrimination* (Elsevier, 2019), pp. 249–274.
21. P. G. Devine, P. S. Forscher, W. T. L. Cox, A. Kaatz, J. Sheridan, M. Carnes, A gender bias habit-breaking intervention led to increased hiring of female faculty in stemm departments. *J. Exp. Soc. Psychol.* **73**, 211–215 (2017).
22. M. Carnes, P. G. Devine, L. B. Manwell, A. Byars-Winston, E. Fine, C. E. Ford, P. Forscher, C. Isaac, A. Kaatz, W. Magua, M. Palta, J. Sheridan, The Effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Acad. Med.* **90**, 221–230 (2015).
23. C. F. Karpowitz, J. Q. Monson, J. R. Preece, How to elect more women: Gender and candidate success in a field experiment. *Am. J. Pol. Sci.* **61**, 927–943 (2017).
24. T. Schmader, T. C. Dennehy, A. S. Baron, Why antibias interventions (need not) fail. *Perspect. Psychol. Sci.* **17**, 1381–1403 (2022).
25. L. F. Pendry, D. M. Driscoll, S. C. Field, Diversity training: Putting theory into practice. *J. Occup. Organ. Psychol.* **80**, 27–50 (2007).
26. A. Kalev, F. Dobbin, E. Kelly, Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *Am. Sociol. Rev.* **71**, 589–617 (2006).
27. D. Dahlerup, L. Freidenvall, Quotas as a 'fast track' to equal representation for women: Why scandinavia is no longer the model. *Int. Fem. J. Polit.* **7**, 26–48 (2005).
28. R. Murray, Quotas for men: Reframing gender quotas as a means of improving representation for all. *Am. Polit. Sci. Rev.* **108**, 520–532 (2014).
29. C. L. Williams, The glass escalator: Hidden advantages for men in the "female" professions. *Soc. Probl.* **39**, 253–267 (1992).
30. M. J. Budig, Male advantage and the gender composition of jobs: Who rides the glass escalator? *Soc. Probl.* **49**, 258–277 (2002).
31. N. Bowles, Push for gender equality in tech? Some men say it's gone too far, *New York Times*, 23 September 2017.
32. L. Sanders, One in three men say males are disadvantaged by gender diversity efforts in the workplace. *YouGovAmerica*, 21 March 2019.
33. T. Mitchell, A century after women gained the right to vote, majority of Americans see work to do on gender equality, *Pew Research Center* (2021).
34. F. Beale, Double Jeopardy: To be Black and female, in *The Black Woman: An Anthology* (New American Library, 1979), pp. 90–100.

35. V. Purdie-Vaughns, R. P. Eibach, Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles* **59**, 377–391 (2008).
36. I. Browne, J. Misra, The intersection of gender and race in the labor market. *Annu. Rev. Sociol.* **29**, 487–513 (2003).
37. D. Pager, H. Shepherd, The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annu. Rev. Sociol.* **34**, 181–209 (2008).
38. L. Quillian, A. H. Midtbøen, Comparative perspectives on racial discrimination in hiring: The rise of field experiments. *Annu. Rev. Sociol.* **47**, 391–415 (2021).
39. A. H. Midtbøen, The invisible second generation? Statistical discrimination and immigrant stereotypes in employment processes in norway. *J. Ethn. Migr. Stud.* **40**, 1657–1675 (2014).
40. J. Guryan, K. K. Charles, Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *Econ. J.* **123**, F417–F432 (2013).
41. S. Schwab, Is statistical discrimination efficient? *Am. Econ. Rev.* **76**, 228–234 (1986).
42. E. S. Phelps, The statistical theory of racism and sexism. *Am. Econ. Rev.* **62**, 659–661 (1972).
43. K. Arrow, The theory of discrimination, in *Discrimination in Labour Markets,* O. Ashenfelter, A. Rees, Eds. (Princeton Univ. Press, 1973), pp. 3–33.
44. P. England, *Comparable Worth: Theories and Evidence.* (Aldine de Gruyter, 1992), p 346.
45. W. T. Bielby, J. N. Baron, Men and women at work: Sex segregation and statistical discrimination. *Am. J. Sociol.* **91**, 759–799 (1986).
46. G. Becker, *The Economics of Discrimination* (University of Chicago Press, 1957).
47. A. H. Eagly, S. J. Karau, Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* **109**, 573–598 (2002).
48. M. E. Heilman, Gender stereotypes and workplace bias. *Res. Organ. Behav.* **32**, 113–135 (2012).
49. P. Glick, S. T. Fiske, Hostile and benevolent sexism: Measuring ambivalent sexist attitudes toward women. *Psychol. Women Q.* **21**, 119–135 (1997).
50. S. J. Correll, C. L. Ridgeway, *Expectation States Theory* (Kluwer, 2003), chap 2, pp. 1–23.
51. A. J. C. Cuddy, S. T. Fiske, P. Glick, Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Adv. Exp. Soc. Psychol.* **40**, 61–149 (2008).
52. M. E. Heilman, Sex bias in work settings: The lack of fit model. *Res. Organ. Behav.* **5**, 269–298 (1983).
53. C. T. Begeny, M. K. Ryan, C. A. Moss-Racusin, G. Ravetz, In some professions, women have become well represented, yet gender bias persist—Perpetuated by those who think it is not happening. *Sci. Adv.* **6**, eaba7814 (2020).
54. L. A. Rudman, P. Glick, Prescriptive gender stereotypes and backlash toward agentic women. *J. Soc. Issues* **57**, 743–762 (2001).
55. J. T. Jost, L. A. Rudman, I. V. Blair, D. R. Carney, N. Dasgupta, J. Glaser, C. D. Hardin, The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Res. Organ. Behav.* **29**, 39–69 (2009).
56. P. G. Devine, Stereotypes and prejudice: Their automatic and controlled components. *J. Pers. Soc. Psychol.* **56**, 5–18 (1989).
57. I. Régner, C. Thinus-Blanc, A. Netter, T. Schmader, P. Huguet, Committees with implicit biases promote fewer women when they do not believe gender bias exists. *Nat. Hum. Behav.* **3**, 1171–1179 (2019).
58. C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, J. Handelsman, Science faculty's subtle gender biases favor male students. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16474–16479 (2012).
59. M. Bertrand, D. Chugh, S. Mullainathan, Implicit discrimination. *Am. Econ. Rev.* **95**, 94–98 (2005).
60. M. McPherson, L. Smith-Lovin, J. M. Cook, Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
61. S. Cho, K. W. Crenshaw, L. McCall, Toward a field of intersectionality studies: Theory, applications, and praxis. *Signs* **38**, 785–810 (2013).
62. C. L. Ridgeway, T. Kricheli-Katz, Intersecting cultural beliefs in social relations: Gender, race, and class binds and freedoms. *Gend. Soc.* **27**, 294–318 (2013).
63. J. A. Bohren, P. Hull, A. Imas, *Systemic Discrimination: Theory and Measurement* (Technical report, National Bureau of Economic Research, 2022).
64. G. Blair, J. Cooper, A. Coppock, M. Humphreys, Declaring and diagnosing research designs. *Am. Polit. Sci. Rev.* **113**, 838–859 (2019).
65. M. J. Page, J. E. Mc Kenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. M. Donald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **372**, n71 (2021).
66. A. M. Ahmed, L. Andersson, M. Hammarstedt, Are gay men and lesbians discriminated against in the hiring process? *Southern Econ. J.* **79**, 565–585 (2013).
67. R. Albert, L. Escot, J. A. Fernández-Cornejo, A field experiment to study sex and age discrimination in the Madrid labour market. *Int. J. Hum. Resour. Manag.* **22**, 351–375 (2011).
68. L. Aldén, S. Bastani, M. Hammarstedt, Ethnic background and the value of self-employment experience: Evidence from a randomized field experiment. *Oxf. Bull. Econ. Stat.* **83**, 1287–1310 (2021).
69. M. Arai, M. Bursell, L. Nekby, The reverse gender gap in ethnic discrimination: Employer stereotypes of men and women with arabic names. *Int. Migr. Rev* **50**, 385–412 (2016).
70. M. Asali, N. Pignatti, S. Skhirtladze, Employment discrimination in a former soviet union republic: Evidence from a field experiment. *J. Comp. Econ.* **46**, 1294–1309 (2018).
71. S. Baert, A.-S. De Pauw, N. Deschacht, Do employer preferences contribute to sticky floors? *ILR Rev.* **69**, 714–736 (2016).
72. S. Baert, S. De Visschere, K. Schoors, D. Vandenberghe, E. Omey, First depressed, then discriminated against? *Soc. Sci. Med.* **170**, 247–254 (2016).
73. A. Banerjee, M. Bertrand, S. Datta, S. Mullainathan, Labor market discrimination in delhi: Evidence from a field experiment. *J. Comp. Econ.* **37**, 14–27 (2009).
74. S. O. Becker, A. Fernandes, D. Weichselbaumer, Discrimination in hiring based on potential and realized fertility: Evidence from a large-scale field experiment. *Labour Econ.* **59**, 139–152 (2019).
75. C. Berson, Does competition induce hiring equity? *Halshs, No. 00718627* (2012).
76. G. E. Birkelund, B. Lancee, E. N. Larsen, J. G. Polavieja, J. Radl, R. Yemane, Gender discrimination in hiring: Evidence from a cross-national harmonized field experiment. *Eur. Sociol. Rev.* **38**, 337–354 (2022).
77. M. Bursell, The multiple burdens of foreign-named men—Evidence from a field experiment on gendered ethnic hiring discrimination in sweden. *Eur. Sociol. Rev.* **30**, 399–409 (2014).
78. M. Bygren, A. Erlandsson, M. Gähler, Do employers prefer fathers? Evidence from a field experiment testing the gender by parenthood interaction effect on callbacks to job applications. *Eur. Sociol. Rev.* **33**, 337–348 (2017).
79. G. Busetta, M. G. Campolo, D. Panarello, Immigrants and italian labor market: Statistical or taste-based discrimination? *Genus* **74**, 4 (2018).
80. M. Carlsson, S. Eriksson, Age discrimination in hiring decisions: Evidence from a field experiment in the labor market. *Labour Econ.* **59**, 173–183 (2019).
81. R. Carlsson, J. Agerström, F. Björklund, M. Carlsson, D.-O. Rooth, Testing for backlash in hiring: A field experiment on agency, communion, and gender. *J. Pers. Psychol.* **13**, 204–214 (2014).
82. B. Capéau, L. Eeman, S. Groenez, M. Lamberts, Two concepts of discrimination: Inequality of opportunity versus unequal treatment of equals. *Ecore Discuss. Pap.* **58**, (2012).
83. S. J. Correll, S. Benard, I. Paik, Getting a job: Is there a motherhood penalty? *Am. J. Sociol.* **112**, 1297–1338 (2007).
84. M. Dahl, N. Krog, Experimental evidence of discrimination in the labour market: Intersections between ethnicity, gender, and socio-economic status. *Eur. Sociol. Rev.* **34**, 402–417 (2018).
85. A. Edo, N. Jacquemet, C. Yannelis, Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications. *Rev. Econ. Househ.* **17**, 349–376 (2019).
86. A. Erlandsson, Do men favor men in recruitment? A field experiment in the Swedish labor market. *Work Occup.* **46**, 239–264 (2019).
87. F. B. Galarza, G. Yamada, Labor market discrimination in Lima, Peru: Evidence from a field experiment. *World Dev.* **58**, 83–94 (2014).
88. F. B. Galarza, G. Yamada, Afro-descendants in Peru: Do beauty and race matter in the labor market? *Rev. Dev. Econ.* **23**, 211–230 (2019).
89. M. J. González, C. Cortina, J. Rodríguez, The role of gender stereotypes in hiring: A field experiment. *Eur. Sociol. Rev.* **35**, 187–204 (2019).
90. L. Hipp, Do hiring practices penalize women and benefit men for having children? Experimental evidence from germany. *Eur. Sociol. Rev.* **36**, 250–264 (2020).
91. M. Jackson, Disadvantaged through discrimination? The role of employers in social stratification 1. *Br. J. Sociol.* **60**, 669–692 (2009).
92. M. Maurer-Fazio, L. Lei, "As rare as a panda": How facial attractiveness, gender, and occupation affect interview callbacks at Chinese firms. *Int. J. Manpow.* **36**, 68–85 (2015).
93. J. M. Nunley, A. Pugh, N. Romero, R. A. Seals, Racial discrimination in the labor market for recent college graduates: Evidence from a field experiment. *BE J. Econ. Anal. Policy* **15**, 1093–1125 (2015).

94. P. Oreopoulos, Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes. *Am. Econ. J. Econ. Policy* **3**, 148–171 (2011).
95. E. Patacchini, G. Ragusa, Y. Zenou, Unexplored dimensions of discrimination in Europe: Homosexuality and physical appearance. *J. Popul. Econ.* **28**, 1045–1073 (2015).
96. D. S. Pedulla, Penalized or protected? Gender and the consequences of nonstandard and mismatched employment histories. *Am. Sociol. Rev.* **81**, 262–289 (2016).
97. P. Petit, The effects of age and family constraints on gender hiring discrimination: A field experiment in the french financial sector. *Labour Econ.* **14**, 371–391 (2007).
98. N. Quadlin, The mark of a woman's record: Gender and academic performance in hiring. *Am. Sociol. Rev.* **83**, 331–360 (2018).
99. M. Ramos, L. Thijssen, M. Coenders, Labour market discrimination against moroccan minorities in the Netherlands and Spain: A cross-national and cross-regional comparison. *J. Ethn. Migr. Stud.* **47**, 1261–1284 (2021).
100. P. A. Riach, J. Rich, An experimental investigation of sexual discrimination in hiring in the English labor market. *BEJ. Econ. Anal. Policy* **6**, 1416 (2006).
101. L. A. Rivera, A. Tilcsik, Class advantage, commitment penalty: The gendered effect of social class signals in an elite labor market. *Am. Sociol. Rev.* **81**, 1097–1131 (2016).
102. D.-O. Rooth, Obesity, attractiveness, and differential treatment in hiring a field experiment. *J. Hum. Resour.* **44**, 710–735 (2009).
103. B. J. Ruffle, Z. Shtudiner, Are good-looking people more employable? *Manage. Sci.* **61**, 1760–1776 (2015).
104. A. Saeed, S. Maqsood, A. Rafique, Color matters: Field experiment to explore the impact of facial complexion in pakistani labor market. *J. Asia Pac. Econ.* **24**, 347–363 (2019).
105. V. Di Stasio, E. N. Larsen, The racialized and gendered workplace: Applying an intersectional lens to a field experiment on hiring discrimination in five european labor markets. *Soc. Psychol. Q.* **83**, 229–250 (2020).
106. K. Wu, Beauty and employment: A field experiment on appearance discrimination in China's labor market (2017);http://dx.doi.org/10.2139/ssrn.3044200.
107. J. E. Yavorsky, Uneven patterns of inequality: An audit analysis of hiring-related practices by gendered and classed contexts. *Soc. Forces* **98**, 461–492 (2019).
108. X. Zhou, J. Zhang, X. Song, *Gender Discrimination in Hiring: Evidence from 19,130 Resumes in China. MPRA Paper No.43543* (University Library of Munich, 2013).
109. N. Drydakis, Sexual orientation discrimination in the cypriot labour market. Distastes or uncertainty? *Int. J. Manpow.* **35**, 720–744 (2014).
110. International Labour Organization, Structure of the international standard classification of occupations (isco-08) (International Labour Organization, 2016).
111. A. Martin-Caughey, What's in an occupation? Investigating within-occupation variation and gender segregation using job titles and task descriptions. *Am. Sociol. Rev.* **86**, 960–999 (2021).
112. A. Abadie, S. Athey, G. W. Imbens, J. Wooldridge, When should you adjust standard errors for clustering? (Working Paper 24003, National Bureau of Economic Research, 2017).
113. A. Agresti, B. A. Coull, Approximate is better than "exact" for interval estimation of binomial proportions. *Am. Stat.* **52**, 119–126 (1998).
114. P. England, The gender revolution: Uneven and stalled. *Gend. Soc.* **24**, 149–166 (2010).
115. P. England, A. Levine, E. Mishel, Progress toward gender equality in the United States has slowed or stalled. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 6990–6997 (2020).
116. T. Incerti, Corruption information and vote share: A meta-analysis and lessons for experimental design. *Am. Polit. Sci. Rev.* **114**, 761–774 (2020).
117. J. Baxter, E. O. Wright, The glass ceiling hypothesis: A comparative study of the United States, Sweden, and Australia. *Gend. Soc.* **14**, 275–294 (2000).
118. M. Hultin, Some take the glass escalator, some hit the glass ceiling? Career consequences of occupational sex segregation. *Work Occup.* **30**, 30–61 (2003).
119. D. M. Butler, C. Crabtree, Moving beyond measurement: Adapting audit studies to test bias-reducing interventions. *J. Exp. Political Sci.* **4**, 57–67 (2017).
120. A. H. Fang, A. M. Guess, M. Humphreys, Can the government deter discrimination? Evidence from a randomized intervention in New York City. *J. Theor. Polit.* **81**, 127–141 (2019).
121. H. Finseraas, Å. Johnsen, A. Kotsadam, G. Torsvik, Exposure to female colleagues breaks the glass ceiling—Evidence from a combined vignette and field experiment. *Eur. Econ. Rev.* **90**, 363–374 (2016).
122. B. S. Bell, K. J. Klein, Effects of disability, gender, and job level on ratings of job applicants. *Rehabil. Psychol.* **46**, 229–246 (2001).
123. L. Blommaert, M. Coenders, F. van Tubergen, Ethnic discrimination in recruitment and decision makers' features: Evidence from laboratory experiment and survey data using a student sample. *Soc. Indic.s Res.* **116**, 731–754 (2014).
124. M. Carlsson, H. Finseraas, A. H. Midtbøen, G. L. Rafnsdóttir, Gender bias in academic recruitment? Evidence from a survey experiment in the nordic region. *Eur. Sociol. Rev.* **37**, 399–410 (2021).
125. A. Bertogg, C. Imdorf, C. Hyggen, D. Parsanoglou, R. Stoilova, Gender discrimination in the hiring of skilled professionals in two male-dominated occupational fields: A factorial survey experiment with real-world vacancies and recruiters in four european countries. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* **72**, 261–289 (2020).
126. G. Mari, R. Luijkx, Gender, parenthood, and hiring intentions in sex-typical jobs: Insights from a survey experiment. *Res. Soc. Stratif. Mobil.* **65**, 100464 (2020).
127. V. Di Stasio, Education as a signal of trainability: Results from a vignette study with italian employers. *Eur. Sociol. Rev.* **30**, 796–809 (2014).
128. V. Di Stasio, H. G. Van de Werfhorst, Why does education matter to employers in different institutional contexts? A vignette study in england and the netherlands. *Soc. Forces* **95**, 77–106 (2016).

Acknowledgments

We thank K. Okura, A. Bertrogg, B. Ruffle, and participants at the 2022 European Consortium for Sociological Research conference and the Yale Gender in the Social Sciences Workshop for helpful feedback. We thank the participants in an Experiments in Governance and Politics feedback session: A. Clayton, M. Mendez-Beck, J. Preece, J. Rickne, D. Teele, S. Tombe, and L. Young. We are grateful to the authors who shared their raw data with us: L. Alden, M. Asali, I. Asanov, B. Balkan, B. Bell, L. Blommaert, M. Bygren, B. Capeau, M. Carlssdon, R. Carlssdon, M. Dahl, A. Edo, A. Erlandsson, L. Escot, M. Gahler, F. B. Galarza, L. Hipp, M. Jackson, A. H. Midtbøen, J. Nunley, P. Petit, B. Ruffle, V. Di Stasio, and K. Thomas. **Funding:** D.R.G. gratefully acknowledges funding support from the Danish National Research Foundation (DNRF144). **Author contributions:** Both authors designed the project, participated in the writing and editing of the manuscript, and collected and analyzed the data, and A.C. produced the visualizations. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All data and code needed to reproduce the figures and tables in the main text and in the Supplementary Materials are available on Dataverse at https://doi.org/10.7910/DVN/5MCRNY.

Submitted 8 September 2022
Accepted 4 April 2023
Published 5 May 2023
10.1126/sciadv.ade7979

Science Advances

Gender composition predicts gender bias: A meta-reanalysis of hiring discrimination audit experiments

Diana Roxana Galos and Alexander Coppock

Sci. Adv., **9** (18), eade7979.
DOI: 10.1126/sciadv.ade7979