



# Who Shortlists? Evidence on Gender Disparities in Hiring Outcomes

Administrative Science Quarterly  
2024, Vol. 69(4)1044–1084  
© The Author(s) 2024



asq-jscu.org

**Almasa Sarabi<sup>1</sup> and Nico Lehmann<sup>2</sup>**

## Abstract

A substantial body of research examines the relevance of hiring as a source of gender disparities in organizations. However, there is limited evidence on how different sets of key organizational decision makers contribute to gender disparities in hiring outcomes. To address this research gap, we exploit the staggered adoption of a new hiring process in a multinational corporation, which transferred from hiring managers to HR departments the task of shortlisting: narrowing a large pool of candidates to a more manageable set before final decision making. Using a difference-in-differences design, we find that the transfer of shortlisting responsibility increased the share of newly hired women. Additional tests based on quantitative and qualitative data are largely consistent with our finding that the transfer of shortlisting from hiring managers to HR departments led to fewer gender disparities in hiring outcomes given the increased expert knowledge in evaluating candidates and reduced opportunity costs for conducting such evaluations. Our setting offers a unique opportunity to help isolate key organizational decision makers' role in contributing to gender disparities in hiring outcomes, and our findings have implications for how to alleviate gender disparities in employment.

**Keywords:** gender disparities, hiring, shortlists, difference-in-differences analysis, field evidence

This study examines whether the transfer of shortlisting from hiring managers to human resources (HR) departments affects gender disparities in hiring outcomes. Gender disparities in the workplace remain a ubiquitous problem in contemporary organizations, and research has examined the role that the demand side, that is, employers' actions and choices, plays in generating gender disparities in different organizational contexts such as pay (e.g., Elvira and Graham, 2002; Abraham, 2017), promotion (e.g., Phillips, 2005; Dencker, 2008), performance evaluations

<sup>1</sup> Amsterdam Business School, University of Amsterdam

<sup>2</sup> Erasmus School of Economics, Erasmus University Rotterdam

## Corresponding author:

Almasa Sarabi, University of Amsterdam, Amsterdam Business School, Plantage Muidergracht 12, Amsterdam 1018 TV, Netherlands

(e.g., Castilla and Benard, 2010), and hiring (e.g., Fernandez-Mateo and Fernandez, 2016; Leung and Koppman, 2018). The hiring context is especially important because it drives both occupational and economic sorting (Bills, Di Stasio, and Gërkhani, 2017; Rivera, 2020).

Most hiring processes involve identifying potential candidates, evaluating candidates' qualifications against job requirements, and negotiating employment terms (Breaugh, 2013; Keller, 2018). HR departments and hiring managers are two sets of key organizational decision makers throughout these processes. An organization's expertise in recruitment and selection is typically concentrated within its HR department; the main task of HR professionals is to select qualified candidates for different kinds of jobs (Dobbin, 2009; Stainback, Tomaskovic-Devey, and Skaggs, 2010). Hiring managers are often revenue-generating professionals who make the final hiring decisions and then work with the hired candidates (Rivera, 2012, 2015). Both sets of organizational decision makers are gatekeepers who facilitate career opportunities for some candidates while blocking entry for others (Rivera, 2020).

A large body of research examines the role of hiring in generating gender disparities. This research commonly argues that the hiring context is prone to disparate outcomes because decisions leading to disparities are made under high uncertainty and are more difficult to detect in the beginning of the employment relationship than at later points in time (e.g., Petersen and Saporta, 2004). Consistent with this notion, evidence suggests that employers are less likely to hire from candidate pools with a large proportion of gender-atypical applicants (Leung and Koppman, 2018), that conditions at initial hire tend to be better for men than for women (Petersen and Saporta, 2004), and that employers are less likely to invite women than men for interviews when responding to otherwise identical résumés with randomly assigned signals of gender (for an overview, see Azmat and Petrongolo, 2014). Evidence further suggests that women are less prevalent than men among hires for higher-paid jobs (Fernandez and Mors, 2008) and, on average, are less likely than men to receive job offers, conditional on being interviewed (Fernandez and Campero, 2017).

Despite the relevance of hiring as a potential source of gender disparities and important insights into employers' role in producing these disparities, we have little evidence about how different sets of key organizational decision makers contribute to disparate hiring outcomes. This lack of differentiation is surprising given substantial variation in HR practices and in types of decision makers across organizations; given the prevalence of HR and hiring managers as two important sets of gatekeepers in most organizations' hiring processes (Rivera, 2012, 2020); and given existing insights that these key decision makers often differ in terms of training, competence, and incentives (Lowe, 1992; McGovern et al., 1997). Evidence about how different sets of key decision makers may contribute to gender disparities in hiring can help us understand how to alleviate, or treat, gender disparities in employment. Understanding the treatment of gender disparities within organizational contexts is important and nontrivial (Reskin, 2003; Kalev, Dobbin, and Kelly, 2006).<sup>1</sup> Differences in key decision makers' roles may

---

<sup>1</sup> Diversity (bias) training and evaluations as prevailing treatments for employment disparities in many organizations are a case in point. While we know from experimental research that unconscious bias is endemic and thus a plausible source of workplace disparities, evidence from the field suggests that diversity trainings—explicitly designed to treat the bias—do not, for example, result in higher managerial diversity in a competitive market setting (Kalev, Dobbin, and Kelly, 2006).

generate valuable insights into variations in the sources of gender disparities, which, in turn, may inform organizational strategies to treat these disparities.

Accordingly, the goal of our study is to empirically isolate the role of key organizational decision makers in shaping gender disparities in hiring outcomes. Our empirical analysis is set in the competitive market setting of a global, technology-focused, multinational corporation, Alpha (pseudonym). In this setting, we examine the transfer of shortlisting, an early stage of the hiring process that narrows a large pool of candidates to a more manageable set before final decision making (Huber, Neale, and Northcraft, 1987; Highhouse, 1997), from hiring managers to HR departments. Hereafter, we refer to this transfer of responsibility as the intervention.<sup>2</sup> The intervention tasked HR with forming shortlists containing a maximum of seven candidates per job queue to submit to hiring managers for further consideration. Hiring managers remained responsible for later stages of the hiring process, such as interviews and final hiring decisions.

Our empirical setting is suitable for our research goal for at least two reasons. First, researchers commonly face data accessibility constraints as organizations restrict insights into their employment processes and do not share details about employment outcomes. For our study, Alpha gave us access to detailed internal hiring data. We also gained access to the results of a regular, company-wide survey administered to Alpha's hiring managers and were able to conduct semi-structured interviews with different stakeholders at Alpha. Second, our setting helps us to overcome key identification challenges that typically arise when researchers examine the effects of organizational practices on gender disparities. Decision making within employment contexts is typically endogenous, which makes it difficult to draw causal inferences (Oyer and Schaefer, 2011; Mocanu, 2022). Also, organizations that adopt new processes (e.g., new hiring practices) often do so at once, which impedes the creation of a control group to approximate the counterfactual (Castilla, 2015; Tolbert and Castilla, 2017). We address these identification challenges by exploiting the fact that Alpha adopted the new hiring process in seven predetermined, staggered time waves across all its worldwide subsidiaries.

Using a sample that consists of 8,750 externally hired candidates over a 24-month period, we examine the staggered adoption of the new hiring process at Alpha, with the help of a difference-in-differences (DiD) design. Our findings show that the intervention increased the share of newly hired women. We conduct various tests to increase our confidence in the causal interpretation of this intervention effect. We show that, consistent with the plausibility of the parallel trends assumption, the intervention effect does not trend ahead of the actual intervention and that the timing of the staggered adoption was plausibly unrelated to gender concerns in Alpha's hiring. We also demonstrate that our findings are robust when we used alternative estimation techniques such as stacked DiD estimation (Cengiz et al., 2019) and imposed a covariate balance via weighting techniques.

We then turn to potential mechanisms situated at the shortlisting stage of the hiring process that could explain the increase in women hired. First, compared to hiring managers, HR professionals might have higher expert knowledge in evaluating candidates and might face lower opportunity costs for conducting such evaluations (a differential evaluation mechanism). Thus, HR professionals may

---

<sup>2</sup> We use the terms "intervention" and "treatment" interchangeably.

be less likely than hiring managers to deviate from predefined shortlisting criteria and to redefine the criteria for job success depending on the candidate's gender. Second, the intervention might reduce the likelihood that social networks and referrals reproduce the gender composition of Alpha's existing men-dominated workforce (a social influence mechanism). Given that HR professionals are more external to the job than hiring managers are, network status might have a lower (informational) value for HR than for hiring managers. Third, the intervention might lead to a change in same-gender preferences (a homophily mechanism). Given that the gender composition of HR departments is, on average, more dominated by women than it is among hiring managers, the transfer of shortlisting might skew shortlists in favor of women.

We probe these mechanisms through additional quantitative analyses, interviews, and survey findings. We also test whether the increase in women hired can be explained by the intervention leading to (behavioral) changes at other stages of the hiring process, such as the job planning and job offer stages. While our analyses cannot causally isolate an individual explanation, the findings are largely consistent with the differential evaluation mechanism, meaning that the intervention led to a change in candidate evaluation given HR departments' higher expert knowledge in evaluating candidates and lower opportunity costs for investing time and effort in doing so. Overall, our findings help isolate the role of key organizational decision makers in contributing to gender disparities in hiring outcomes and inform the debate about how to alleviate gender disparities in employment.

## **THE EFFECT OF TRANSFERRING SHORTLISTING TO HR ON GENDER DISPARITIES**

We expect that the transfer of shortlisting from hiring managers to HR reduces gender disparities in hiring outcomes. We argue that the transfer can lead to shortlists that are less prone to gender disparities for at least three reasons: changes in the way candidates are evaluated in shortlisting (a differential evaluation mechanism), reduced importance of social networks and referrals used in shortlisting (a social influence mechanism), and changes in same-gender preferences in shortlisting (a homophily mechanism).

### **Differential Evaluation Mechanism**

The transfer of shortlisting from hiring managers to HR might affect the way candidates are evaluated by increasing expert knowledge in shortlisting and reducing opportunity costs for conducting such evaluations. Scholars commonly refer to differential treatment as one reason for gender disparities in employment; they suggest that organizational decision makers treat equally qualified individuals differently because of their gender (e.g., Reskin, 1998; Pager and Shepherd, 2008). This reasoning is based on the idea that intrapsychic factors, such as social categorization, in-group preference, or stereotyping, shape individual actions and behaviors that, in turn, cause disparate outcomes (Pager and Shepherd, 2008). Intrapsychic factors are expected to play out in organizational contexts in which evaluation processes face the challenge of subjective preferences interfering with decision making, and they often lead to a disadvantage for women (Bielby, 2000; Reskin, 2000). Decision makers might, for

example, use gender as a stereotypic indicator to infer the relative quality of candidates' future performances, as they might expect more-competent performances from men than from women in most settings (Berger et al., 1977; Correll and Ridgeway, 2003).

Given that an organization's expertise in recruitment and selection is typically concentrated within HR departments (e.g., Dobbin, 2009), HR professionals are more likely than hiring managers to be consistent in the way they approach responsibilities such as candidate selection (e.g., McGovern et al., 1997). This consistent approach makes HR more likely than hiring managers to commit to previously agreed-upon selection criteria prior to shortlisting candidates. Committing to selection criteria reduces ambiguity in decision making and avoids shifting bias (Uhlmann and Cohen, 2005; Correll, 2017), which means that HR departments are less likely than hiring managers to redefine the criteria for success in the job depending on a candidate's gender. In addition, HR professionals likely face lower opportunity costs than hiring managers do for spending time and effort on shortlisting candidates. Opportunity costs, in turn, likely contribute to the overall costs of candidate selection, that is, costs incurred to reduce uncertainty about candidates' quality (Podolny, 1993, 1994). While sorting through a large number of candidate profiles is a task layered on top of hiring managers' main professional responsibilities, it is one of HR's main responsibilities (e.g., McGovern et al., 1997). Prior evidence suggests that when these costs are high in evaluation situations, decision makers are more apt to rely on observable, stereotypic indicators of expected quality, such as gender (e.g., Botelho and Abraham, 2017).

### **Social Influence Mechanism**

The transfer of shortlisting from hiring managers to HR might reduce the application of specific hiring procedures that could otherwise contribute to reproducing the existing demographic composition of an organization's workforce. Prior research points out that disparate impact can cause gender disparities in employment; disparate impact occurs when groups are treated equally within a given set of organizational rules and procedures, but the rules and procedures still confer advantages to one group over another (Reskin, 1998). The organizational rules and procedures may not have explicit discriminatory content but nevertheless exacerbate disparities among groups because they are applied in a context with a pre-existing difference that affects minorities in a negative way (Small and Pager, 2020). The role of social networks in hiring is a case in point. Using networks in hiring, such as via employee referrals, is a long-established and taken-for-granted practice that is generally viewed as an efficient strategy for matching workers to employers (e.g., Granovetter, 1995; Fernandez, Castilla, and Moore, 2000). However, given that networks are often homophilous (e.g., McPherson, Smith-Lovin, and Cook, 2001), their use is likely to reproduce the company's existing demographic composition and exclude members of groups that are not already well represented (Braddock and McPartland, 1987).

There are at least three reasons that HR professionals might differ from hiring managers in their use of social networks and employee referrals when shortlisting candidates. First, as employers are often not able to observe candidates' quality (or productivity) directly, they might rely on easily observable signals, such as referrals, that are expected to correlate with quality (e.g.,

Fernandez, Castilla, and Moore, 2000). If HR professionals have lower opportunity costs than hiring managers do for shortlisting candidates, HR professionals might be less likely to rely on network status or referrals as an easily accessible but imprecise signal. Second, the information advantage theory of referrals argues that network hiring (and the use of referrals) creates an information flow between the employer and the candidate: the referrer enriches the candidate's information set about the job, and the referrer shares with the employer informal, difficult-to-access information about the candidate's quality, such as personality traits (Rees, 1966). Given that HR is plausibly more external to the job vacancy and the referrer, referrals might have a lower informational value for HR vis-à-vis hiring managers; the two groups likely differ in their access to the information passed on by the referrer. Third, and related, social networks commonly encourage trust and support (Kanter, 1977; Baron and Pfeffer, 1994), which makes a candidate's network status potentially more important for hiring managers, who will have to work with the hired candidates, than for HR.

### Homophily Mechanism

The transfer of shortlisting from hiring managers to HR might also lead to a reconfiguration of same-gender preferences. In-group preference is widespread (Tajfel and Turner, 1979) and may influence judgment (Baron and Pfeffer, 1994; Reskin, 2000). Prior research suggests that decision makers may prefer working with candidates of their own gender or rate in-group members more highly than members of other groups, which in turn might influence their evaluation of candidates (e.g., Gorman, 2005). Given that in most large organizations the gender composition of HR professionals is more dominated by women than it is among hiring managers, the transfer of shortlisting from hiring managers to HR might skew shortlists in favor of women by replacing hiring managers' same-gender preferences with HR's same-gender preferences.

## METHODS

### Introducing the Intervention and Data Access

This study is set within a leading technology-focused multinational corporation, Alpha, that globally changed its hiring process by transferring the shortlisting task from hiring managers to HR (the intervention). Alpha has a workforce of over 50,000 employees in over 60 countries. Its three key geographical regions are Europe, the Middle East, and Africa (EMEA); Asia and Australia; and the Americas (North and South America). We collected data by cooperating with Alpha's global HR team and used four data sources.

First, we gained detailed access to internal hiring data, which include key information about the hired candidate (e.g., gender, age, education, job family).<sup>3</sup> Second, we received access to a less comprehensive dataset about Alpha's job listings (see Online Appendix Table A1). This dataset corresponds to only a subset of all hirings that occurred at Alpha during our sample period (2,796 job listings from October 2018 to May 2019 across 49 countries), and we cannot match candidates' characteristics (e.g., gender or education) to the job-listing

---

<sup>3</sup> Given data availability restrictions, we have information only about employees' binary gender and cannot make any inferences about employees' actual gender identity.

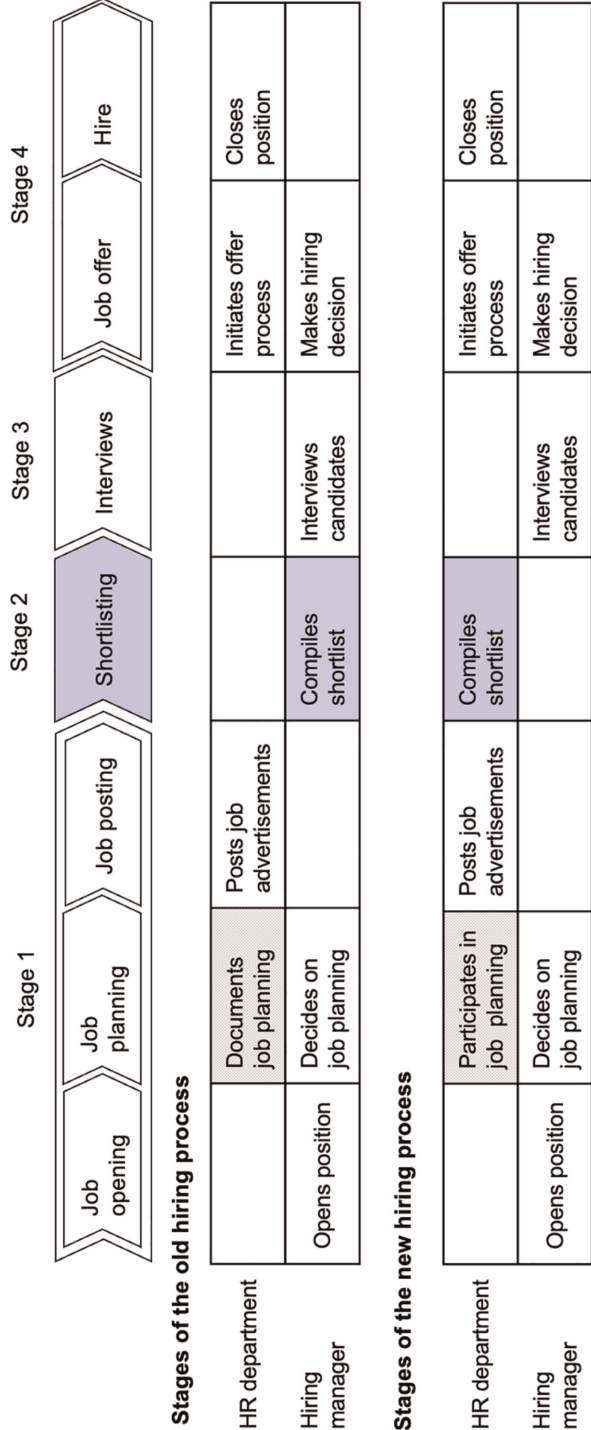


entries. Nevertheless, this second dataset has two useful features for our study: it provides additional information about Alpha's hiring process, including information about the duration of filling positions (time-to-fill), recruitment process (e.g., use of employee referrals), and candidate pools; and it provides the necessary information to test the intervention's effect on these additional job-listing characteristics. Third, we accessed the results of a regular company-wide survey that Alpha administered to its hiring managers about their overall work experiences (see Online Appendix Table A2 for information on the survey design).

Fourth, we conducted semi-structured interviews with members of Alpha's global HR team, HR departments, hiring managers, and newly hired candidates to learn more about their work experiences. For information on the sample and exemplary quotes, see Tables A3 and A4 in the Online Appendix. We conducted the semi-structured interviews throughout the adoption of the new hiring process to learn more about both hiring managers' and HR professionals' work experiences. One of the authors met regularly with representatives of the global HR team to discuss the advent of the intervention and was given access to internal company documentation. The interviews lasted, on average, 34 minutes, were conducted either in English or in the native language used in Alpha's headquarters, and spanned February to August 2019. We used a semi-structured interview format and asked about the interviewees' experiences with the hiring process in terms of (changes in) their respective roles and responsibilities, their satisfaction with the overall process, and differences between the old and the new hiring processes, when appropriate. Access to the interviewees was provided by Alpha's global HR team and depended both on interviewees' availability and on our attempt to capture voices from all three core geographic areas where Alpha does business. The interviews were conducted in private via video conferencing tools or telephone (ensuring that no one else could hear the interviewees' responses), recorded, and transcribed. We analyzed the transcribed interviews, using the MAXQDA qualitative software package.

**Alpha's new hiring process.** Alpha's new hiring process consisted of four stages: job opening, planning, and posting; shortlisting; interviewing; and offering/hiring (see Figure 1). The intervention introduced a key change in the second stage. Under the old hiring process, HR forwarded all candidate profiles to hiring managers, who reviewed the profiles and formed shortlists. Under the new process, HR was tasked with forming shortlists for hiring managers that consisted of a maximum of seven candidate profiles per job queue. To equip HR with the necessary information to perform their new role, the intervention also extended the first stage of the hiring process. Under the old process, job planning meetings primarily involved HR documenting hiring managers' job requirements; the new hiring process put more emphasis on hiring managers explaining their job requirements to HR. The intervention did not formally alter the last two stages of the hiring process. In the third stage of both the old and the new hiring process, shortlisted candidates were invited to interviews and were interviewed by hiring managers. After the interviews, in stage four, selected candidates received job offers and were hired. Under both the old and the new hiring

Figure 1. Old Versus New Hiring Process at Alpha: HR Departments’ and Hiring Managers’ Roles and Responsibilities





process, hiring managers decided which candidates to extend the job offers to, and HR initiated the job offer process before positions were closed.

**Staggered adoption of Alpha's new hiring process.** Alpha adopted the new hiring process in seven staggered treatment waves across its different country locations in 2018 and 2019. Important for our identification strategy, the starting date of all seven treatment waves and the assignment of individual country locations to treatment waves were set in advance (in 2017) by the global HR team. The intervention in treatment wave 1, for example, occurred in May 2018. This means that subsidiaries in countries assigned to treatment wave 1 started hiring under the new hiring process beginning on the first day of the respective month: May 1, 2018. The other six treatment waves were adopted in October 2018, November 2018, February 2019, April 2019, July 2019, and October 2019.<sup>4</sup> Table 1 (panel A) provides an overview of the different treatment waves and tabulates the assignment of countries to specific treatment waves.

**Intention behind Alpha's new hiring process.** Conversations with members of the global HR team indicated that the intervention aimed to address one key problem: avoiding unnecessary delays (reducing time-to-fill) that would influence candidates' experience in a negative way, especially for top candidates who had been jumping ship given overlong hiring periods.<sup>5</sup> Important for our identification strategy, these conversations also clarified that the staggered adoption of the new hiring process was not related to any concerns about gender disparities and that the assignment of countries to treatment waves did not follow any particular pattern. Table 1 (panel B) shows that the treatment waves are geographically heterogeneous, that is, made up of countries of different geographical regions (for example, India and South Africa were both in treatment wave 1). In Online Appendix Figure A1, we also document that Alpha's key geographical regions do not overlap with the timing of the intervention; we find country locations assigned to different treatment waves in each of Alpha's key geographical regions. Figure A1 further shows that the timing of the intervention was unrelated to Alpha's average time-to-fill for its job positions per country.

<sup>4</sup> As our data access ended in September 2019, the last treatment wave has no post-intervention observations.

<sup>5</sup> Although it is beyond the scope of this article, we used the job listings dataset to document that the intervention did not succeed in decreasing Alpha's time-to-fill (Online Appendix Table A1). The results reveal a positive and marginally significant intervention effect suggesting an increase in Alpha's time-to-fill. At the same time, we do not find that Alpha's recruitment practices (size of candidate pools, use of recruitment agencies or employee referrals) or the contract types offered (permanent position, full-time position, additional compensation benefits) changed with the intervention. This reduces concerns that the increase in time-to-fill is a result of broader changes in Alpha's hiring practices. We also probed whether the increase in time-to-fill relates to Alpha's propensity to hire women. Since our job listings dataset does not include any candidate characteristics, we constructed a variable that reflects the average time-to-fill at the country level and merged it with our original hiring dataset (for hirings from 49 countries). Models 3 to 6 (panel B, Online Appendix Table A1) show that our treatment effect is virtually unaffected when we include time-to-fill as an additional control variable. We also do not find that time-to-fill is able to explain cross-sectional variation in our treatment effect.

**Table 1. Overview of the Adoption of the New Hiring Process\***

Panel A. New hiring process across treatment waves							
	Treatment wave 1	Treatment wave 2	Treatment wave 3	Treatment wave 4	Treatment wave 5	Treatment wave 6	Treatment wave 7
October 2017	0	0	0	0	0	0	0
November 2017	0	0	0	0	0	0	0
December 2017	0	0	0	0	0	0	0
January 2018	0	0	0	0	0	0	0
February 2018	0	0	0	0	0	0	0
March 2018	0	0	0	0	0	0	0
April 2018	0	0	0	0	0	0	0
May 2018	1	0	0	0	0	0	0
June 2018	1	0	0	0	0	0	0
July 2018	1	0	0	0	0	0	0
August 2018	1	0	0	0	0	0	0
September 2018	1	0	0	0	0	0	0
October 2018	1	1	0	0	0	0	0
November 2018	1	1	1	0	0	0	0
December 2018	1	1	1	0	0	0	0
January 2019	1	1	1	0	0	0	0
February 2019	1	1	1	1	0	0	0
March 2019	1	1	1	1	0	0	0
April 2019	1	1	1	1	1	0	0
May 2019	1	1	1	1	1	0	0
June 2019	1	1	1	1	1	0	0
July 2019	1	1	1	1	1	1	0
August 2019	1	1	1	1	1	1	0
September 2019	1	1	1	1	1	1	0

Panel B. Additional information about countries assigned to treatment waves							
Example countries assigned to treatment waves	India, South Africa	China, Taiwan	Italy, United States	Malaysia, Poland	Belgium, Morocco	Korea, Argentina	Japan, United Kingdom
Number of countries assigned to treatment waves	3	3	6	6	22	12	10

\* In Panel A, 0 represents the pre-treatment period (hiring under the old hiring process), and 1 represents the post-treatment period (hiring under the new hiring process). In Panel B, we share only a few example countries in order to guarantee Alpha’s anonymity.

Empirical Methodology

**DiD design.** To examine the effects of the intervention on hiring outcomes, we used a generalized DiD approach. We estimated the following model at the individual (newly hired employee) level:<sup>6</sup>

$$y = \beta_1 + \beta_2 Post \times Treatment + \sum \beta_j Controls_j + \sum \beta_k Fixed\ Effects_k + \varepsilon$$

(1)

<sup>6</sup> We followed prior studies in (financial) economics (e.g., Giroud, 2013; Christensen, Hail, and Leuz, 2016) and chose the unit of analysis based on the level of the expected treatment effect instead of the aggregate level of the intervention.

where  $y$  is the dependent variable *Men* (gender of the focal newly hired employee, which equals 1 for men and 0 for women), and  $Post \times Treatment$  is a dummy variable that equals 1 for all new employees hired under the new process and 0 otherwise. *Controls* is a vector of control variables. We controlled for newly hired employees' human capital: *Age* and *Education*. On the organizational level, we controlled for three major occupational groups at Alpha (*Management*, *Operations*, and *Experts*) as well as job families, distinguishing among five categories (*Engineering and R&D*, *Manufacturing*, *Customer services*, *Compliance and communication*, and *Sales and marketing*). *Fixed effects* include month-by-region fixed effects and treatment wave fixed effects. They also subsume the separate indicators for *Post* and *Treatment*. The main coefficient of interest is  $\beta_2$ , which measures the effect of the intervention. We estimated equation (1) using a linear probability model and heteroscedasticity-robust standard errors clustered at the month-by-region level.<sup>7</sup>

**Idea of the identification strategy.** The following example illustrates the identification strategy. Alpha adopted the new hiring process for its subsidiaries in Italy in November 2018 (see Table 1, panel B). For these subsidiaries, the  $Post \times Treatment$  variable is 0 for all employees hired before November 2018 under the old hiring process and 1 for all employees hired after November 1, 2018 under the new hiring process. To measure the effect of the intervention in Italy on the gender composition of newly hired employees, one could simply compare the share of newly hired men before and after November 1, 2018.

However, such a simple before–after comparison might be confounded by an overall positive time trend concerning the awareness of gender disparities in employment.<sup>8</sup> To account for such a time trend, the DiD approach uses hirings during 2018 from subsidiaries located in other European countries that fall into a different treatment wave (such as Belgian subsidiaries, which adopted the new hiring process in April 2019) as control observations. We then compared the difference in the share of newly hired men in Italian subsidiaries before and after November 1, 2018 with the difference in the share of newly hired men in control

<sup>7</sup> We assumed that hiring decisions within a multinational corporation are likely spatially correlated (e.g., driven by a seasonal hiring demand across key regions or the occurrence of company-wide hiring freezes). To accommodate variations in spatial correlations across Alpha's key strategic regions, we used month-by-region clustering, where the region variable reflects Alpha's three key geographic regions. This choice also accommodates the primary trade-off when selecting the level of clustering: selecting a small number of large groups for clustering (e.g., at the regional level) to accommodate more appropriately the various dependencies in the data versus selecting a cluster level with a modest number of groups that more likely meets the homogeneity restriction, such as clustering by month-by-country (Petersen, 2009; Gow, Ormazabal, and Taylor, 2010; Conley, Gonçalves, and Hansen, 2018). In addition, our cluster choice results in a nested fixed effects structure (for the month-by-region fixed effects). The non-nested treatment wave fixed effects are of less concern, as the number of observations to estimate such a fixed effect is sufficiently large and the number of treatment waves relatively low (Conley, Gonçalves, and Hansen, 2018: 1175). In Figure 3, we show that our results hold when we used different fixed effects and cluster specifications.

<sup>8</sup> For instance, there might have been a positive time trend in Italy and the EU (where Alpha is headquartered) concerning the awareness of gender disparities as reflected by regulatory initiatives like the EU-wide disclosure mandates for public interest entities about sustainability and (gender) diversity aspects, which came into effect in 2018 (e.g., Fiechter, Hitz, and Lehmann, 2022). Such a positive time trend might have led over time to changes in the labor market (and thus changes in the supply side of women) and to changes in Alpha's demand for women, which in turn might have decreased the share of Alpha's newly hired men.

countries before and after November 1, 2018. The difference between the two differences is the estimated treatment effect.

**Threats to the identification strategy.** The staggered adoption of the new hiring process across many countries mitigates concerns that typically arise in studies of single intervention events. It is unlikely, for example, that shocks unrelated to the intervention affected the gender composition of Alpha's hirings and coincided in time with the adoption pattern of the seven treatment waves (see Table 1). For the findings to be spurious or to reflect other shocks, a large set of countries would have needed to experience differently timed local (demand or supply) labor shocks (within short intervals of one to three months). Alpha's global HR team would have needed to predict these shocks and assign different adoption dates to countries in 2017 (eight months prior to the start of the first treatment wave and over two years prior to the start of the last treatment wave). These adoption dates would then have needed to coincide with country-specific future changes in hirings of women in 2018 and 2019. Such predictive ability by the global HR team seems highly unlikely. Likewise, given the predetermined treatment dates, it seems highly unlikely that the timing of the seven treatment waves unintentionally or incidentally correlated with local (demand or supply) labor shocks that, in turn, affected Alpha's propensity to hire more women in 2018 and 2019.

Nevertheless, to mitigate concerns that differently timed labor shocks confound the findings, we included month-by-region fixed effects in all DiD regressions. These fixed effects control for month-specific time trends in the key geographical regions in which Alpha operates. To illustrate the usefulness of month-by-region fixed effects, suppose that Alpha purposefully assigned all Asian countries to one specific treatment wave while randomly allocating the remaining countries to the other six treatment waves. Month-by-region fixed effects exclude the possibility that time-variant factors in Asia (such as changes in labor market conditions or regulatory changes in 2018 and 2019), which might have led to the clustered treatment wave assignment and the specific timing of the treatment waves in the first place, confound the treatment estimation as an omitted factor. In the presence of month-by-region fixed effects, the treatment effects in such an example would be solely estimated through the remaining six treatment waves with randomly assigned countries.

## Sample

The final sample consists of 8,750 externally hired employees over a 24-month period (October 2017 to September 2019) across more than 60 countries. For these newly hired employees, we have information on whether each employee was hired under the old or the new hiring process in their respective country. Using this sample, we tested the effect of the intervention on gender disparities in hiring outcomes. Table 2 provides summary statistics and gives insights into the composition of the newly hired employees by gender. Overall, 64 percent of all newly hired employees are men. Alpha is a men-dominated company; the majority of its newly hired workforce during our sample period was men across all but one job family (the exception being *Compliance and communication*, with

**Table 2. Summary Statistics\***

	Summary Statistics of Key Variables (Across Time)				Gender Distribution	
	2017 mean	2018 mean	2019 mean	Overall mean	Number of all newly hired men	Percentage of newly hired men
Men	0.65	0.65	0.64	0.64		
Age	34.43	33.36	33.25	33.43		
<b>Educational degree</b>						
Master's degree	0.11	0.15	0.12	0.13	1,147	58.59%
Bachelor's degree	0.27	0.31	0.34	0.32	2,784	64.69%
Others	0.62	0.54	0.54	0.55	4,819	65.76%
<b>Major occupational groups</b>						
Occupation management	0.09	0.08	0.07	0.07	645	73.02%
Occupation operations	0.31	0.24	0.24	0.25	2,180	62.71%
Occupation prof. and expert	0.60	0.68	0.69	0.68	5,925	64.20%
<b>Job families</b>						
Engineering and R&D	0.19	0.22	0.20	0.21	1,820	72.25%
Manufacturing	0.15	0.13	0.11	0.12	1,068	73.60%
Customer services	0.27	0.30	0.36	0.32	2,844	66.60%
Sales and marketing	0.22	0.20	0.19	0.20	1,724	58.93%
Compliance and communication	0.17	0.15	0.14	0.15	1,294	48.76%
<b>Key geographical regions</b>						
EMEA	0.42	0.34	0.35	0.35	1,897	64.84%
Asia	0.40	0.45	0.42	0.43	3,810	65.80%
Americas	0.18	0.21	0.23	0.22	3,043	62.59%
Number of observations	953	4,097	3,700	8,750		

\* *Men* equals 1 for newly hired employees who are men and 0 for newly hired employees who are women. *Age* is measured in years. For education, the category *Others* includes employees with a vocational background, other country-specific degrees, or no degree. *Occupation management* denotes newly hired employees working in quality, project, or general management positions such as team leaders, supervisors, heads, or managers. *Occupation operations* denotes newly hired employees working in technical or administrative operation positions, such as service technicians or administrative support. *Occupation prof. and expert* denotes newly hired employees working as professionals in positions such as various kinds of engineers, finance professionals, IT professionals, and R&D professionals. The sample includes all externally hired employees between October 2017 and September 2019 with the necessary data available. For example, we exclude observations with duplicated identifiers in the original dataset (33) and without sufficient candidate information available such as information about entry date, age, or location of employment (848).

48.76 percent newly hired men). Table A5 in the Online Appendix provides correlation statistics for the control variables and the dependent variable *Men*.

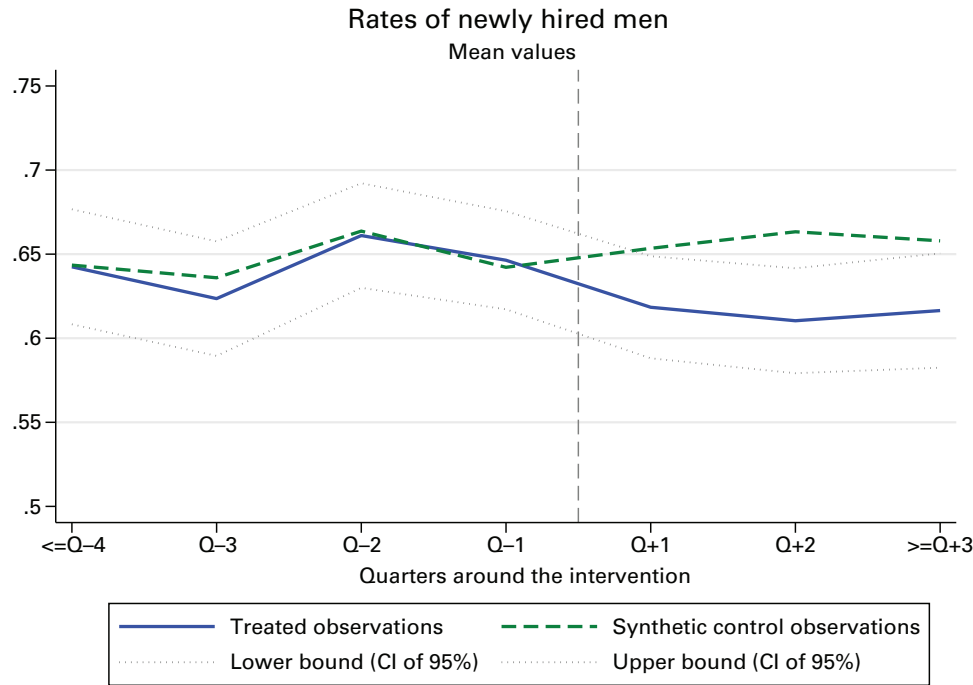
## RESULTS

### Graphical Analyses of the Rate of Newly Hired Men

We started by plotting the share of newly hired men around the intervention. The blue solid line in panel A of Figure 2 represents treated observations, whereas the green dashed line represents a synthetic control group (e.g., Abadie, Diamond, and Hainmueller, 2010). The synthetic control group is estimated based on a stacked dataset from which we can separately identify all control observations for each treatment wave. The stacked dataset is the sum of wave-specific datasets that include observations from the focal treatment waves

Figure 2. Gender Disparities Around the Intervention

Panel A. Graphical analysis\*



Panel B. Univariate (DiD) analysis

	Rates of newly hired men (mean)		
	Treated	Synthetic control <sup>†</sup>	Unmatched control
Pre-intervention (old hiring process)	64.3%	64.6%	68.3%
Post-intervention (new hiring process)	61.5%	65.8%	69.1%
Difference	-2.8%	+1.2%	+0.8%
Univariate DiD: based on synthetic control	-4%		
Univariate DiD: based on unmatched control	-3.6%		

\* In Panel A,  $\leq Q-4$  pools all observations that fall into Q-7, Q-6, Q-5, and Q-4. Likewise,  $\geq Q+3$  pools all observations that fall into Q+5, Q+4, and Q+3. We do so to avoid having sample imbalances on both sides of the timeline affect the graphical analysis. Online Appendix Table A9 documents these imbalances (e.g., Q+5 includes observations from only 3 of 52 potential treatment countries).

<sup>†</sup> Along several dimensions, our setting differs from the standard application of synthetic control group matching (e.g., Abadie and Gardeazabal, 2003; Abadie, Diamond, and Hainmueller, 2010; Abadie, 2021). For example, given our sample requirements, we can neither exploit a long time series of pre-treatment periods nor assess the predictive power based on sufficiently long training and validation periods.

but exclude post-observations from later-treated or already-treated control waves (see Online Appendix Table A6). For the pre-intervention period (through Q-1), the pattern in this figure shows that by construction, the shares of newly hired

men in both the treatment and synthetic control groups move in parallel. In the post-intervention period (after Q–1), the figure documents a decrease in the share of men hired for the treatment group relative to the synthetic control group.

Panel B of Figure 2 summarizes the distribution of newly hired men across the pre- and post-intervention periods. It shows that under the old hiring process, 64.3 percent of all newly hired employees are men. Under the new hiring process, this number falls to 61.5 percent. This results in a pre–post change of – 2.8 percentage points for the treatment group. In comparison, the same pre–post change for the synthetic control group amounts to +1.2 percentage points. At face value, these numbers translate into a univariate DiD effect of – 4.0 percentage points (or – 6.2 percent). Using an unmatched control group, we observe a comparable effect of – 3.6 percentage points (or – 5.6 percent). Collectively, the graphical analyses in panel A of Figure 2 and the mean comparisons in panel B indicate a decrease in the share of newly hired men for treated hirings post-intervention.

### DiD Analyses

In Table 3, we estimated two regression specifications: in Model 1, a regression with month-by-region fixed effects and treatment wave fixed effects but without the individual- and organizational-level control variables; and in Model 2, a fully specified regression as defined in equation (1). The findings offer three main insights. First, the estimated average treatment effect  $Post \times Treatment$  is negative and significant for each specification (coefficients of – 0.056 and – 0.059 and  $p$ -values < 0.05), which suggests that the intervention decreased the share of newly hired men. Second, moving from the first model to the fully specified model does not seem to materially affect the economic significance of the results, which suggests low sensitivity of the treatment effect toward the choice of control variables. Third, the results are also economically meaningful, indicating a relative decrease in the share of newly hired men of around 5.9 percentage points, or 9.2 percent, based on the sample mean of treated observations pre-intervention. In Figure 3, we plotted  $Post \times Treatment$  for 14 DiD specifications that vary in terms of fixed effects and clustering choice. The statistical and economic significance levels of  $Post \times Treatment$  in these specifications (with an average coefficient of – 0.056 and an average  $p$ -value of 0.018) are in line with the findings in Table 3.

### Treatment Effect Heterogeneity

Staggered DiD designs provide a weighted average of many different treatment effects (depending on the number and weights of the different treatment waves). For example, one of these treatment effects stems from the fact that already-treated observations act as control units for later-treated observations. In the case of treatment effect heterogeneity—if the already-treated observations exhibit particular post-treatment dynamics vis-à-vis later-treated observations—the overall weighted treatment effect might be biased (for an overview see



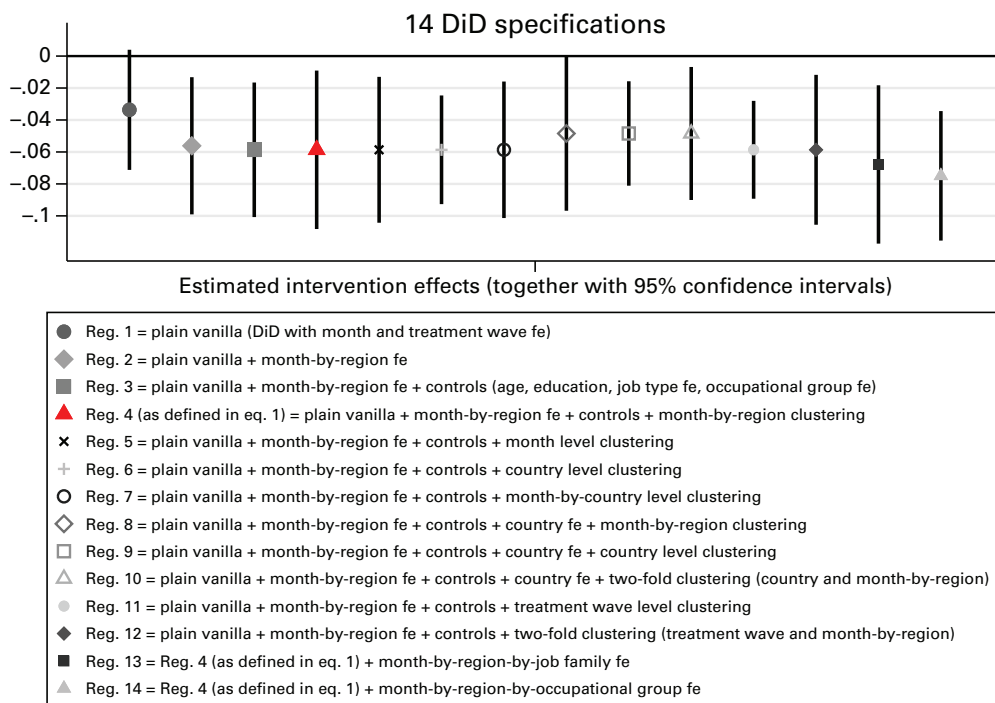
Table 3. Baseline DiD Results\*

	Dependent Variable: Men	
	Model 1	Model 2
<b>Treatment effect</b>		
Post × Treatment	−0.056* (0.026)	−0.059* (0.025)
<b>Control variables</b>		
Age		0.002* (0.001)
Master’s degree		−0.108*** (0.019)
Bachelor’s degree		−0.051*** (0.015)
Occupation management		0.123*** (0.021)
Occupation operations		−0.055* (0.027)
Engineering and R&D		0.145*** (0.020)
Manufacturing		0.189*** (0.026)
Customer services		0.119*** (0.018)
Compliance and communication		−0.088*** (0.024)
<b>Fixed effects</b>		
Month-by-region fixed effects	Included	Included
Treatment wave fixed effects	Included	Included
Adj. R <sup>2</sup>	0.016	0.055
N	8,750	8,750

+  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; two-tailed.  
\* Standard errors (in parentheses) are one-way clustered at the month-by-region level. Month-by-region clusters comprise 72 groups (24 months times the three regions Alpha does business in). Our inferences are robust to alternative clustering choices such as one-way clustering at the month-by-country level, month level, or country level (see Figure 3). To account for the extensive fixed-effects structure and increase the readability of the economic magnitudes of the treatment effects, we estimate the regression models based on a linear probability model (for a similar approach, see, for example, Kowaleski, Sutherland, and Vetter, 2020; Egan, Matvos, and Seru, 2019). Statistical inferences remain virtually unchanged when we use logit or probit specifications.

Baker, Larcker, and Wang, 2022).<sup>9</sup> Recent work in econometrics proposes several solutions to mitigate this potential bias. One central feature of these solutions is to modify the set of control observations in a way that ensures the estimation process is not contaminated by treatment effect heterogeneity. We followed Cengiz et al. (2019) and Deshpande and Li (2019) and estimated the DiD analysis based on a stacked regression sample. As noted, the idea is to create event- or, in our case, wave-specific datasets that include observations from the focal treatment waves but exclude post-observations from later-treated or

<sup>9</sup> The post-intervention downward trend in our treatment observations in Figure 2 might already mitigate some concerns about treatment effect heterogeneity, e.g., that a reverse trend (hire more men) in early-treated countries, and hence negative weights in the computation of the weighted average treatment effect, drives our documented intervention effect in Table 3.

**Figure 3. Alternative Cluster Choice, Fixed Effects, and DiD Specifications\***

\* This figure provides point estimates of the intervention effects based on 14 DiD regressions (together with 95 percent confidence intervals). The point estimates are generated by estimating variants of the following regression model:

$$Y = \beta_0 + \beta_1 Post \times Treatment + \sum \beta_j Controls_j + \sum \beta_k Fixed\ Effects_k + \varepsilon$$

The regressions are based on the final sample (N=8,750). *Post* × *Treatment* effects are significant at a 5-percent level with two exceptions: the *p*-value of *Post* × *Treatment* in regression model 1 is 0.079 and in regression model 8 is 0.050.

already-treated control waves. The final DiD regression estimation is then performed on a stacked dataset (N=23,903), which includes all wave-specific datasets and aligns the timeline of these datasets in a relative way (for example, *t*–1 or *t*+1 relative to the intervention date). In a final step, to account for the stacked nature of the final dataset, it is necessary to saturate the fixed effects and cluster dimensions with indicators of the specific stacked datasets. Online Appendix Tables A6 and A7 (panel A) provide insights into the construction of the stacked dataset.

Consistent with the baseline findings in Table 3, the stacked regression estimator in panel B of Online Appendix Table A7 is negative and significant with a comparable effect size and significance level (coefficient of –0.076 and *p*-value < 0.01).<sup>10</sup> In Online Appendix Table A8, we followed an alternative approach to account for

<sup>10</sup> In untabulated tests, we did not find evidence that the coefficient estimates in the stacked DiD estimation are significantly different from the coefficient estimate of our baseline effect as documented in Table 3 (with *p*-values between 0.53 and 0.77, indicating insignificant differences in the effect sizes).

treatment effect heterogeneity and implemented the Callaway and Sant'Anna (2021) estimator. The intervention effect is negative and marginally significant ( $p$ -value  $< 0.10$ ) with a comparable effect size of  $-0.069$ . As suggested in this literature (e.g., Sant'Anna and Zhao, 2020; Baker, Larcker, and Wang, 2022), we also estimated both the stacked DiD and the Callaway and Sant'Anna (2021) estimators without time-varying covariates as a benchmark. The estimated treatment effects are again negative and significant ( $p$ -values  $< 0.05$ ), with comparable effect sizes mitigating concerns that estimation issues arise from the inclusion of covariates.

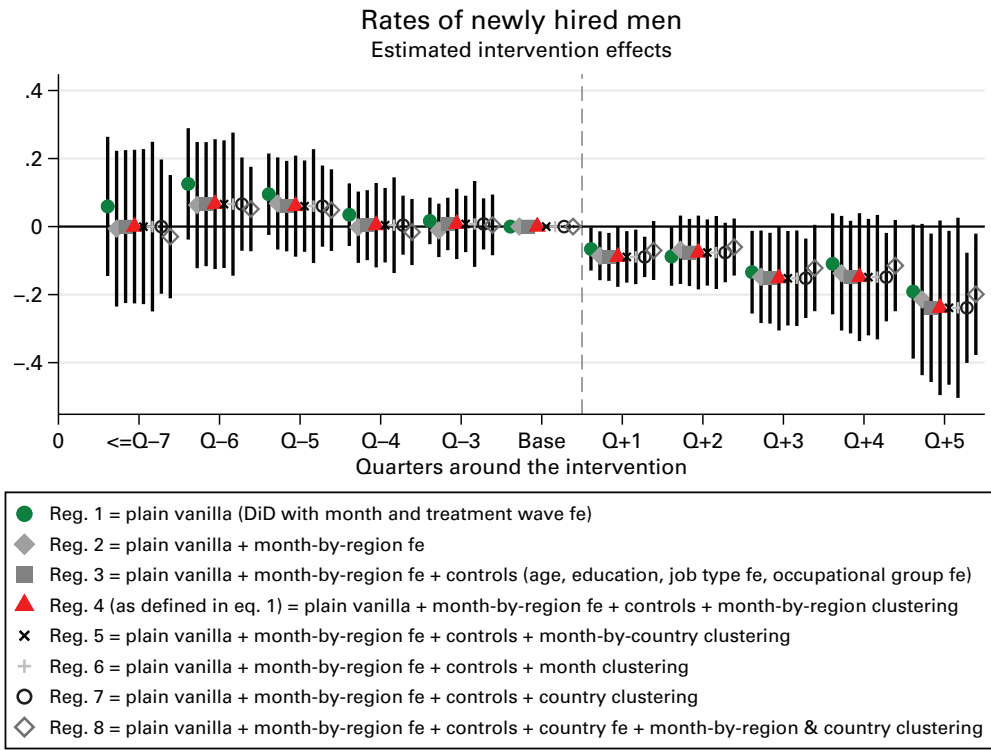
### Parallel Trends Assumption

A key assumption underlying DiD designs is the parallel trends assumption, which is not directly testable (Angrist and Pischke, 2009; Atanasov and Black, 2016; Wing, Simon, and Bello-Gomez, 2018). This assumption requires that the trends (or changes) in the outcome variable across treated and control observations are the same absent the treatment. We performed three analyses to gauge the plausibility of the parallel trends assumption: pre-trends, timing of the staggered intervention, and covariate balance.

**Pre-trends.** Prior research commonly gauges the plausibility of the parallel trends assumption by examining the pre-trends in the outcome variable across treatment and control observations (e.g., Atanasov and Black, 2016). Significant pre-trends would indicate a violation of the parallel trends assumption. We followed this literature and estimated an event study specification of equation (1) with quarter-specific treatment effects (for a similar approach, see Giroud, 2013; Christensen, Hail, and Leuz, 2016; Darendeli et al., 2022). Figure 4 illustrates the results by plotting the point estimates of all quarterly treatment effects for eight different DiD specifications. Online Appendix Table A9 provides the sample characteristics of the event study design (panel A) and tabulates the full set of regression results (panel B). The graphical inspection of Figure 4 reveals three main insights. First, the treatment effects in the pre-intervention period are insignificant across all specifications. This finding indicates that the outcome variable (*Men*) does not trend ahead of the actual intervention, which is consistent with the parallel trends assumption. Second, moving from a model without control variables to different versions of a fully specified model (with different sets of fixed effects specifications) does not seem to affect the insignificant pre-trends. This result indicates a low sensitivity of the pre-trend analyses toward the choice of control variables and fixed effects. Third, the treatment effect of the intervention turns out sufficiently sharp (Atanasov and Black, 2016): in Q+1, the treatment effect is significantly negative in seven out of eight specifications.<sup>11</sup>

<sup>11</sup> In the online appendix, we performed two additional sets of tests to assess pre-trends in the data. First, we estimated placebo interventions in a sample that excludes observations from the actual post-intervention period (for a similar approach, see Lehmann, 2019). The results show insignificant treatment effects across all placebo treatment dates (Figure A2), which are consistent with the insignificant pre-trends in Figure 3. Second, we estimated placebo interventions in a sample that excludes observations from the actual pre-intervention period. The results show insignificant treatment effects across all placebo intervention dates (Figure A3). The findings are again consistent with Figure 3, suggesting that the treatment effect of the actual intervention appears to be sufficiently sharp. They further mitigate concerns that the treatment effect is driven by a more secular time trend in the data.

Figure 4. Event Study Specifications\*



\* The point estimates are generated by estimating variants of the following regression model (without subscripts):

$$Y = \beta_0 + \beta_1 Treated_{Q-7} + \beta_2 Treated_{Q-6} + \beta_3 Treated_{Q-5} + \beta_4 Treated_{Q-4} + \beta_5 Treated_{Q-3} + \beta_6 Treated_{Q+1} + \beta_7 Treated_{Q+2} + \beta_8 Treated_{Q+3} + \beta_9 Treated_{Q+4} + \beta_{10} Treated_{Q+5} + \sum \beta_l Controls_l + \sum \beta_k Fixed\ Effects_k + \varepsilon$$

As we omitted the DiD estimators for Q-1 and Q-2, the first six months prior to the intervention serve as the benchmark period. Thus, each point estimate reflects a quarter-specific treatment effect in comparison to the benchmark period. The inferences hold when we used only Q-1 as the benchmark period. Q+1 includes the month with the intervention (month t0) and the two following months (months t+1 and t+2). The other quarters are defined accordingly. The shapes indicate point estimates, and the lines represent 95 percent confidence intervals. We chose a quarterly instead of a monthly time frame to allow for sufficiently large sample sizes. The results also hold with monthly time specifications.

**Timing of the staggered intervention.** Given the staggered nature of the treatment assignments, the parallel trends assumption in our setting requires that the timing of the adoption of the new hiring process be independent of factors that might affect gender disparities in hiring outcomes. Our pre-trend analysis in Figure 4 provides some reassurance that the timing might indeed be independent of these factors, as an endogenous treatment timing would often result in pre-trends. However, an endogenous treatment timing might also affect the trends in the post-period, which is not directly testable. For example, the parallel trends assumption might be violated in the post-period if the timing of the intervention lined up with the labor force supply of women or demand shocks. However, given that Alpha’s global HR team set up the starting dates of all

seven treatment waves in advance (in 2017), it is less likely that this timing (un)intentionally correlates with local (demand or supply) labor shocks in 2018 and 2019 that, in turn, affected Alpha's propensity to hire more women. Nevertheless, to shed light on the timing of the staggered intervention, we followed prior research and modeled the timing decision (e.g., Carlin, Umar, and Hanyi, 2023). We modeled the timing decision at the country level since the timing of the treatment waves also occurred at the country level (see Table 1). We further focused on country-level labor market characteristics that might correlate with gender disparities in employment.<sup>12</sup>

In Figure 5, we show graphically that the timing of the treatment waves is unrelated to many gender-related labor market variables, including the country-level United Nations *Economic gender social norms index* (GSNI), which measures how social beliefs obstruct gender equality in the context of work and employment; the country-level rate of women's participation in the labor force (*Women labor participation rate*); and the country-level rate of the population of women with at least some secondary education (*Women education rate*). We also do not find that the timing decision correlates with broader measures of country-level institutional environments: the *Regulatory quality* country index and the *Rule of law* country index. The latter findings mitigate concerns that Alpha might have timed the intervention in anticipation of broader economic shocks to its supply or demand of talented women. Overall, our findings illustrate that country locations with similar (gender-related labor) market characteristics are assigned to different treatment waves, whereas country locations with different characteristics are included in the same treatment waves.

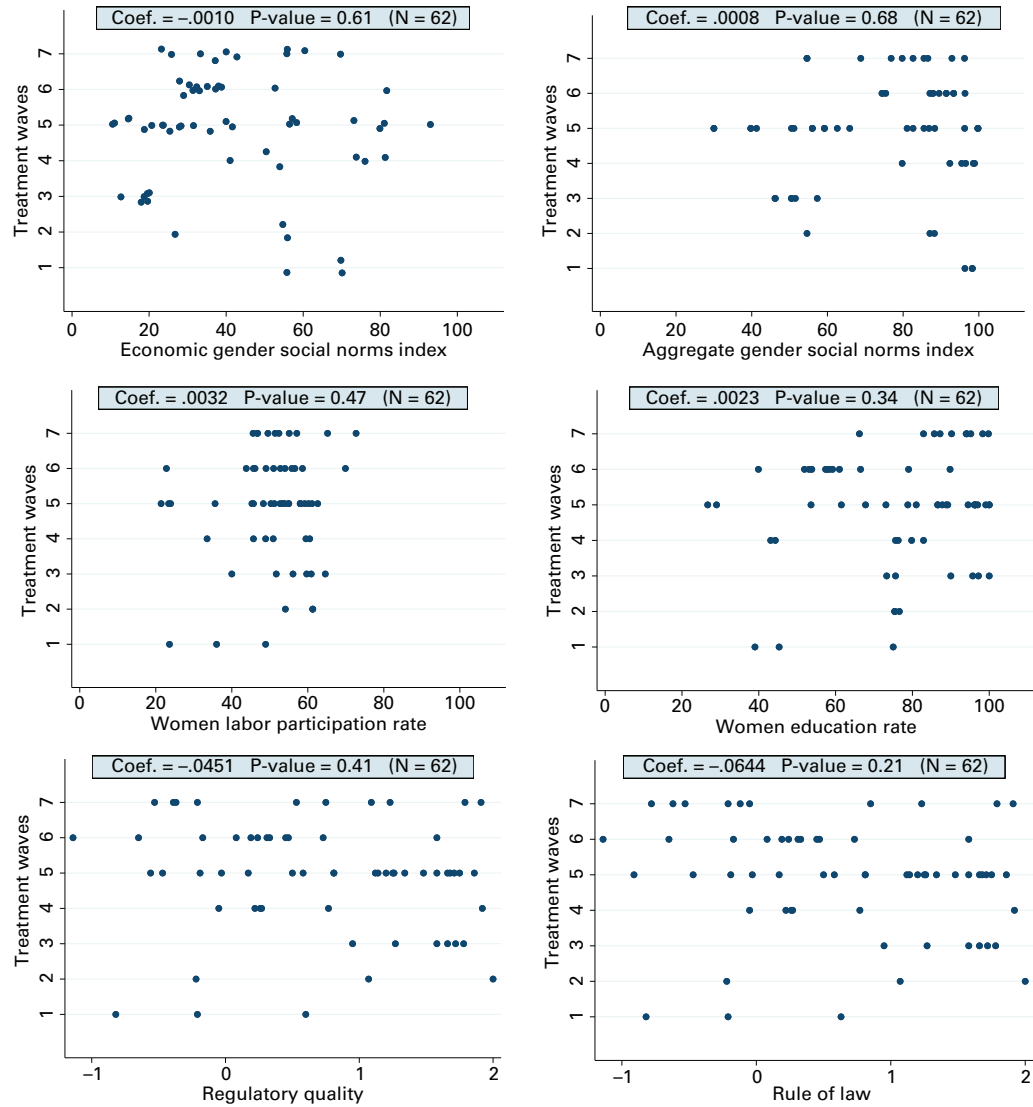
**Covariate balance.** A common expectation of DiD designs is that the treatment assignment should result in a reasonable covariate balance between treated and control observations (e.g., Atanasov and Black, 2016). Although, in a strict sense, a covariate balance (in levels) is not an identifying assumption of a DiD design, it can increase the likelihood that the trends (changes) in the outcome variable across treated and control observations are the same absent the treatment. To assess the covariate balance, we used a stacked dataset (N= 23,903), which includes all wave-specific datasets and in which we can separately identify all control observations for each treatment wave (see panel A of Online Appendix Table A7). Panel A of Online Appendix Table A10 reports the covariate balance across the unmatched and matched (weighted) samples. The findings suggest that some covariates are significantly different across treatment and control groups.

Given these imbalances, we employed entropy balancing weights, which impose a perfect covariate balance across the included characteristics (e.g., Hainmueller, 2012; McMullin and Schonberger, 2020; Darendeli et al., 2022). This approach addresses concerns of treatment group selection based on observable characteristics. If observable and unobservable characteristics are related, matching based on entropy balancing weights provides a way to assess the severity of potential selection effects (Altonji, Elder, and Taber, 2005; Bode,

---

<sup>12</sup> If the timing is unrelated to these characteristics, it is less plausible to assume that it correlates with confounding shocks (e.g., changes in the talent supply of women or changes in Alpha's hiring demand for women talent), as these shocks commonly cluster across countries with similar (labor market) characteristics.

Figure 5. Do Country Characteristics Predict the Timing of the Intervention?\*



\* This figure shows scatterplots of the timing of the country-wise adoption of the new hiring process. Both GSNi measures are provided by the United Nations; the aggregate measure reflects levels across work and employment, education, politics, and physical integrity. *Women labor participation rate*, which is based on % of ages 15 and older, and *Women education rate*, which is based on % of ages 25 and older, are obtained from the 2020 Human Development Perspectives Report by the United Nations. *Regulatory quality* and *Rule of law* are widely used country-level indices for differences in the institutional environment of firms. Both indices are taken from Kaufmann, Kraay, and Mastruzzi (2009). Reported coefficients and p-values are generated by estimating variants of the following Poisson regression model (unit of observation=country):  $Treatment\ Wave = \beta_0 + \beta_1 Country\ Characteristic + \varepsilon$ . Regression results are tabulated in Online Appendix Table A13.

Singh, and Rogan, 2015; Christensen et al., 2017). The results in panel A of Online Appendix Table A10 (Models 2 and 3) show that the entropy balancing weights fully absorb the remaining covariate imbalance across treatment and

control observations. In panel B of Online Appendix Table A10, we then re-estimated the stacked DiD regressions based on entropy balancing weights. The resulting weighted treatment effects of  $-0.073$  (with a  $p$ -value  $< 0.01$ ) and  $-0.058$  (with a  $p$ -value  $< 0.01$ ) are comparable to the unweighted treatment effect of  $-0.076$ , as well as to the baseline treatment effect of  $-0.059$  ( $p$ -value  $< 0.05$ ) in Table 3.<sup>13</sup>

Given that the changes in the intervention effect after we applied weights are only marginal, any potential selection on unobservable characteristics would either need to have little correlation with the included observable characteristics, such as job type, candidate age, or candidate education, or be quite large to explain the entire intervention effect (Altonji, Elder, and Taber, 2005). To provide further context for the covariate imbalance, we show that our covariates (e.g., age composition, educational composition, and job profile composition of newly hired employees) did *not* change with the adoption of the new hiring process for the treatment groups relative to the control groups (Online Appendix Table A11). These findings suggest similar trends in our covariates across treatment and control groups, which increases the likelihood that the trends in the outcome variable (*Men*) across treated and control observations are the same absent the treatment.

## INSIGHTS INTO EXPLANATIONS AT THE SHORTLISTING STAGE

So far, the findings suggest that the intervention reduced the share of men hired into the organization. Although the tests in the previous section corroborate a causal interpretation of the treatment effect, they do not allow us to pin down likely explanations. Given that the intervention is situated at the shortlisting stage of the hiring process, at least three shortlisting-related explanations could explain our documented treatment effect: changes in the way candidates are evaluated (a differential evaluation mechanism), reduced importance of social networks and referrals (a social influence mechanism), and a change in same-gender preferences (a homophily mechanism). Given the different implications of each explanation, particularly in the context of informing extant research about how to potentially alleviate gender disparities in hiring, it is important to empirically distinguish these explanations. To do so, our empirical strategy involves additional quantitative tests that exploit cross-sectional variation in the treatment effect as well as the use of alternative outcome variables. To triangulate the quantitative findings, we also draw insights from the survey that Alpha administered to its hiring managers, and turn, when appropriate, to the qualitative data to shed light on hiring managers' decision making compared to that of HR professionals.

### Quantitative Insights

**Differential evaluation mechanism.** We argue that the intervention is likely to change the way candidates are evaluated in shortlisting due to differences in expert knowledge and opportunity costs between HR and hiring managers. We

<sup>13</sup> In untabulated tests, we did not find evidence that the coefficient estimates in these different DiD specifications are significantly different compared to the coefficient estimate in our baseline effect, as documented in Table 3 (with  $p$ -values between 0.61 and 0.99, indicating insignificant differences in the effect sizes).



thus examined whether the treatment effect provides plausible cross-sectional variations consistent with gender disparity effects due to differential evaluation. To do so, we leveraged our worldwide setting and used the GSNI of the United Nations, which measures how social beliefs obstruct gender equality at the country level (Mukhopadhyay, Rivera, and Tapia, 2019; UNDP, 2019). The index comprises four sub-indices: political, educational, physical integrity, and economic. The latter is particularly interesting as it measures gender-related beliefs, biases, and prejudices in the context of work and employment (hereafter “gender-related social beliefs”). For the economic sub-index, the survey questions consist of two statements: “Men should have more right to a job than women” and “Men make better business executives than women do.”<sup>14</sup>

While we acknowledge the simplicity of the survey questions and the construction of the GSNI, we perceive this index as a useful proxy to shed light on whether cross-country variations in gender-related social beliefs might explain cross-sectional variations in our treatment effect. We would expect a stronger treatment effect in countries with higher levels of social beliefs that obstruct gender equality. The idea is that HR’s higher expert knowledge in and lower opportunity costs for candidate evaluation will especially matter in evaluation contexts that are more exposed to gendered social beliefs. In contexts less exposed to beliefs obstructing gender equality, we expect the intervention to be less effective. In the latter case, both sets of decision makers (independent of their expert knowledge and opportunity costs) have a lower likelihood of letting gender factor into their shortlisting decisions, as gender is generally less likely to serve as an observable stereotypical indicator of candidate quality.<sup>15</sup>

We first included the GSNI index on work and employment as an additional control variable in our baseline regression model (Model 1, panel A, Table 4). Consistent with the intuition behind this proxy, the coefficient estimate of GSNI turns out positive but only marginally significant ( $p$ -value < 0.10). This finding suggests that—after we control for applicant and job characteristics—relatively more men are hired in countries in which gender disparities in terms of social beliefs and cultural norms are high. At the same time, the inclusion of the GSNI as an additional control variable in the main model does not alter the baseline intervention effect, which remains negative and significant (coefficient of  $-0.059$ ,  $p$ -value < 0.05). We then extended equation (1) by interacting the treatment effect ( $Post \times Treatment$ ) with GSNI to test for cross-sectional variations in the treatment effect. The findings are tabulated in Model 2 (panel A

<sup>14</sup> The use of the GSNI is further motivated by prior research, which suggests that gender becomes salient in contexts in which gender disparities in cultural norms are high (Dore, 1983; Mukhopadhyay, Rivera, and Tapia, 2019). In these high-salience contexts, decision makers are expected to be more apt to let gender factor into their evaluation decisions (e.g., Ridgeway, 1997; Wagner and Berger, 1997). Empirical evidence on gender bias in evaluation processes, such as investment recommendations (Botelho and Abraham, 2017) or networking among entrepreneurs (Abraham, 2020), corroborates the effect of gender salience on decision making.

<sup>15</sup> While we acknowledge that the predicted treatment effect variation along high vs. low GSNI countries is likely not unique to our first explanation (differential evaluation), we perceive it as less plausible that the other two explanations (social influence and homophily) would easily predict such a variation. For example, decision makers might rely on social networks and referrals regardless of whether they themselves are prone to gender-related social beliefs entering their decision making (Small and Pager, 2020). Likewise, same-gender preferences are commonly expected to be a distinct form of intrapsychic factors and should occur regardless of whether decision makers are prone to gender-related social beliefs.

**Table 4. Insights into Plausible Explanations\***

Panel A. Differential evaluation mechanism			
	Dependent Variable: Men		
	Model 1	Model 2	Model 3
Country-level moderator	Economic GSNi		Overall GSNi
Post × Treatment	−0.059* (0.025)	−0.046+ (0.024)	−0.059* (0.025)
Moderator	0.001+ (0.001)	0.002* (0.001)	0.001** (0.001)
Post × Treatment × Moderator		−0.002* (0.001)	−0.002+ (0.001)
Control variables	Included	Included	Included
Month-by-region fixed effects	Included	Included	Included
Treatment wave fixed effects	Included	Included	Included
Adj. R <sup>2</sup>	0.056	0.056	0.057
N	8,750	8,750	8,750
Panel B. Social influence mechanism			
	Dependent Variable		
	Connected candidate	Men	
	Model 1	Model 2	Model 3
Post × Treatment	−0.003 (0.008)	−0.059* (0.025)	−0.060* (0.025)
Connected candidate		−0.012 (0.035)	−0.031 (0.063)
Post × Treatment × Connected candidate			0.027 (0.083)
Control variables	Included	Included	Included
Month-by-region fixed effects	Included	Included	Included
Treatment wave fixed effects	Included	Included	Included
Adj. R <sup>2</sup>	0.061	0.055	0.055
N	8,750	8,750	8,750
Panel C. Homophily mechanism			
	Dependent Variable: Men		
	Model 1	Model 2	Model 3
Post × Treatment	−0.054+ (0.027)	−0.053+ (0.027)	−0.053+ (0.027)
Differential gender rate		−0.059* (0.028)	−0.059+ (0.031)
Post × Treatment × Diff. gender rate			−0.000 (0.041)
Control variables	Included	Included	Included
Month-by-region fixed effects	Included	Included	Included
Treatment wave fixed effects	Included	Included	Included
Adj. R <sup>2</sup>	0.054	0.055	0.055
N	7,471	7,471	7,471

(continued)

Table 4. (continued)

Panel D. Changes in the demand for specific candidates or job types

Dependent Variable	Age	Master's degree	Occupation operations	Engineering and R&D
	Model 1	Model 2	Model 3	Model 4
Post × Treatment	−0.151 (0.422)	0.017 (0.021)	−0.004 (0.025)	0.031 (0.036)
Remaining control variables	Included	Included	Included	Included
Month-by-region fixed effects	Included	Included	Included	Included
Treatment wave fixed effects	Included	Included	Included	Included
Adj. R <sup>2</sup>	0.220	0.098	0.378	0.220
N	8,750	8,750	8,750	8,750

Panel E. Changes in job offers made

	Dependent Variable		
	Permanent contract		Men
	Model 1	Model 2	Model 3
Post × Treatment	−0.001 (0.021)	−0.059* (0.025)	−0.060+ (0.033)
Permanent contract		−0.008 (0.026)	−0.009 (0.027)
Post × Treatment × Permanent contract			0.001 (0.033)
Control variables	Included	Included	Included
Month-by-region fixed effects	Included	Included	Included
Treatment wave fixed effects	Included	Included	Included
Adj. R <sup>2</sup>	0.692	0.055	0.055
N	8,750	8,750	8,750

+  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; two-tailed.  
\* Standard errors (in parentheses) are one-way clustered at the month-by-region level. In panel A, to facilitate the interpretation of the interacted DiD estimator, we mean-centered the moderating variables. In panel C, to ease the interpretation of the interacted treatment effect, we defined *Differential gender rate* as a dummy variable (1=country-level *Differential gender rate* is above the sample mean, 0 otherwise). The results hold when we used *Differential gender rate* as a numerical variable instead. In panel D, we did not include *Men* as a control variable to avoid a potential “bad control” problem (e.g., Wing, Simon, and Bello-Gomez, 2018). The results are virtually identical when we added *Men* as an additional control variable. In panel E, the variable *Permanent contract* is defined as a dummy variable (1=the newly hired employee’s contract is permanent, 0=the contract is temporary). Control variables and fixed effects are the same as those used in the main specification (Table 3).

of Table 4) and show that the interaction effect ( $Post \times Treatment \times GSNI$ ) is negative and significant ( $p\text{-value} < 0.05$ ). This finding suggests that the intervention effect on gender disparities in hiring outcomes is stronger in countries where gender-related social beliefs are high. The result also holds when we use the overall GSNI index (Model 3).

**Social influence mechanism.** Next, we examined whether the intervention changes the use of social networks and referrals during shortlisting. Although it

is inherently difficult to empirically identify the use of these hiring practices, we exploited data about the sourcing of newly hired employees. We constructed the variable *Connected candidate*, which denotes whether a newly hired employee had some connection to Alpha prior to their hiring, either through a re-entry into Alpha after having had an outside position or a transfer from a company that was a former parent company of Alpha. Although it is a rather coarse proxy, we used it to measure how connected the newly hired employee was at the time of hire and, by implication, how likely it is that the employee was identified through Alpha's internal networks. The findings are tabulated in panel B of Table 4. Most important, we do not find a significant intervention effect when using *Connected candidate* as the outcome variable (Model 1). This suggests that the likelihood of hiring connected candidates does not change with the intervention.

Although our comprehensive hiring dataset does not include any further information about the sourcing of candidates, we were able to leverage the second, less comprehensive job listings dataset, which provides information about the use of employee referral programs (ERPs). ERPs constitute a common practice in hirings in which employers draw on the social networks of their employees. In Online Appendix Table A1 (panel C, Model 3), we estimated the intervention effect by using the ERPs as the dependent variable. If the intervention reduced the importance of social networks and referrals in the hiring process, we would expect to see a decrease in the use of ERPs as a sourcing strategy post-intervention. However, we do not find any significant changes in the use of ERPs post-intervention.

**Homophily mechanism.** Finally, we examined whether the intervention leads to a change in same-gender preferences during shortlisting. We constructed the variable *Differential gender rate*, which is defined—at the country level—as the pre-treatment difference between the share of women HR professionals and the share of women hiring managers. The mean value of this variable is 36 percent, which suggests that prior to the intervention, HR departments employed 36 percentage points more women than were employed as hiring managers. We used this variable to examine whether the treatment effect exhibits plausible cross-sectional variation consistent with a same-gender preferences effect. In essence, same-gender preferences would predict stronger treatment effects in countries with higher values of the *Differential gender rate* variable. Online Appendix Table A12 provides a simplified two-country example to illustrate the idea behind this variable. We extended equation (1) by interacting the treatment effect ( $Post \times Treatment$ ) with *Differential gender rate*. The findings are tabulated in panel C of Table 4 and show that the interaction effect  $Post \times Treatment \times Differential\ gender\ rate$  remains insignificant. This suggests that potential changes in the preferences to shortlist women are less likely to serve as an alternative explanation for our findings.

## Qualitative Insights

**Insights from interviews with hiring managers and HR.** To triangulate the quantitative findings, we turned next to the qualitative data to shed light on hiring managers' decision making compared to that of HR professionals. The qualitative data reveal clear differences in the shortlisting process between hiring

managers and HR departments. The interviews show that, pre-intervention, hiring managers felt pressed for time: “In the past, I used to have about 500 CVs to work through . . . and you know, my desk was full” (Hiring manager C7, EMEA) or “I don’t have the time to look through 50 applications for a position” (Hiring manager C2, Asia). To economize on these costs, hiring managers allowed subjective preferences to interfere with their shortlisting decisions:

For instance, when I look at a CV, I’m not only—I’m trying to see from a paper, what quality this person has. So—it sounds stupid to another person—but what sticks out for me . . . is like activities and interests. So, if he [a potential candidate] was competing in any sports on a national level that means to me that this person has got stamina. He’s focused. He knows what he wants and will work hard for it. These are the top things that I look for. If I get people with exactly the same qualifications, this guy will have the edge. (Hiring manager C7, EMEA)

These observations are consistent with prior research (e.g., Rivera, 2012, 2015), which also indicates that hiring managers’ evaluation decisions can be made under time constraints and are often susceptible to subjective preferences. In addition, we found a few instances that point toward hiring managers’ reliance on their own networks for candidate selection pre-intervention:

What I always missed was for [HR] to take my hand and show me suitable candidates. I’ve always had to go through my colleagues [other hiring managers] who gave me tips on good candidates. We looked at them, we interviewed them, and then picked someone. (Hiring manager C2, Asia)

This reliance went so far that, in one pre-intervention case, HR was merely informed that a candidate had been “found” (Global HR team B2, EMEA).

In contrast, and informed by job planning meetings, HR shortlisted candidates by following pre-specified and seemingly more objective criteria:

So, in the beginning we have a discussion about the job requirements. I know, for example, we’re looking for a specialist in women’s health. The hiring manager clarifies that we need a radiography and mammography specialist—and these criteria are non-negotiable. Candidates must have both qualifications. So, for me that’s easy as it eliminates candidates who only have radiography but no mammography experience, for example. And then the further pre-screening would depend on whatever other criteria the manager requires. It could be a matter of three to five years of experience, and we have to see whether it’s negotiable or not. That goes down to quite specific technical questions important for the job. (HR B4, EMEA)

This in-depth knowledge of job requirements was necessary for HR to understand the candidate profiles—in terms of required candidate knowledge, experience, and/or abilities—and to make informed shortlisting decisions. The interviews also revealed that HR professionals perceived the evaluation of candidates to be an important part of their jobs, which highlights the potentially lower opportunity costs for HR departments vis-à-vis hiring managers: “We’re the experts, we know the criteria we need to pay attention to. . . . If I know directly from the hiring manager who we are looking for, what the exact job requirements are, what the person will have to know—there is no way I would sort somebody out [who] the hiring manager would have perceived as qualified” (HR B2, EMEA). Consistent

with the two groups experiencing different opportunity costs of candidate evaluation, survey evidence further reveals that hiring managers acknowledged the importance of HR's support in the post-intervention period: "[HR] helped me run the process while I was able to focus on the day-to-day business . . ." (Hiring manager, Americas).

**Insights from surveying hiring managers.** Insights from the survey that Alpha administered to its hiring managers show that they were satisfied overall with HR's shortlisting efforts post-intervention. In fact, 82 percent of the 341 surveyed hiring managers in the post-intervention period acknowledged the new role of HR and/or were generally satisfied with the HR department's work (panel A of Table A3). This satisfaction rate increased by 22 percentage points from 60 percent in the pre-intervention period. Hiring managers' high satisfaction rates with HR's shortlists is noteworthy for at least two reasons. First, the high satisfaction rate could be consistent with the notion that hiring managers largely appreciated HR taking over shortlisting, which reduced their own time and effort spent on shortlisting (opportunity costs). Second, it might also suggest that hiring managers were largely happy with the work being done by HR and the quality of the shortlists they produced.

### **Summary of the Findings: Changes at the Shortlisting Stage of the Hiring Process**

Our tests in this section shed light on the question of whether the intervention triggered changes at the shortlisting stage of the hiring process, which in turn led to changes in hiring outcomes. While our tests cannot causally isolate an individual explanation, the findings are largely consistent with the differential evaluation mechanism: the intervention led to reduced gender disparities in hiring outcomes given a change in candidate evaluation at the shortlisting stage through an influx of HR expert knowledge in shortlisting and through reduced opportunity costs related to that process. Especially our qualitative data reveal instances consistent with this notion. The interviews support the idea that busy hiring managers prioritize their own day-to-day business at the potential expense of consistency in candidate shortlisting and thus are susceptible to subjective preferences in decision making. We do not find evidence that changes in the role of social networks and referrals (a social influence mechanism) could explain our results, nor do we find that the intervention triggered a change in same-gender preferences for shortlist candidates (a homophily mechanism).

### **INSIGHTS INTO EXPLANATIONS AT OTHER STAGES OF THE HIRING PROCESS**

Although the intervention is situated at the shortlisting stage of the hiring process, we cannot rule out that the documented intervention effect can also be explained by the intervention leading to (behavioral) changes at *other* stages of the hiring process. We thus go beyond the shortlisting stage in this section and discuss potential explanations related to changes at the application, job planning, and hiring stages.

### Supply-Side Dynamics at the Application Stage

First, the transfer of shortlisting from hiring managers to HR might trigger supply-side reactions contributing indirectly to gender disparities in hiring outcomes. If, for example, women fear discrimination in the hiring process, they might be less willing to apply. Studies find that after being rejected in a context dominated by men, women are less likely to reapply within that context (Brands and Fernandez-Mateo, 2017; Fernandez-Mateo, Rubineau, and Kuppawamy, 2022). Research further suggests that women place greater weight than men do on the perceived fairness of recruitment and selection processes (Brands and Fernandez-Mateo, 2017). Baron et al. (2007) associated the presence of HR professionals with formalized organizational practices, which are likely to signal processual fairness to candidates. In the same vein, Mocanu (2022) showed that the adoption of formal screening tools in the hiring process leads to changes in the applicant pool, in particular by attracting women. Thus, the transfer of shortlisting to HR might signal more fairness and equal treatment, especially to women, and result in higher rates of women in the applicant pools and eventually more women being hired.

We examined whether the intervention changes candidates' perceived fairness perceptions about Alpha's hiring process, thus attracting more women to apply. We again leveraged the job listings dataset as it includes information about the size of the candidate pool for each job listing. In Online Appendix Table A1 (panel C, Model 1), we estimated the intervention effect in this dataset, using the size of candidate pools as the dependent variable. All else equal, the perceived fairness argument would imply an increase in the size of candidate pools concurrent with the intervention as more women would be motivated to apply. We do not find any significant changes in the size of candidate pools post-intervention.

In addition, the interviews with candidates hired both under the old and new hiring processes do not indicate any differences in the ways candidates talked about their recruitment experience. In all cases, job offers were made relatively fast (Candidate D4, Asia; Candidate D3, Americas), candidates felt listened to and taken seriously (Candidate D5, Asia), and they were generally happy with their job offers (Candidate D4, Asia). Considering these data, we do not find any instances of candidates learning about and responding differently to the new hiring process. This is consistent with conversations with Alpha's global HR team in which members told us that they did not communicate the change in hiring process externally or internally beyond the involved HR professionals and hiring managers.

### Demand-Side Dynamics at the Application Stage

Second, we shed light on the possibility that Alpha's demand for certain job profiles, or its ability to attract talent, changed concurrently with the intervention. To explain our findings, a potential change in Alpha's demand would need to overlap in time with the adoption pattern of the seven treatment waves and be correlated with gender disparities in hiring outcomes. These requirements raise the bar for demand shocks that are unrelated to the intervention, such as Alpha adopting a concurrent strategy to expand in or withdraw from various markets or technologies, which in turn affects its demand for or its ability to attract certain



candidates. However, it is plausible that the intervention itself might change Alpha's demand for or ability to attract certain candidates. HR's shortlisting might, for example, change the likelihood of certain positions being filled (e.g., for difficult-to-fill jobs), which in turn might affect the gender composition of Alpha's hirings post-intervention.<sup>16</sup>

We examined the treatment effect on various alternative outcome variables such as candidates' education level or job openings in engineering and R&D. The idea was to test whether candidate characteristics other than gender might change concurrently with the intervention. The results in panel D of Table 4 show that the age composition, educational composition, and job profile composition of newly hired employees do not change with the adoption of the new hiring process. The results thus do not suggest that Alpha's demand for or its ability to attract certain candidates or job profiles changes concurrently with the intervention.

### **Other Stages of the Hiring Process (Job Planning Stage and Job Offer Stage)**

Finally, transferring the shortlisting from hiring managers to HR might affect hiring managers' behavior at other stages of the hiring process. The intervention, for example, likely increases the interaction between HR professionals and hiring managers via the job planning meetings. The latter could have clarified the actual evaluation criteria, which would have made them more salient and reduced the risk of shifting bias (Correll, 2017) in later stages of the hiring process for both parties. Alternatively, and given HR's involvement, hiring managers might feel more accountable to HR in terms of the predefined evaluation criteria as laid out in their job planning meetings or, more broadly, in their overall decision making throughout the hiring process. Prior research suggests that accountability, the expectation that one can be called on to justify one's actions to others (Lerner and Tetlock, 1999), can reduce the scope of subjective assessment (Castilla, 2015).

The interview data show a few instances in which HR professionals felt that the intervention introduced a new level of accountability for hiring managers: "I don't want to say that hiring managers feel like they're [on] the radar now but they know, okay, there is somebody who will ask, there is somebody in [HR] who will look into the system, and if I do not proceed as I'm supposed to they'll rap my knuckles" (HR B2, EMEA). However, hiring managers themselves were more agnostic about the changes. No hiring managers we interviewed talked about a change in accountability vis-à-vis HR, substantial changes in the hiring process other than HR taking over shortlisting, or an increase in HR's influence at other stages of the hiring process, such as in the job planning meetings. It is also important to highlight again that the intervention was not related to any

---

<sup>16</sup> It is questionable whether a higher likelihood of filling difficult-to-fill positions could explain an increase in newly hired women. Difficult-to-fill positions are more likely to require high degrees of specialization and expertise such as in engineering and technology, which are more likely jobs dominated by men (e.g., Cardador, 2017). Given Alpha's positioning as a technology-focused organization, it is plausible to assume that, on average, high-potential candidates are more likely to be men than women. Men are, for example, more likely to have the relevant education background and to select themselves into these kinds of occupations (e.g., Correll, 2001; Barbulescu and Bidwell, 2013).

concerns about gender disparities and that no DEI-related recruiting initiatives were launched during the study window. This mitigates the risk of gender-related changes in less-formal hiring practices surrounding the intervention.

Although we do not have quantitative data about the individual hiring stages, we have some information about the job offer stage in terms of the type of contract issued to newly hired employees. We know whether the contract accepted by newly hired employees was a temporary or permanent contract. Admittedly, this information provides only a very limited snapshot into the job offer stage. However, it allows us to mitigate concerns that HR taking over shortlisting—and with that a potential increase in accountability—might have increased hiring managers' propensity to offer permanent contracts, especially to women, which in turn might have affected gender disparities in offers accepted post-intervention. Panel E of Table 4 documents an insignificant intervention effect when we used *Permanent contract* as the outcome variable (Model 1). Considering our second, less comprehensive job listings dataset (Online Appendix Table A1, panels C and D), we do not find any pre–post intervention differences in Alpha's recruitment practices (e.g., use of ERPs) or offered contracts (e.g., type of contract and additional compensation benefit).

### Summary of the Findings: Changes at Other Stages of the Hiring Process

While we acknowledge the shortcomings of our data to comprehensively test the explanations as laid out in this section, we do not find evidence that supply-side reactions (candidates changing their application-related behavior) or demand-side reactions (Alpha changing its demand for specific job or candidate types) can explain our documented intervention effect. The results also do not suggest that the increase in women hired can be easily explained by the intervention leading to changes in job requirements or increased accountability for hiring managers.

## DISCUSSION

We examined an intervention that transferred the shortlisting of candidates from hiring managers to HR, and our main analyses established that the share of women hired into the organization increased after the intervention. Several mechanisms might produce this result. As the intervention was situated at the shortlisting stage of the hiring process, we first examined whether changes at the shortlisting stage could lead to changes in hiring outcomes. We distinguished three potential mechanisms at the shortlisting stage: the differential evaluation, social influence, and homophily mechanisms. Our empirical data are consistent with the differential evaluation mechanism: the intervention led to a change in candidate evaluations given HR professionals' higher expert knowledge in and lower opportunity costs for shortlisting candidates. Our data are not consistent with changes in the role of social networks and referrals (a social influence mechanism) or the intervention triggering a change in same-gender preferences to shortlist candidates (a homophily mechanism). To examine whether the intervention resulted in (behavioral) changes at other stages of the hiring process, we tested explanations for the intervention effect at the application, job planning, and job offer stages and did not find that these mechanisms are very plausible.

### Who Is Shortlisting?

Research on the role of hiring in generating gender disparities provides evidence that the hiring context is prone to disparate outcomes (e.g., Petersen and Saporta, 2004; Fernandez and Mors, 2008; Fernandez and Campero, 2017; Leung and Koppman, 2018). Our study contributes to this research by isolating the role of key decision makers in shaping gender disparities in hiring outcomes. The key implication we draw from our research is that *who* is shortlisting candidates may influence the demographic composition of the workforce. By isolating the *who* in shortlisting, we can zoom in on the differences between key decision makers, enabling us to derive implications for alleviating gender disparities in hiring.

A first implication of our findings could be that organizational practices that early on allocate responsibility to decision makers with high HR-related expert knowledge or that provide training to enhance this knowledge might reduce gender disparities in hiring. What appears to contribute to more women entering the organization in our setting is decision makers' HR-related expert knowledge and opportunity costs at an early stage of evaluation in the hiring process. Our qualitative evidence, for example, shows that hiring managers approached their shortlisting responsibilities in a less structured and more intuitive way, compared to HR. These observations corroborate prior research that indicates a lack of appropriate training and competences to handle HR responsibilities among managers, who take on such responsibilities in addition to their main jobs (McGovern et al., 1997; Nehles et al., 2006).

A second implication of our findings might be that organizations can reduce gender disparities in hiring by establishing or reinforcing organizational measures that encourage decision makers to place greater priority on their HR-related activities in recruitment, selection, or evaluation. Our qualitative evidence broadly corroborates the image of busy hiring managers, who prioritize their own full-time professional work at the potential expense of candidate shortlisting. This is consistent with Rivera's (2015: 1354) qualitative findings, which show that hiring managers face a trade-off between "time, effort, and evaluative rigor," especially as they need to balance candidate evaluation with their full-time professional work. Our findings are also consistent with a case study on HR practices showing that managers are generally not incentivized to take over HR-related responsibilities (McGovern et al., 1997). A formal institutionalization of HR tasks could make these responsibilities an inherent part of hiring managers' jobs, instead of extra tasks that are layered on top of their main responsibilities (McGovern et al., 1997). Formal incentives would signal the importance an organization attaches to HR tasks executed by hiring managers.

A third implication of our findings might be that organizational practices do not necessarily have to target explicit diversity issues or goals to reduce potential gender disparities. Prior research examines the impact of organizational practices that are specifically designed to reduce disparities (e.g., Reskin and McBrier, 2000; Kalev, Dobbin, and Kelly, 2006; Abendroth et al., 2017). This literature provides insights into, for example, the ways that managerial responsibility, transparency, and accountability can reduce disparities in pay and promotion contexts (Kalev, Dobbin, and Kelly, 2006; Castilla, 2015); the (limited) effectiveness of diversity initiatives (Kalev, Dobbin, and Kelly, 2006); and the effects of formalized HR practices, such as hiring procedures or written performance evaluations, on reducing disparities in managerial staffing and pay (Reskin and

McBrier, 2000; Abendroth et al., 2017). The intervention in our study was not aimed at reducing disparities, and the fact that it increased the share of women entering the organization suggests that organizational practices do not necessarily have to target explicit diversity issues or goals, such as programs targeting managerial bias or organizational structures allocating explicit responsibility and accountability for diversity goals (e.g., Kalev, Dobbin, and Kelly, 2006; Castilla, 2015). This finding is particularly important, as the success of many organizational practices targeting explicit diversity issues or goals often depends on the acceptance of such practices by the respective decision makers, who often contest them (e.g., Kalev, Dobbin, and Kelly, 2006; Dobbin, Schrage, and Kalev, 2015).

### **Organization-External Decision Makers**

Our findings help contextualize the role of organization-external decision makers: those who are situated outside the organization, such as employment agencies or executive search firms. The few studies that distinguish different sets of decision makers in the hiring process commonly focus on the interplay between organization-external and organization-internal decision makers. These studies suggest that organization-external decision makers who identify and evaluate candidates will use gender as a criterion to rank candidates for jobs in the attempt to anticipate their client firms' (the organization-internal decision makers') preferences (Fernandez-Mateo and King, 2011; Fernandez-Mateo and Fernandez, 2016); these first screeners will try to cater to the gender-based preferences of organization-internal decision makers who have the final say in hiring. These prior findings may not seem to align with our theorizing, as employment agencies or executive search firms are experts in HR-related activities and presumably have low opportunity costs for HR services, which constitute their main business. We suggest that different dynamics may be at play depending on whether first screeners of candidates are external or internal to the organization. When first screeners are external and are contracted by organizational clients, they focus on maintaining good relationships (Fernandez-Mateo and King, 2011), which means trying to anticipate clients' preferences. In contrast, internal first screeners generally do not face the same contractual incentives and might approach the screening task with greater independence. Our qualitative evidence broadly corroborates that the HR department seemed to act as an independent first set of screeners, in some cases even going against hiring managers' preferences. Further research is warranted to shed light on the independence of decision makers as a potential boundary condition when scholars examine their role in generating and constraining gender disparities in hiring outcomes.

### **Devolution of HR Responsibilities**

Our findings and theorizing also have broader implications for the literature on the devolution of HR responsibilities to managers. In recent decades, we have seen more and more HR responsibilities being transferred from HR professionals to managers (Purcell and Hutchinson, 2007; Steffensen et al., 2019). In many organizational settings, managers are directly involved in the active execution of operational HR activities that concern their own employees. Despite the

consensus that managers are critical to the recognition and integration of HR practices within organizations, our understanding of managers' impact on these HR practices is still limited (Steffensen et al., 2019). Research shows that transferring HR activities to managers can have both beneficial effects, such as efficiency and cost savings (e.g., Renwick, 2003; Kulik and Perry, 2008), and adverse effects, such as the occurrence of increased role stressors and higher workloads (Maxwell and Watson, 2006). Against this backdrop, one interpretation of our findings could be that the devolution of HR responsibilities to managers may induce adverse effects for workforce diversity. But our theorizing might also offer a more subtle interpretation, which is that, at least in the context of diversity, it may not matter who takes on HR responsibilities as long as decision makers are properly trained and incentivized.

### Limitations and Future Outlook

Our findings are subject to several limitations. First, while we find some evidence consistent with an increase in expert knowledge and reduction of opportunity costs as the most plausible explanation for our findings, our analyses cannot ultimately isolate an individual explanation. Since our intervention effect is measured at the point of hire and we lack more-direct data on the gender composition of shortlists, we cannot exclude that concurrent changes at other stages of the hiring process might explain (part of) our findings as well. Although it is reassuring that our various tests do not detect changes in Alpha's hiring process pre-post intervention (e.g., size of candidate pools, the use of ERPs, contract design), the tests capture only snapshots of the overall hiring process and do not account for more subtle and informal changes, such as those triggered by increased interaction between HR and hiring managers in the job planning meetings.

Second, as with other field studies, the setting and sample features need to be considered when generalizing the findings and interpreting the effect size of the intervention. For example, our cross-sectional tests suggest that the intervention effect varies consistently across country locations and institutional differences in terms of gender-related social beliefs. We also caution about generalizing our findings to other types of disparities, such as age and race. Although our theoretical arguments could potentially be expanded to other forms of disparities, future research is warranted to examine these forms and their responses to similar organizational interventions.

Last, we caution that our study offers only a snapshot into an intervention at the point of hire; it does not give any indication as to whether employees hired under the new hiring process fare better in their subsequent organizational lives, compared to employees hired under the old hiring process. Nor do the findings show whether the new hiring process was more beneficial for the organization in terms of higher productivity levels and actual (team) performance. Research has found that similarity in demographic characteristics, including gender, between supervisors and employees can positively influence outcomes such as turnover within work teams (e.g., O'Reilly, Caldwell, and Barnett, 1989) and employees' organizational attachment (e.g., Tsui, Egan, and O'Reilly, 1992). Thus, more studies examining whether specific hiring practices produce satisfactory hires that are more productive and efficient (Cappelli, 2019), that is, tying an

individual's entry into an organization to subsequent organizational and individual outcomes, would yield additional important insights.

## Acknowledgments

We are extremely appreciative of associate editor Chris Rider for his invaluable guidance and deep engagement throughout the whole review process. We thank the three anonymous reviewers for their very constructive feedback and Joan Friedman and Ashleigh Imus for excellent copyediting expertise. This article has benefited from the comments of Emilio Castilla, Andrew Knight, Kamal Munir, and conference and workshop participants at the EGOS Colloquium, the Edinburgh University Paper Development Workshop, the AMJ Paper Development Workshop, the ASA Annual Meeting, the University of Goettingen, and the Amsterdam Business School. We also thank members of the Global HR team at Alpha for the time and insights they shared with us. We particularly would like to thank one member for excellent assistance in the data generation process, without whose efforts this project would not have been feasible. We unfortunately cannot thank that person by name here to ensure Alpha's anonymity. Almasa Sarabi is very grateful to the Theo and Friedl Schöller-Foundation for financial support.

## REFERENCES

- Abadie, A.  
2021 "Using synthetic controls: Feasibility, data requirements, and methodological aspects." *Journal of Economic Literature*, 59: 391–425.
- Abadie, A., A. Diamond, and J. Hainmueller  
2010 "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American Statistical Association*, 105: 493–505.
- Abadie, A., and J. Gardeazabal  
2003 "The economic costs of conflict: A case study of the Basque Country." *American Economic Review*, 93: 113–132.
- Abendroth, A.-K., S. Melzer, A. Kalev, and D. Tomaskovic-Devey  
2017 "Women at work: Women's access to power and the gender earning gap." *ILR Review*, 70: 190–222.
- Abraham, M.  
2017 "Pay formalization revisited: Considering the effects of manager gender and discretion on closing the gender wage gap." *Academy of Management Journal*, 60: 29–54.
- Abraham, M.  
2020 "Gender-role incongruity and audience-based gender bias: An examination of networking among entrepreneurs." *Administrative Science Quarterly*, 65: 151–180.
- Altonji, J. G., T. E. Elder, and C. R. Taber  
2005 "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools." *Journal of Political Economy*, 113: 151–184.

**Angrist, J., and J.-S. Pischke**

2009 *Mostly Harmless Econometrics*. Princeton University Press.

**Atanasov, V., and B. Black**

2016 "Shock-based causal inference in corporate finance and accounting research." *Critical Finance Review*, 5: 207–304.

**Azmat, G., and B. Petrongolo**

2014 "Gender and the labor market: What have we learned from field and lab experiments?" *Labour Economics*, 30: 32–40.

**Baker, A., D. Larcker, and C. Wang**

2022 "How much should we trust staggered difference-in-differences estimates?" *Journal of Financial Economics*, 144: 370–395.

**Barbulescu, R., and M. Bidwell**

2013 "Do women choose different jobs from men? Mechanisms of application segregation in the market for managerial workers." *Organization Science*, 24: 737–756.

**Baron, J. N., M. T. Hannan, G. Hsu, and O. Kocak**

2007 "In the company of women: Gender inequality and the logic of bureaucracy in start-up firms." *Work and Occupations*, 34: 35–66.

**Baron, J. N., and J. Pfeffer**

1994 "The social psychology of organizations and inequality." *Social Psychology Quarterly*, 57: 190–209.

**Berger, J., M. H. Fisek, F. Z. Norman, and M. Zelditch**

1977 *Status Characteristics and Social Interaction: An Expectation States Approach*. Elsevier.

**Bielby, W. T.**

2000 "Minimizing gender and racial bias." *Contemporary Sociology*, 29: 120–129.

**Bills, D., V. Di Stasio, and K. Gërxhani**

2017 "The demand side of hiring: Employers in the labor market." *Annual Review of Sociology*, 43: 291–310.

**Bode, C., J. Singh, and M. Rogan**

2015 "Corporate social initiatives and employee retention." *Organization Science*, 26: 1702–1720.

**Botelho, T. L., and M. Abraham**

2017 "Pursuing quality: How search costs and uncertainty magnify gender-based double standards in a multistage evaluation process." *Administrative Science Quarterly*, 62: 698–730.

**Braddock, J. H., and J. M. McPartland**

1987 "How minorities continue to be excluded from equal employment opportunities: Research on labor market and institutional barriers." *Journal of Social Issues*, 43: 5–39.

**Brands, R. A., and I. Fernandez-Mateo**

2017 "Leaning out: How negative recruitment experiences shape women's decisions to compete for executive roles." *Administrative Science Quarterly*, 62: 405–442.

**Breaugh, J. A.**

2013 "Employee recruitment." *Annual Review of Psychology*, 64: 389–416.

**Callaway, B., and P. H. Sant'Anna**

2021 "Difference-in-differences with multiple time periods." *Journal of Econometrics*, 225: 200–230.

**Cappelli, P.**

2019 "Your approach to hiring is all wrong." *Harvard Business Review*, May–June.

**Cardador, M. T.**

2017 "Promoted up but also out? The unintended consequences of increasing women's representation in managerial roles in engineering." *Organization Science*, 28: 597–617.



**Carlin, B., T. Umar, and Y. Hanyi**

2023 "Deputizing financial institutions to fight elder abuse." *Journal of Financial Economics*, 149: 557–577.

**Castilla, E. J.**

2015 "Accounting for the gap: A firm study manipulating organizational accountability and transparency in pay decisions." *Organization Science*, 26: 311–333.

**Castilla, E. J., and S. Benard**

2010 "The paradox of meritocracy in organizations." *Administrative Science Quarterly*, 55: 543–576.

**Cengiz, D., A. Dube, A. Lindner, and B. Zipperer**

2019 "The effect of minimum wages on low-wage jobs." *The Quarterly Journal of Economics*, 134: 1405–1454.

**Christensen, H. B., E. Floyd, L. Y. Liu, and M. Maffett**

2017 "The real effects of mandated information on social responsibility in financial reports: Evidence from mine-safety records." *Journal of Accounting and Economics*, 64: 284–304.

**Christensen, H. B., L. Hail, and C. Leuz**

2016 "Capital-market effects of securities regulation: Prior conditions, implementation, and enforcement." *Review of Financial Studies*, 29: 2885–2924.

**Conley, T., S. Gonçalves, and C. Hansen**

2018 "Inference with dependent data in accounting and finance applications." *Journal of Accounting Research*, 56: 1139–1203.

**Correll, S. J.**

2001 "Gender and the career choice process: The role of biased self-assessments." *American Journal of Sociology*, 106: 1691–1730.

**Correll, S. J.**

2017 "Reducing gender biases in modern workplaces: A small wins approach to organizational change." *Gender & Society*, 31: 725–750.

**Correll, S. J., and C. L. Ridgeway**

2003 "Expectation states theory." In J. Delamater (ed.), *Handbook of Social Psychology*, 29–51. Kluwer Academic/Plenum Publishers.

**Darendeli, A., P. Fiechter, J.-M. Hitz, and N. Lehmann**

2022 "The role of corporate social responsibility (CSR) information in supply-chain contracting: Evidence from the expansion of CSR rating coverage." *Journal of Accounting and Economics*, 74: 1–26.

**Dencker, J. C.**

2008 "Corporate restructuring and sex differences in managerial promotion." *American Sociological Review*, 73: 455–476.

**Deshpande, M., and Y. Li**

2019 "Who is screened out? Application costs and the targeting of disability programs." *American Economic Journal: Economic Policy*, 11: 213–248.

**Dobbin, F.**

2009 *Inventing Equal Opportunity*. Princeton University Press.

**Dobbin, F., D. Schrage, and A. Kalev**

2015 "Rage against the iron cage: The varied effects of bureaucratic personnel reforms on diversity." *American Sociological Review*, 80: 1014–1044.

**Dore, R.**

1983 "Goodwill and the spirit of market capitalism." *British Journal of Sociology*, 34: 459–482.

**Egan, M., G. Matvos, and A. Seru**

2019 "The market for financial adviser misconduct." *Journal of Political Economy*, 127: 233–295.

**Elvira, M. M., and M. E. Graham**

2002 "Not just a formality: Pay system formalization and sex-related earnings effects." *Organization Science*, 13: 601–617.

**Fernandez, R. M., and S. Campero**

2017 "Gender sorting and the glass ceiling in high-tech firms." *ILR Review*, 70(1): 73–104.

**Fernandez, R. M., E. J. Castilla, and P. Moore**

2000 "Social capital at work: Networks and employment at a phone center." *American Journal of Sociology*, 105: 1288–1356.

**Fernandez, R. M., and M. L. Mors**

2008 "Competing for jobs: Labor queues and gender sorting in the hiring process." *Social Science Research*, 37: 1061–1080.

**Fernandez-Mateo, I., and R. M. Fernandez**

2016 "Bending the pipeline? Executive search and gender inequality in hiring for top management jobs." *Management Science*, 62: 3636–3655.

**Fernandez-Mateo, I., and Z. King**

2011 "Anticipatory sorting and gender segregation in temporary employment." *Management Science*, 57: 989–1008.

**Fernandez-Mateo, I., B. Rubineau, and V. Kuppuswamy**

2022 "Reject and resubmit: A formal analysis of gender differences in reapplication and their contribution to women's presence in talent pipelines." *Organization Science*, 34: 1554–1576.

**Fiechter, P., J.-M. Hitz, and N. Lehmann**

2022 "Real effects of a widespread CSR reporting mandate: Evidence from the European Union's CSR directive." *Journal of Accounting Research*, 60: 1499–1549.

**Giroud, X.**

2013 "Proximity and investment: Evidence from plant-level data." *The Quarterly Journal of Economics*, 128: 861–915.

**Gorman, E. H.**

2005 "Gender stereotypes, same-gender preferences, and organizational variation in the hiring of women: Evidence from law firms." *American Sociological Review*, 70: 702–728.

**Gow, A. D., G. Ormazabal, and D. J. Taylor**

2010 "Correcting for cross-sectional and time-series dependence in accounting research." *The Accounting Review*, 85(2): 483–512.

**Granovetter, M.**

1995 *Getting a Job: A Study of Contacts and Careers*. Chicago: University of Chicago Press.

**Hainmueller, J.**

2012 "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies." *Political Analysis*, 20: 25–46.

**Highhouse, S.**

1997 "Understanding and improving job-finalist choice: The relevance of behavioral decision research." *Human Resource Management Review*, 7: 449–470.

**Huber, V. L., M. A. Neale, and G. B. Northcraft**

1987 "Decision bias and personnel selection strategies." *Organizational Behavior and Human Decision Processes*, 40: 136–147.

**Kalev, A., F. Dobbin, and E. Kelly**

2006 "Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies." *American Sociological Review*, 71: 589–617.

**Kanter, R. M.**

1977 "Some effects of proportions on group life: Skewed sex ratios and responses to token women." *American Journal of Sociology*, 82: 965–990.

**Kaufmann, D., A. Kraay, and M. Mastruzzi**

2009 "Governance matters VIII: Aggregate and individual governance indicators, 1996–2008." World Bank Policy Research Paper No. 4978.

**Keller, J. R.**

2018 "Posting and slotting: How hiring processes shape the quality of hire and compensation in internal labor markets." *Administrative Science Quarterly*, 63: 848–878.

**Kowaleski, Z. T., A. G. Sutherland, and F. W. Vetter**

2020 "Can ethics be taught? Evidence from securities exams and investment adviser misconduct." *Journal of Financial Economics*, 138: 159–175.

**Kulik, C. T., and E. Perry**

2008 "When less is more: The effect of devolution on HR's strategic role and construed image." *Human Resource Management*, 47: 541–558.

**Lehmann, N.**

2019 "Do corporate governance analysts matter? Evidence from the expansion of governance analyst coverage." *Journal of Accounting Research*, 57: 721–761.

**Lerner, J. S., and P. E. Tetlock**

1999 "Accounting for the effects of accountability." *Psychological Bulletin*, 125: 255–275.

**Leung, M. D., and S. Koppman**

2018 "Taking a pass: How proportional prejudice and decisions not to hire reproduce gender segregation." *American Journal of Sociology*, 124: 762–813.

**Lowe, J.**

1992 "Locating the line: The front-line supervisor and human resource management." In P. Blyton and P. Turnbull (eds.), *Reassessing Human Resource Management*: 148–168. Sage.

**Maxwell, G. A., and S. Watson**

2006 "Perspectives on line managers in human resource management: Hilton International's UK hotels." *International Journal of Human Resource Management*, 17: 1152–1170.

**McGovern, P., L. Gratton, V. Hope-Hailey, P. Stiles, and C. Truss**

1997 "Human resource management on the line." *Human Resource Management Journal*, 7: 12–29.

**McMullin, J. L., and B. Schonberger**

2020 "Entropy-balanced accruals." *Review of Accounting Studies*, 25: 84–119.

**McPherson, M., L. Smith-Lovin, and J. M. Cook**

2001 "Birds of a feather: Homophily in social networks." *Annual Review of Sociology*, 27: 415–444.

**Mocanu, T.**

2022 "Designing gender equity: Evidence from hiring practices and committees." Working paper, University College London.

**Mukhopadhyay, T., C. Rivera, and H. Tapia**

2019 "Gender inequality and multidimensional social norms." Working paper, United Nations Development Programme. Human Development Report Office.

**Nehles, A. C., M. van Riemsdijk, I. Kok, and J. K. Looise**

2006 "Implementing human resource management successfully: A first-line management challenge." *Management Review*, 17: 256–273.

**O'Reilly, C. A., D. F. Caldwell, and W. P. Barnett**

1989 "Work group demography, social integration, and turnover." *Administrative Science Quarterly*, 34: 21–37.

**Oyer, P., and S. Schaefer**

2011 "Personnel economics: Hiring and incentives." In D. Card and O. Ashenfelter (eds.), *Handbook of Labor Economics*, Vol. 4, Part B: 1769–1823. Elsevier.

**Pager, D., and H. Shepherd**

2008 "The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets." *Annual Review of Sociology*, 34: 181–209.

**Petersen, M. A.**

2009 "Estimating standard errors in finance panel data sets: Comparing approaches." *The Review of Financial Studies*, 22: 435–480.

**Petersen, T., and I. Saporta**

2004 "The opportunity structure of discrimination." *American Journal of Sociology*, 109: 852–901.

**Phillips, D. J.**

2005 "Organizational genealogies and the persistence of gender inequality: The case of Silicon Valley law firms." *Administrative Science Quarterly*, 50: 440–472.

**Podolny, J. M.**

1993 "A status-based model of market competition." *American Journal of Sociology*, 98: 829–872.

**Podolny, J. M.**

1994 "Market uncertainty and the social character of economic exchange." *Administrative Science Quarterly*, 39: 458–483.

**Purcell, J., and S. Hutchinson**

2007 "Front-line managers as agents in the HRM-performance causal chain: Theory, analysis and evidence." *Human Resource Management Journal*, 17: 3–20.

**Rees, A.**

1966 "Information networks in labor markets." *American Economic Review*, 56: 559–566.

**Renwick, D.**

2003 "Line manager involvement in HRM: An inside view." *Employee Relations*, 25: 262–280.

**Reskin, B. F.**

1998 *The Realities of Affirmative Action in Employment*. American Sociological Association.

**Reskin, B. F.**

2000 "The proximate causes of employment discrimination." *Contemporary Sociology*, 29: 319–328.

**Reskin, B. F.**

2003 "Including mechanisms in our models of ascriptive inequality." *American Sociological Review*, 68: 1–21.

**Reskin, B. F., and D. B. McBrier**

2000 "Why not ascription? Organizations' employment of male and female managers." *American Sociological Review*, 65: 210–233.

**Ridgeway, C. L.**

1997 "Interaction and the conservation of gender inequality: Considering employment." *American Sociological Review*, 62: 218–235.

**Rivera, L. A.**

2012 "Hiring as cultural matching: The case of elite professional service firms." *American Sociological Review*, 77: 999–1022.

**Rivera, L. A.**

2015 "Go with your gut: Emotion and evaluation in job interviews." *American Journal of Sociology*, 120: 1339–1389.

**Rivera, L. A.**

2020 "Employer decision making." *Annual Review of Sociology*, 46: 215–232.

**Sant'Anna, P. H. C., and J. Zhao**

2020 "Doubly robust difference-in-differences estimators." *Journal of Econometrics*, 219: 101–122.

**Small, M. L., and D. Pager**

2020 "Sociological perspectives on racial discrimination." *Journal of Economic Perspectives*, 34: 49–67.

**Stainback, K., D. Tomaskovic-Devey, and S. Skaggs**

2010 "Organizational approaches to inequality: Inertia, relative power, and environments." *Annual Review of Sociology*, 36: 225–247.

**Steffensen, D. S., B. P. Ellen, G. Wang, and G. R. Ferris**

2019 "Putting the 'management' back in human resource management: A review and agenda for future research." *Journal of Management*, 45: 2387–2418.

**Tajfel, H., and J. C. Turner**

1979 "An integrative theory of intergroup conflict." In W. G. Austin and S. Worchel (eds.), *The Social Psychology of Intergroup Relations*: 33–37. Brooks/Cole.

**Tolbert, P., and E. J. Castilla**

2017 "Editorial essay: Introduction to a special issue on inequality in the workplace ('What works?')." *ILR Review*, 70: 3–15.

**Tsui, A. S., T. D. Egan, and C. A. O'Reilly**

1992 "Being different: Relational demography and organizational attachment." *Administrative Science Quarterly*, 37: 549–579.

**Uhlmann, E. L., and G. L. Cohen**

2005 "Constructed criteria: Redefining merit to justify discrimination." *Psychological Science*, 16: 474–480.

**UNDP**

2019 *Human Development Report 2019: Statistical Annex*. Palgrave Macmillan.

**Wagner, D. G., and J. Berger**

1997 "Gender and interpersonal task behaviors: Status expectation accounts." *Sociological Perspectives*, 40: 1–32.

**Wing, C., K. Simon, and R. A. Bello-Gomez**

2018 "Designing difference in difference studies: Best practices for public health policy research." *Annual Review of Public Health*, 39: 453–469.