Behaviorally designed training leads to more diverse hiring

By **Cansın Arslan[1], Edward H. Chang[2], Siri Chilazi[3], Iris Bohnet[3,4] and Oliver P. Hauser[1,5]***

Many organizations have shown interest in increasing the diversity of their workforce for various reasons. Collectively, they have spent millions of dollars and countless employee hours on diversity training. Yet there is little empirical evidence that such training increases diversity in organizations (*1*, *2*). Diversity training may be ineffective because it is commonly implemented in ways that are not conducive to changing behavior: It takes place far in advance of consequential decisions and tends to be generic and not specific to the behaviors that organizations may want to affect (e.g., hiring). Results of our large-scale field experiment in a global telecommunications and engineering firm show that behaviorally designed diversity training can influence hiring decisions. This training increased the shortlisting of women and non-national applicants and increased the hiring of non-national applicants relative to business as usual. This proof-of-concept demonstrates that new approaches to diversity training can deliver results.

Although meta-analyses have found that diversity training can help people acquire knowledge about diversity issues (*3*), previous studies of the typical diversity trainings conducted in organizations find largely null effects on behavior (*4, 5*). As such, for organizations interested in increasing diversity, it might be reasonable to conclude that diversity training does not work, so they should instead invest their resources into other sorts of diversity initiatives.

But if training as thus far implemented has been largely ineffective, it is important to ask why. Research on decision-making may offer some clues: Past studies have shown that timeliness is critical for behavioral interventions (*6*), that effects of prejudice-reduction interventions can decay quickly (*7*), that the decision context and appropriate authority figures matter for behavior change (*8*), and that making diversity salient can spur equity-enhancing decisions (*9, 10*). Yet in practice, organizational diversity trainings are typically created without following principles of behavioral design (*11*). As a result, employees often participate in generic diversity training at onboarding or as part of an annual process, without regard to when consequential workplace decisions like hiring and promotions occur, and the content may span myriad topics without focusing on specific decisions.

The company with which we partnered was interested in encouraging managers to hire more people from underrepresented groups. To have the best chance to accomplish this goal, the diversity training drew on principles of behavioral design (*11*). First, the training was timely: hiring managers were asked to watch a 7-minute diversity training video immediately prior to shortlisting candidates (i.e., selecting whom to interview). Second, the content was tailored to hiring decisions: the training told managers to base their assessment on skills and think about how to maximize the collective intelligence of their team. It also asked managers to reflect on what perspectives and characteristics were missing on their team. Third, the training invoked appeals to authority and accountability: the video featured two senior executives of the company discussing these topics. Finally, the training explicitly affirmed that the company valued diversity and inclusion (e.g., by telling managers, "We want to be an organization that reflects the diversity of our society, both today and in the future. Try to bring someone into the team with a background, experience, or perspective that is currently underrepresented").

In addition to the diversity training, following the company's recommendation, the field experiment also included another condition featuring a version of the training that did not explicitly mention diversity but focused

[1] Department of Economics, University of Exeter Business School, Exeter, United Kingdom
[2] Harvard Businesss School, Boston, United States
[3] Harvard Kennedy School, Cambridge, United States
[4] Harvard Radcliffe Institute, Cambridge, United States
[5] Institute for Data Science and Artificial Intelligence, University of Exeter, Exeter, United Kingdom
*Corresponding author. Email: o.hauser@exeter.ac.uk

on team effectiveness instead (hereafter referred to as the "effectiveness training"). This condition allowed us to explore an alternative framing that did not make diversity explicit, potentially assuaging concerns around reactance to explicit diversity trainings (*12*). The content of the effectiveness training was similar to that of the diversity training (e.g., it was of similar length, it included the same senior executives, and the timing of the delivery—i.e., before shortlisting candidates—was the same), but it instead framed hiring people with different perspectives and backgrounds as being beneficial for team effectiveness, as opposed to making any explicit references to diversity or inclusion. For example, the effectiveness training [vs. diversity training] asked managers to "Reflect on what's needed for the role but also what's needed to build a more effective [vs. diverse] team" and invited them to consider "building a high-performing [vs. inclusive] culture," respectively. Comparing these two conditions allows us to isolate the effect of making salient both diversity and the company's commitment to diversity in the training.

To test the effects of these different trainings, all hiring managers at the company (80% men; listing job vacancies across 110 countries) were randomly assigned to one of three conditions prior to their raising the job requisition: the diversity training condition; the effectiveness training condition; or the no-treatment ("business-as-usual") control condition (see Supplementary Figure S1 for a depiction of our study design).

Prior research suggests interventions that are perceived as controlling or constraining can lead to backlash or reactance (*12).* Thus, watching the video was voluntary. In addition, we conducted intention-to-treat analyses (i.e., we analyzed hiring decisions for all 10,433 job requisitions raised by 3,385 managers in the experimental timeframe of July 2022 through December 2022, regardless of whether the hiring manager watched the video), thereby avoiding self-selection issues and increasing the generalizability of our results. For additional information about the experimental conditions, the sample, our analysis strategy, and robustness checks, see the Supplementary Materials (SM). Unless noted otherwise, all analyses were pre-registered (AsPredicted #101182).

We examined the effects of the diversity training on who was shortlisted and who was hired. As illustrated in Figure 1, it was 12% (3.0 percentage points) more likely for a woman to be on the shortlist in the diversity training condition relative to the control condition (p = 0.008, 95% CI: [0.008, 0.053]). However, we did not see significant effects of the diversity training on the likelihood that a woman was hired, relative to the control condition (p = 0.100). The likelihood of shortlisting and hiring female applicants in the effectiveness training condition was not significantly different from the control condition (p = 0.088 and p = 0.238, respectively) or the diversity training (p = 0.394 and p = 0.662, respectively). For a breakdown of results, see Supplementary Table S2.

We pre-registered our intent to examine effects of the treatment on race/ethnicity if the data were available. Collecting such data can be complicated for a multinational company because in some countries it is illegal to collect race/ethnicity data while in others it is uncommon for people to disclose such information. In the end, the engineering firm could only make the nationality of job applicants available to us. A higher representation of non-nationals and women was part of its internal equity, diversity, and inclusion (EDI) goals. Thus, we conducted non-preregistered analyses of whether a "non-national"—that is, a job candidate who is a national of a country that is different from the country of the job opening—was shortlisted or hired (using the same preregistered regression specifications and controls). As illustrated in Figure 1, it was 13% (2.7 percentage points) more likely for a non-national to be on the shortlist in the diversity training condition relative to the control (p = 0.024, 95% CI: [0.004, 0.051]). The diversity training also increased the likelihood that a non-national was hired by 20% (2.1 percentage points) relative to the control (p = 0.025, 95% CI: [0.003, 0.040]). Again, the rates of shortlisting and hiring non-nationals in the effectiveness training condition were not significantly different from the control condition (p = 0.213 and p = 0.362, respectively) or the diversity training condition (p = 0.294 and p = 0.173, respectively).

In further, non-preregistered analyses, we also examined treatment effects on the shortlisting and hiring of

people from multiple underrepresented groups—in our case, female non-nationals. As illustrated in Figure 1, it was 28% (2.1 percentage points) more likely for a female non-national to be on the shortlist in the diversity training condition relative to the control (p = 0.007, 95% CI: [0.006, 0.037]). The diversity training also increased the likelihood that a female non-national applicant was hired by 41% (1.2 percentage points) relative to the control (p = 0.018, 95% CI: [0.002, 0.023]). The increase in shortlisting in the diversity training condition was also significantly higher relative to the effectiveness training condition (p = 0.010, 95% CI: [0.005, 0.037]) but it was not significant for hiring (p = 0.055). The effectiveness training condition was not significantly different with respect to shortlisting and hiring relative to the control (p = 0.973 and p = 0.652, respectively).

The findings for non-nationals were not just driven by Europeans moving around Europe or Canadians being hired in the US. In further, non-preregistered analyses, we found similar results when examining whether applicants from the Global South were shortlisted or hired (Table S3).

We also conducted various heterogeneity analyses to explore potential moderators. We found that the effects of diversity training were generally larger in higher income countries (Table S4), in countries that have higher levels of gender equality (Table S5), and in the wealthy, relatively gender-equal headquarters country of the company (Tables S6-S8). We did not find significant moderation by individual characteristics of the manager such as manager gender or seniority (Table S9 and S10). These results suggest that there is considerable variation in response to diversity training by country and cultural context but not by individual factors.

The results of this field experiment show that a behaviorally designed diversity training can lead to consequential behavior change. While the effect sizes on hiring were at times small in absolute terms, they were often large in relative terms (e.g., it was 114% more likely for a female non-national candidate to be hired in the headquarters country in the diversity training vs. control condition), which reflects low baseline rates of hiring members of historically underrepresented groups. In addition, our results were stronger in the headquarters country where societal gender equality is high. It is plausible that diversity interventions work better when aligned with cultural norms (*3*) or that our intervention induced higher levels of conformity among managers closer to the center of the organization. In contrast, the effectiveness training did not significantly affect shortlisting or hiring compared to the control and also did significantly worse for shortlisting outcomes than diversity training when looking at female non-nationals, consistent with past research showing that making diversity salient can spur equity-enhancing decisions (*9, 10*).

As our goal was to test a behaviorally designed diversity training that consisted of many differences to past trainings—including diversity salience, timeliness, task specificity, and the delivery by an organizational authority figure—we are unable to single out the specific active ingredient that led to the observed behavior change. It is possible that any, all, or some combination of these factors were necessary to see our effects, and we encourage future research to experimentally isolate each of these factors. Our approach also made salient that managers were expected to contribute to an "organization that reflects the diversity of our society" and, as such, managers may have reweighted the decision criteria ahead of their shortlisting and hiring decisions or simply complied with a request from a senior leader (*13*).

We view the current research as a proof-of-concept that diversity training can "work" to change the diversity of hires but that its success relies on carefully considered design choices and the decision context. While the outcome of interest in this study was demographic representation, we cannot exclude the possibility that the trainings also affected other outcomes such as team performance, job satisfaction, or turnover.

Going forward, we hope that researchers, policymakers, and organizational leaders invest in rigorous randomized controlled trials (*14*) to study when and how diversity training can be implemented in ways that lead to actual behavior change across a wide range of relevant organizational outcomes. Such evidence might be particularly important at a time when the merits and impacts of diversity initiatives are debated. Many organizations

have committed to broadening their representation and want to access the full range of the talent pool. Our research offers a practical tool for leaders interested in changing behavior to meet this commitment.

REFERENCES AND NOTES

1. A. Kalev, F. Dobbin, E. Kelly, *American Sociological Review* **71**, 589 (2006).
2. E. L. Paluck, R. Porat, C. S. Clark, D. P. Green, *Annual Review of Psychology* **72**, 533 (2021).
3. K. Bezrukova, C. S. Spell, J. L. Perry, K. A. Jehn, *Psychological Bulletin* **142**, 1227 (2016).
4. E. H. Chang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7778 (2019).
5. C. K. Lai, J. A. Lisnek, *Psychological Science* **34**, 424 (2023).
6. J. Austin, S. O. Sigurdsson, Y. S. Rubin, *Environment and Behavior* **38**, 140 (2006).
7. C. K. Lai *et al.*, *Journal of Experimental Psychology: General* **145**, 1001 (2016).
8. R. B. Cialdini, N. J. Goldstein, *Annual Review of Psychology* **55**, 591 (2004).
9. E. H. Chang, E. L. Kirgios, A. Rai, K. L. Milkman, *Management Science* **66**, 2752 (2020).
10. E. L. Kirgios, A. Rai, E. H. Chang, K. L. Milkman, *Nature Human Behavior* **6**, 383 (2022).
11. I. Bohnet, S. Chilazi, *Make Work Fair*, Harper Collins (forthcoming: 2025).
12. F. Dobbin, D. Schrage, A. Kalev, *American Sociological Review* **80**, 1014 (2015).
13. J. Axt, J. To. *Personality and Social Psychology Review* (2024).
14. O. P. Hauser, E. Linos, T. Rogers, *Research in Organizational Behavior* **37**, 185 (2017).
15. C. Arslan, E. H. Chang, S. Chilazi, I. Bohnet, O. P. Hauser. Code and data access instructions for "Behaviorally designed training leads to more diverse hiring", https://doi.org/10.5061/dryad.9cnp5hqvt

Acknowledgments: All code for replication and information about how to request access to the proprietary dataset can be found in the Supplementary Materials as well as at (*15*). We are grateful to the global telecommunications and engineering company that allowed us to study these research questions and to MoreThanNow, a behavioral science consultancy, that assisted the company with the implementation of the randomized controlled trial. We wish to thank, in particular, J. Elfer, K. Wright, and G. Lindemann for the trial implementation, as well as E. Kirgios, K. Milkman, and A. Duckworth for feedback on the manuscript. Neither the field partner nor MoreThanNow had any role in the data analysis, the publishing decision, or the manuscript preparation. O.P.H. gratefully acknowledges funding from the UKRI Future Leaders Fellowship (MR/T020253/1). For the purpose of open access, the authors have applied a CC-BY license to any Author Accepted Manuscript version arising.

10.1126/science.ads5258

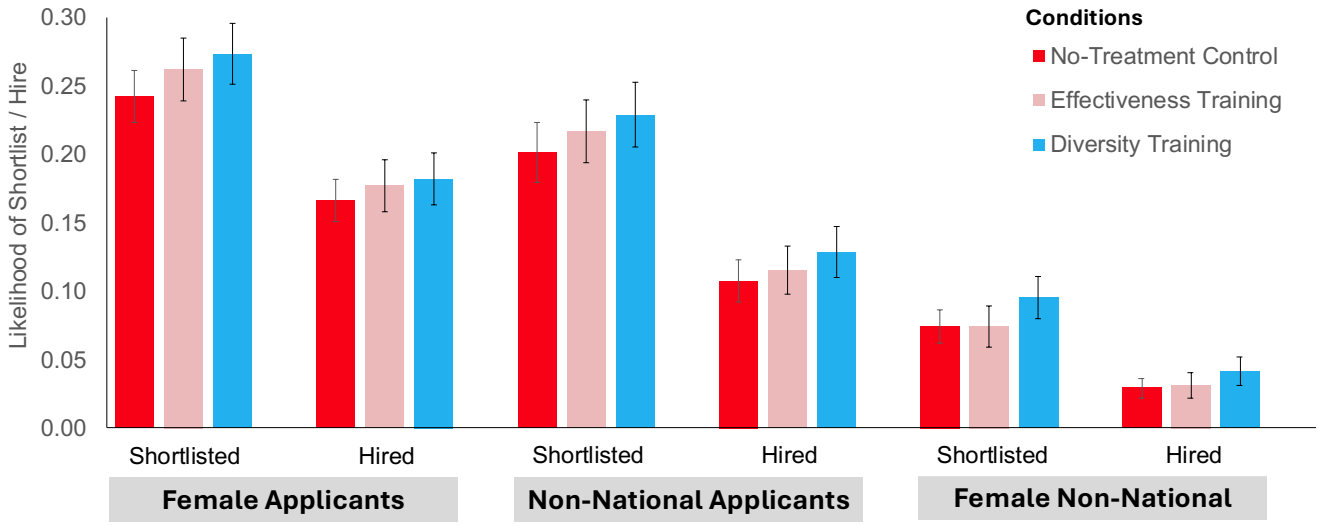

Fig. 1. Analyses reveal that the percentages of female, non-national, and female non-national applicants being shortlisted and hired are highest in the diversity training condition. Likelihood estimates of conditions include pre-registered control variables and fixed effects, see Supplementary Materials. Error bars represent 95% confidence intervals.

Supplementary Materials for

Behaviorally Designed Training Leads to More Diverse Hiring

Cansın Arslan, Edward H. Chang, Siri Chilazi, Iris Bohnet, and Oliver P. Hauser

Corresponding author: o.hauser@exeter.ac.uk

The PDF file includes:

Materials and Methods

Context

We analyzed data from a field experiment conducted at a large (roughly 100,000 employees), global telecommunications and engineering company operating in over 100 countries. Their Equity, Diversity, and Inclusion (EDI) strategy highlights the importance of being evidence-based and incorporating insights from behavioral science into their decision-making processes. One of the firm's EDI priorities has been to increase the representation of people from traditionally underrepresented backgrounds, which includes women and non-nationals at this company.

Procedures

At this company, the recruitment process typically begins with a hiring manager raising a job requisition. An internal recruiter is then assigned to the requisition to work with the hiring manager to post a job advertisement and monitor job applications. Once the deadline has passed or the recruiter determines that candidates are ready to be reviewed, the recruiter changes the status of the requisition to allow the hiring manager to review the applications. This is the point at which the intervention (treatments) took place.

In the diversity training and effectiveness training conditions, hiring managers received an email that stated:

> "Dear [name of hiring manager],
>
> You're hiring for a new role on your team. It's an exciting time and one of the most important decisions a manager can make – not just for you and the people around you, but for [company name] as a whole. As you consider who to interview, please watch this short video featuring our thoughts on recruitment at [company name].
>
> Here's the link to watch. We hope you find it useful.
>
> Best regards,
> [Executives at the company]"

The link in the email directed to a video that discussed the importance of hiring for the company. In the diversity training condition, the video explicitly mentioned diversity and inclusion and explicitly asked the hiring manager to consider diversity in their hiring decision. In the effectiveness training video, the video contained much of the same content, but did not explicitly mention diversity or inclusion.

The different conditions emerged in part due to conversations with our field partner. Broadly, they were interested in trying to encourage their managers to hire more people from underrepresented groups to help diversify the organization. Thus, we designed a diversity training that we thought had the best chance of accomplishing this goal. We incorporated principles of behavioral design in our diversity training: we made the diversity training timely, tailored the content to be behaviorally specific and relevant to hiring, and invoked appeals to authority and accountability by including senior members of the organization in the video. In addition, based on the reasoning and theorizing provided in the paper, we thought that explicitly mentioning diversity would lead to a more impactful intervention (e.g., following work on diversity salience; Chang et

al., 2020; Kirgios et al., 2022). For example, the diversity training explicitly said that "Diversity of thought is a strength" and includes explicit statements about how the company values diversity (e.g., "At [company name], we want to be an organization that reflects the diversity of our society, both today and in the future.").

However, in discussions with our field partner, an alternative proposal also emerged, which was to not talk about diversity explicitly and instead talk about the benefits of hiring people with different backgrounds and perspectives to make teams more effective, related to scholarly work on how different perspectives can lead to enhanced team functioning (e.g., Hong and Page, 2004; Woolley et al., 2010). As such, the effectiveness training condition did not explicitly mention diversity or inclusion and instead trained people on hiring to create effective teams (e.g., by including statements like, "If you hire similar people, your group thinking and decision-making will be weaker as you will be missing out on some important perspectives"). The effectiveness training [vs. the diversity training] also included statements like, "Reflect on what's needed for the role but also what's needed to build a more effective [vs. diverse] team" and "building a high-performing [vs. inclusive] culture."

Because it was plausible that either training might be more impactful (e.g., some research suggests that talking explicitly about diversity might lead to backlash or reactance; Legault, Gutsell, and Inzlicht, 2011; Kidder et al., 2004), we implemented a three-condition study, where we could compare each of the two conditions to the no-treatment "business-as-usual" control condition and to each other. While our primary focus is on the diversity training condition, we thought that the effectiveness training could have its own effects, which is why our pre-registered analysis strategy also explicitly tests for its effect relative to the business-as-usual control.

In Table S1, we show the redacted scripts of the videos for both conditions. Watching the video was not mandatory. However, to avoid issues of differential selection, we include all managers who raised a requisition during the experimental time frame in our analyses, regardless of whether they watched the video or not.

A reminder email was sent to all managers assigned to the diversity training or effectiveness training conditions if they did not click on the video link within the first two weeks of the original email invite. No additional treatment emails were sent to managers who hired for multiple roles in the study period.

In the control condition, hiring managers did not receive any email.

Sample

Because the company did not know ex ante which managers at the company would hire in the coming months, all managers at the company (~7,500 managers) were randomly assigned to one of the three conditions, with randomization stratified by manager gender, tenure, job level, and unit. While our pre-registration stated that the company expected to run this field experiment for three months, they ended up running it for just over six months. The final sample of managers includes only those who ended up raising at least one requisition during the experimental period (June 2022 to December 2022).

In total, 3,644 managers raised 13,691 hiring requisitions (i.e., job openings). Of these, 3,063 managers ended up successfully hiring for 9,787 roles. Managers' likelihood of successfully hiring for the opening was not statistically different across conditions ($p = 0.975$ diversity training relative to control; $p = 0.934$ diversity training relative to effectiveness training; $p = 0.910$ effectiveness training relative to control).

There were 367,805 applications made to the available job openings during the study period. 22% of applicants were women, 35% of applicants were non-nationals, and 8% of applicants were female non-nationals. On average, each requisition received 27 applications. We see no statistically significant differences in the number of applications (p = 0.166 diversity training relative to control; p = 0.410 diversity training relative to effectiveness training; p = 0.616 effectiveness training relative to control) or in the demographic (gender and nationality) composition of applicants across conditions (all p-values are larger than 0.5).

We preregistered not analyzing data that had missing values in key variables. As such, for analyses regarding gender, we do not consider candidates with missing gender information (7% of candidates) or requisitions where the person hired is of unknown gender (22% of requisitions). As a result, we are left with 10,433 requisitions when analyzing how our treatments influenced the hiring of women. We acknowledge that not analyzing applicants with missing gender implies that our study may not be able to speak to applicants who do not conform to binary gender identities or prefer not to disclose their gender for other reasons. We encourage more research in the future that studies the experiences of non-binary people. For analyses regarding nationality, we do not consider candidates with missing nationality information (10% of candidates) or requisitions where the person hired is of unknown nationality (39% of requisitions). As a result, we are left with 7,759 requisitions when analyzing how our treatments influenced the hiring of non-nationals. Similarly, for analyses that focus on female non-national applicants, the above data exclusions on both gender and nationality simultaneously leave us with 7,549 applicants.

For our main analyses, we included all requisitions regardless of whether an applicant was ultimately hired. Thus, our analytical strategy is intention-to-treat (ITT), which avoids potential self-selection issues if managers are affected by the treatment and decide not to hire anyone (note that we do not find evidence of differential manager behavior across conditions in likelihood of hiring; ps > 0.585). However, a portion of requisitions in our dataset (29%) ended up unfilled, either because no one applied or no one was hired. According to our field partner, unfilled requisitions are not unusual and can simply be due to internal changes or requirements of the vacancy. In an intention-to-treat analysis, by definition, unfilled requisitions are coded as 0; only requisitions that are filled and that hire a female applicant (or, depending on the outcome variables of interest, a non-national or female non-national applicant) are coded as 1. Thus, for further robustness checks, we also analyzed the data excluding requisitions where no one was hired.

Data

The data comprises all requisitions raised and all candidates who applied for at least one of the requisitions raised during the study period where we did not have missing data on key variables, following our preregistration. Available manager characteristics include gender (binary), age (continuous in years), tenure at the company (continuous in years), job level (categorical), job function (categorical), job unit (categorical), and country (categorical). Job requisition characteristics include level (categorical), function (categorical), unit (categorical), and country (categorical). Among the job requisitions, there were five job levels (ranging in seniority), 35 functions (e.g., finance, marketing, sales), 18 units, and 110 countries represented. Functions, units, and countries with very few observations (i.e., less than 15 observations) were grouped under an "other" category so as not to drop them with the inclusion of preregistered fixed effects.[1]

[1] While the grouping into the "other" category for functions, units and countries with very few observations was not pre-registered, we realized only after receiving the data that this was necessary to avoid losing too many

Candidate-level variables include whether the candidate was shortlisted (binary), whether the candidate was hired (binary), gender (binary), age (continuous in years), and country of nationality (categorical). For analyses regarding non-nationals, we created a binary variable for each candidate comparing the nationality of the candidate to the country of the job requisition the candidate had applied for.

For our preregistered heterogeneity analyses, we relied on two external data sources. To test whether treatment effects vary with respect to gender norms of the country of requisition, we used the global gender gap index (2023) of the World Economic Forum.[2] We created a binary variable that took the value of one if a country has above-median gender equality and zero otherwise. Secondly, we used World Bank country-level income classifications to analyze whether treatment effects differed with respect to the income level of the country of the job requisition. We created a binary variable that took the value of one if a country is a high-income country and zero otherwise. Analyses using continuous variables instead of binary variables yield similar results, so we present the results of the analyses using binary variables for ease of interpretation.

Data Access

The data from this experiment is of a proprietary nature from a third-party global engineering and telecommunications company and, due to our data sharing agreement with this company, we are not allowed to make this data publicly available. We have deposited our code and analysis files at Dryad (https://doi.org/10.5061/dryad.9cnp5hqvt) to demonstrate transparency and clarity for reproducibility, and we have also agreed with the intermediary company (MoreThanNow) to set up a specifically monitored email address[3], which interested readers can use to confidentially discuss the process of obtaining data access. Researchers interested in accessing the data used in our study are referred to the last page of the Supplementary Materials, which contains instructions for how to submit a data access request form to MoreThanNow who handle the data access requests on behalf of the field partner.

Outcome Variables

- *Proportion of Shortlisted Applicants Who Are Female.* For each requisition, we calculated the proportion of the shortlisted applicants who were female (0 to 1, continuous).
- *Shortlisting of Female Applicants.* For each requisition, we coded whether the manager shortlisted at least one female applicant (1 if so; 0 otherwise).
- *Hiring of Female Applicants.* For each requisition, we coded whether the manager hired a female applicant (1 if so; 0 otherwise).

We preregistered looking at the same outcome variables as above for members of historically underrepresented ethnic/racial groups if the data were available. Collecting such data can be complicated for a multinational company because in some countries, it is illegal to collect race/ethnicity data, whereas in others, it is uncommon for people to disclose such information. In the end, the engineering firm could only make the nationality of job applicants available to us. Since we only had data on job candidate nationality, we also considered the following outcome variables in non-preregistered analyses:

observations when using fixed effects in our regression framework. However, all results are robust to not using any groupings, and to changing how the groupings are constructed.

[2] Page 11: https://www3.weforum.org/docs/WEF_GGGR_2023.pdf

[3] Email: experiment@morethannow.co.uk

- *Proportion of Shortlisted Applicants Who Are Non-Nationals.* For each requisition, we calculated the proportion of the shortlisted applicants who were non-nationals (0 to 1, continuous).
- *Shortlisting of Non-National Applicants.* For each requisition, we coded whether the manager shortlisted at least one non-national (1 if so; 0 otherwise).
- *Hiring of Non-National Applicants.* For each requisition, we coded whether the manager hired a non-national applicant (1 if so; 0 otherwise).

In additional non-preregistered analyses, we also considered the intersection of gender and nationality for these three outcome variables.

- *Proportion of Shortlisted Applicants Who Are Female Non-Nationals.* For each requisition, we calculated the proportion of the shortlisted applicants who were female non-nationals (0 to 1, continuous).
- *Shortlisting of Female Non-National Applicants.* For each requisition, we coded whether the manager shortlisted at least one female non-national applicant (1 if so; 0 otherwise).
- *Hiring of Female Non-National Applicants.* For each requisition, we coded whether the manager hired a female non-national applicant (1 if so; 0 otherwise).

Finally, in non-preregistered analyses, we considered the above non-national outcomes restricting to applicants from the Global South.[4]

- *Proportion of Shortlisted Applicants Who Are Non-Nationals from the Global South.* For each requisition, we calculated the proportion of the shortlisted applicants who were non-nationals from the Global South (0 to 1, continuous).
- *Shortlisting of Non-National Applicants from the Global South.* For each requisition, we coded whether the manager shortlisted at least one non-national from the Global South (1 if so; 0 otherwise).
- *Hiring of Non-National Applicants from the Global South.* For each requisition, we coded whether the manager hired a non-national applicant from the Global South (1 if so; 0 otherwise).
- *Proportion of Shortlisted Applicants Who Are Female Non-Nationals from the Global South.* For each requisition, we calculated the proportion of the shortlisted applicants who were female non-nationals from the Global South (0 to 1, continuous).
- *Shortlisting of Female Non-National Applicants from the Global South.* For each requisition, we coded whether the manager shortlisted at least one female non-national applicant from the Global South (1 if so; 0 otherwise).
- *Hiring of Female Non-National Applicants from the Global South.* For each requisition, we coded whether the manager hired a female non-national applicant from the Global South (1 if so; 0 otherwise).

Analysis Strategy

Following our preregistered analysis strategy, we ran ordinary least squares (OLS) regressions predicting the effect of condition assignment on each of the outcome variables

[4] We designate countries as being in the Global South based on the Organization for Women in Science for the Developing World's list: https://owsd.net/sites/default/files/OWSD%20138%20Countries%20-%20Global%20South.pdf

specified above. Analyses were conducted at the job requisition level, with robust standard errors clustered at the manager level (the unit of random assignment). We use ordinary least squares (OLS) to estimate the following equation:

$$Y_{im} = \alpha + \beta_1 D_m + \beta_2 E_m + X'_m \delta + \mu_i + \Omega_i + \eta_i + \theta_i + u_{im}$$

where Y_{im} is the outcome of job opening (hiring requisition) i raised by manager m, where the outcome refers to one of the variables described above (e.g. proportion of shortlisted applicants who are female non-nationals). D is an indicator variable that takes the value 1 if the manager m is assigned to the diversity training condition and E is an indicator variable that takes the value 1 if the manager m is assigned to the effectiveness training condition, and 0 otherwise. u_{im} is the error term. Under the identifying assumption that treatment assignments are orthogonal to the error term (holds by randomization design), the coefficients β_1 and β_2 are estimates of the causal effect of the respective treatment assignments.

X'_m is a vector of preregistered controls including manager gender (binary), manager age (continuous), and tenure within the organization (continuous in years), tenure squared, and job-level (categorical). In addition, our specification incorporates various requisition-level fixed effects (FEs), including μ_i for job level, Ω_i for job function, η_i for job unit, and θ_i for job country. OLS results for baseline specification with no controls and pre-registered full specification for all outcome variables are presented in Table S2, and we follow the same pattern when presenting the results of our exploratory analyses described below.

Supplementary Text

Preregistered Heterogeneity Analyses

As pre-registered, we conducted various heterogeneity analyses to explore the potential moderators.

Different countries: We found heterogeneous treatment effects with respect to the country of requisition. Specifically, we found that the effects were larger in higher income countries (Table S4), in countries that have higher levels of gender equality (Table S5), and in the headquarters country which is both a high-income country and a country with high levels of gender equality (Table S6).

Headquarters country: We further investigated whether the treatment effects could be particularly driven by the headquarters country, which is one of the most significant units of economic activity for the company (15% of the open jobs during the study period were posted there) and—by virtue of its location—is closest to the central HR team's initiatives. As shown in Table S6, the effects of the diversity training were significantly larger in the headquarters country than in non-headquarters countries (p-values of interaction terms: 0.055 for female, 0.006 for non-national, and 0.002 for female non-national).

Consistently, treatment effects are larger and more significant if we restrict the analysis to headquarters country alone (Table S7). In an additional analysis focusing only on requisitions at the headquarters, we found that it was 32% more likely for a woman to be hired in the diversity training condition relative to effectiveness training (p = 0.024) and the control conditions (p = 0.027). Similarly, it was 51% more likely for a non-national applicant to be hired (p = 0.004 relative to effectiveness training; p = 0.004 relative to control), and it was 115% more likely for a

female non-national applicant to be hired ($p < 0.000$ relative to effectiveness training; $p = 0.001$ relative to control). Additionally, we found similar significant effects for hiring non-nationals and female non-nationals from the Global South for roles in the headquarters country (Table S8).

Manager gender: We examined how manager gender interacts with our treatments. While we found that female managers were significantly more likely to hire female applicants in the control condition (Table S9), we did not find that female managers were significantly more likely to hire female applicants in the diversity training or effectiveness training conditions. There were also no significant interaction effects of manager gender and treatment on the hiring of non-national or female non-national applicants (Table S9).

Seniority: As presented in Table S10, we did not find any significant heterogeneous treatment effects with respect to the seniority of the hiring manager (defined as having an above-median job-level). We did find, however, that senior managers are less likely to hire people from a traditionally underrepresented background.

Team diversity: We did not find a significant interaction effect between the hiring manager's existing team (gender) diversity and the diversity training condition (Table S11). We also no longer see a significant main effect of the diversity training condition in this specification, suggesting that the diversity training did not have a significant effect on female hiring when the manager's team had below median gender diversity to start. However, because there was no significant interaction, we cannot make the claim that the diversity training was significantly less effective for less gender diverse teams as compared to more gender diverse teams. We did find a significant interaction between the existing team (gender) diversity and the effectiveness training condition, but because there are no additional indications that the effectiveness training condition increased diversity in our main analyses, we do not explore (or interpret the meaning of) this single significant interaction further. Future research could aim to unpack how effectiveness trainings interact with existing team diversity. Note that we do not have data on team diversity in terms of nationality (we only have data on the country of the job requisition and the nationalities of job applicants, not existing employees) and thus analyze team diversity only in terms of gender.

Applicant pool diversity: As Table S12 shows, we found significant interactions between the diversity of the applicant pool and the diversity training condition when considering non-nationals and female non-nationals, suggesting that when more non-nationals were in the applicant pool, the diversity training was more effective. This may, however, be a mechanical result, as managers cannot shortlist or hire non-nationals when they are not in the applicant pool, and the baseline proportions of non-nationals in the applicant pools were low.

Multiple hires: Some managers in our sample raised multiple hiring requisitions in the experimental time frame, so we tested whether managers behaved differently for the first job requisition (when they received the treatment) versus later requisitions. We did not find significant heterogeneous treatment effects (Table S13). While we anticipated that the effect of our behavioral intervention might decay over time, it is possible that our intuition that the timeliness of the intervention matters may not be as critical as we had assumed. One potential explanation is that managers who were spurred to interview more diverse candidates and (in some cases) hire more diverse candidates as a result of the training now have been "exposed" to diversity in a way that could change their beliefs about the quality of candidates from underrepresented groups (e.g., via intergroup contact) and led them to change their hiring practices. Another potential explanation is that managers exposed to the diversity training have been made aware of the company's increased emphasis of hiring diversely and the possible accountability to senior management that comes with

it. Both of these explanations have in common that some form of learning has taken place, albeit through different mechanisms and for different reasons.

Robustness Checks

As pre-registered, we also estimated local average treatment effects (LATE) using an instrumental variable strategy where we used the random assignment of condition as an instrument for those who affirmatively watched the video.

Although our field partner provided us with information about employees who may have clicked on the video link, we were unable to test and verify that the tracking works as intended. We therefore think that the play rate provided to us may be a lower bound on the number of managers who watched the video (e.g. managers who have had ad or content blocking software may not have been tracked correctly). We can confidently say that some managers affirmatively did watch the video, but we do not think we can confidently say that other managers did not watch the video. Therefore, we caution against overinterpreting the LATE analyses because of the potential challenges associated with our ability to be sure of who may have watched the video but was not tracked accordingly.

Additionally, given that we have only one instrument (i.e., treatment assignment), we can only estimate LATE for one treatment condition at a time. Hence, we estimated the effect of one condition (the diversity training condition or the effectiveness training condition) against the control by dropping the other condition observations from the sample so as to be able to estimate each treatment effect separately. Results are presented in Table S14. As expected, the coefficients of diversity training are consistent with that of the ITT estimations in terms of direction and significance, and larger because we focus on the treated managers (i.e., managers who affirmatively watched the video).

For robustness, we also repeated all analyses that involve a binary outcome variable with logit models instead of the pre-registered OLS models. We find that all results are qualitatively similar and robust when using this alternative estimation method (Table S15).

For further robustness checks, we also analyzed the data excluding requisitions where no one was hired. We find that our results are robust to this alternative sample (Table S16).

Supplementary Figures and Tables

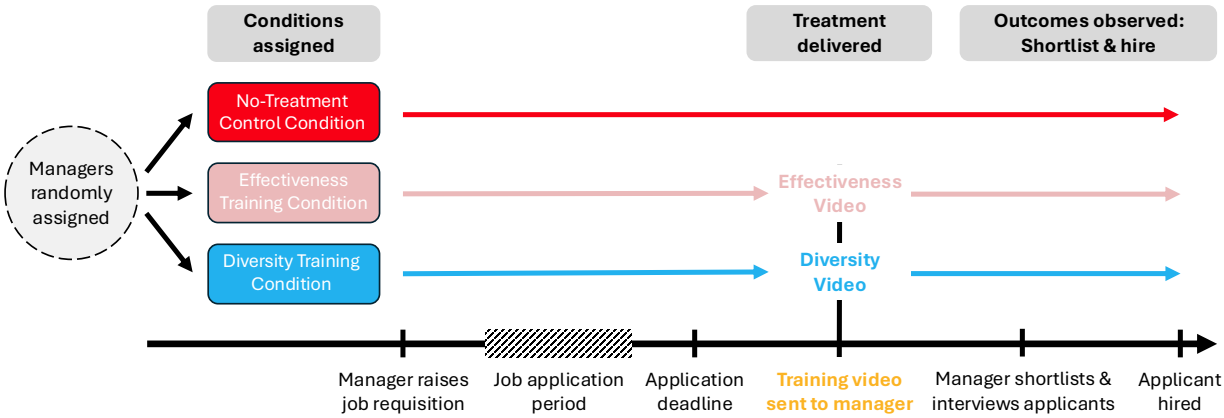

Figure S1. The study design and timeline start with all hiring managers being randomly assigned to one of the three conditions. After managers raise a job requisition and the job advertisement period closes, managers in the diversity training and effectiveness training conditions are sent an email with a link to the respective videos. The pre-registered outcomes include shortlisting and hiring decisions. We analyze both shortlisting and hiring decisions for all managers who raised a requisition regardless of whether they watched the video to avoid self-selection and increase generalizability (i.e., intention-to-treat analysis).

Table S1. Scripts for treatment videos. Differences between scripts are in blue text.

SCRIPT FOR DIVERSITY TRAINING CONDITION	SCRIPT FOR EFFECTIVENESS TRAINING CONDITION
Leader script	**Leader script**
Welcome, I am [executive at company]. I heard you are about to recruit a new team member, and I'm delighted to be talking to you about such an important decision: For you, for your team, and for our organization. I know this can feel like a big decision, but no worries, we are in this together. An important thing I have learned in hiring is to challenge myself to ensure I drive more effective and inclusive hiring: • What do I mean by effective? Well, we need to ensure we hire for the future. How? When you're interviewing and screening CVs, I want you to consider a broad range of skills. Technical skills always have and always will be important at [company name], but non-technical skills are important too, for example, the ability to collaborate with others and communicate clearly. Don't forget to prioritize those. And to look not only at the skills someone demonstrates now, but evidence of their potential to contribute further in the future. • And what about inclusive? Well, I want you to think about what your team needs to be stronger. If you hire everyone with similar characteristics and experiences, your group thinking	Welcome, I am [executive at company]. I heard you are about to recruit a new team member, and I'm delighted to be talking to you about such an important decision: For you, for your team, and for our organization. I know this can feel like a big decision, but no worries, we are in this together. An important thing I have learned in hiring is to challenge myself to ensure I drive more effective and team-based hiring: • What do I mean by effective? Well, we need to ensure we hire for the future. How? When you're interviewing and screening CVs, I want you to consider a broad range of skills. Technical skills always have and always will be important at [company name], but non-technical skills are important too, for example, the ability to collaborate with others, and communicate clearly. Don't forget to prioritize those. And to look not only at the skills someone demonstrates now, but evidence of their potential to contribute further in the future. • And what about team-based? Well, I want you to think about what your team needs to be stronger. If you hire similar people, your group thinking and decision-making will be weaker as you will

and decision-making will be weaker as you will be missing out on some important perspectives. Diversity of thought is a strength. Hire someone who will make a unique contribution to your team and our company! And remember, every hire matters because it contributes to the talent and the inclusive culture of our organization. Your decision will make a difference!	be missing out on some important perspectives. Hire someone who will make a unique contribution to your team and our company! And remember, every hire matters, because it contributes to the talent of our organization. Your decision will make a difference!
Expert Script	**Expert Script**
That's right, [executive name]. We consider hiring one of the critical parts of our people's experience as it shapes our business and our culture. It is up to you now to make a hiring decision that makes your team stronger and helps us continue building a more inclusive [company name]. As [executive name] mentioned, making the right hiring decision is never easy. We're all busy, and that means being tempted to choose a safe option. Usually, someone who reminds us of ourselves, shares our characteristics and experiences. But what else can you do to ensure a diverse talent pool? How can we put into practice [executive name]'s advice? Let me share two of the best practices I have learned throughout my experience as a Talent Acquisition expert. These are practical and something you can use again in the future. The first starts with a quick reflection exercise. Are you ready? Okay…. **ON SCREEN: ONE - REFLECTION EXERCISE: WHAT'S MISSING?**	That's right, [executive name]. We consider hiring one of the critical parts of our people's experience as it shapes our business and our culture. It is up to you now to make a hiring decision that makes your team stronger. As [executive name] mentioned, making the right hiring decision is never easy. We're all busy, and that means being tempted to choose a safe option. But what else can you do to hire well? How can we put into practice [executive name]'s advice? Let me share two of the best practices I have learned throughout my experience as a Talent Acquisition expert. These are practical and something you can use again in the future. The first starts with a quick reflection exercise. Are you ready? Okay… **ON SCREEN: ONE - REFLECTION EXERCISE: WHAT'S MISSING?**

"Picture your team and think of yourselves as a collective unit. Now think about: • Where are you strong together? • Where do you need to develop? • What unique characteristics, experiences, and perspectives are you missing that might make you a more diverse and higher-performing team? Pause this video, write those things down and think about how to integrate them into what you're looking for in your next hire.	"Picture your team and think of yourselves as a collective unit. Now think about: • Where are you strong together? • Where do you need to develop? • What is missing that might make you a higher-performing team? Pause this video, write those things down and think about how to integrate them into what you're looking for in your next hire.
ON SCREEN: TWO - HIRING TIPS: HIRE TO CONTRIBUTE TO YOUR TEAM	**ON SCREEN: TWO - HIRING TIPS: HIRE TO CONTRIBUTE TO YOUR TEAM**
Welcome back, now let me ask you – how many times have you heard someone say they're looking for a 'mini-me' when really, they should be thinking about the skills, experiences, and characteristics they don't have.	Welcome back, now let me ask you – how many times have you heard someone say they're looking for a 'mini-me' when really, they should be thinking about the skills they don't have.
Go back to the list you wrote down a moment ago and consider these questions when screening your CVs: • Does this person have valuable strengths, skills, and experiences that are lacking on my team? • Have I considered what we value in terms of diversity, technical skills, and talent in our new hires for each candidate? • Does this person contribute to the diversity of the team?	Go back to the list you wrote down a moment ago and consider these questions when screening your CVs: • Does this person have valuable strengths, skills, and experiences? • Have I considered what we value in terms of technical skills and talent in our new hires for each candidate?
At [company name], we want to be an organization that reflects the diversity of our society, both today and in the future. Try to bring someone into the team with a	

background, experience, or perspective that is currently underrepresented. **Recruiter Script** Thanks for watching this video, and remember: • Reflect on what's needed for the role but also what's needed to build a more diverse team. • Be open-minded about who you choose to interview. Seek out the difference. • Assess your candidates across a broad range of skills, perspectives, and experiences. As both [previous names] highlighted, we all tend to gravitate to people similar to us, and for that reason, building an inclusive culture with more effective teams needs this proactive approach. This is right for you, right for your team, and right for the future of [company name]. Thanks so much for listening and playing your part as a hiring manager. The talent acquisition team is here to help when you need us. Good luck with the next steps.	**Recruiter Script** "Thanks for watching this video, and remember: • Reflect on what's needed for the role but also what's needed to build a more effective team. • Be open-minded about who you choose to interview. • Assess your candidates across a broad range of skills. As both [previous names] highlighted, we all tend to gravitate to people similar to us, and for that reason, building a high-performing culture with more effective teams needs this proactive approach. This is right for you, right for your team, and right for the future of [company name]. Thanks so much for listening and playing your part as a hiring manager. The talent acquisition team is here to help when you need us. Good luck with the next steps.

Table S2. OLS Regressions Predicting Shortlisting and Hiring. Full Sample.

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.013	0.026**	0.016	0.030**	0.005	0.016
	(0.012)	(0.010)	(0.014)	(0.011)	(0.011)	(0.010)
Effectiveness Training	0.012	0.017	0.016	0.020	0.009	0.011
	(0.012)	(0.010)	(0.014)	(0.012)	(0.012)	(0.010)
Control Mean	0.196		0.242		0.166	
Full Controls Included		Y		Y		Y
Observations	10,433	10,433	10,433	10,433	10,433	10,433
R-squared	0.000	0.102	0.000	0.103	0.000	0.090

	% Non-Nationals Shortlisted		Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.026	0.025*	0.030	0.027*	0.023	0.021*
	(0.015)	(0.011)	(0.017)	(0.012)	(0.013)	(0.010)
Effectiveness Training	0.018	0.014	0.019	0.015	0.013	0.008
	(0.015)	(0.010)	(0.017)	(0.012)	(0.011)	(0.009)
Control Mean	0.167		0.201		0.107	
Full Controls Included		Y		Y		Y
Observations	7,759	7,759	7,759	7,759	7,759	7,759
R-squared	0.001	0.250	0.001	0.259	0.001	0.160

	% Female Non-Nationals Shortlisted		Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.012	0.015**	0.017	0.021**	0.012*	0.012*
	(0.007)	(0.006)	(0.010)	(0.008)	(0.006)	(0.005)
Effectiveness Training	-0.002	-0.000	-0.002	0.000	0.003	0.002
	(0.006)	(0.005)	(0.009)	(0.008)	(0.005)	(0.005)
Control Mean	0.046		0.074		0.029	
Full Controls Included		Y		Y		Y
Observations	7,549	7,549	7,549	7,549	7,549	7,549
R-squared	0.001	0.108	0.001	0.130	0.001	0.068

Notes: Ordinary Least Squares (OLS) estimations. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S3. OLS Regressions Predicting Shortlisting and Hiring of Global South Non-Nationals. Full Sample.

	% Global South Non-Nationals Shortlisted		Global South Non-National Shortlisted (0/1)		Global South Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.019	0.019*	0.026	0.028*	0.014	0.013
	(0.013)	(0.010)	(0.015)	(0.011)	(0.010)	(0.008)
Effectiveness Training	0.011	0.013	0.013	0.013	0.010	0.009
	(0.011)	(0.009)	(0.014)	(0.011)	(0.009)	(0.007)
Control Mean	0.115		0.149		0.071	
Full Controls Included		Y		Y		Y
Observations	7,759	7,759	7,759	7,759	7,759	7,759
R-squared	0.001	0.185	0.001	0.194	0.000	0.116

	% Female Global South Non-Nationals Shortlisted		Female Global South Non-National Shortlisted (0/1)		Female Global South Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.010	0.011*	0.012	0.014*	0.010*	0.009*
	(0.005)	(0.005)	(0.008)	(0.007)	(0.005)	(0.004)
Effectiveness Training	0.002	0.003	0.000	0.002	0.006	0.006
	(0.005)	(0.004)	(0.007)	(0.007)	(0.004)	(0.004)
Control Mean	0.030		0.054		0.017	
Full Controls Included		Y		Y		Y
Observations	7,549	7,549	7,549	7,549	7,549	7,549
R-squared	0.001	0.072	0.001	0.094	0.001	0.043

Notes: Ordinary Least Squares (OLS) estimations. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S4. OLS Regressions Predicting Hiring Behavior Using an Interaction between Treatment Conditions and Income Level of Requisition Country

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.003	0.005	0.007	0.006	-0.007	-0.005
	(0.017)	(0.014)	(0.020)	(0.016)	(0.016)	(0.014)
Effectiveness Training	0.018	0.018	0.020	0.018	0.008	0.007
	(0.019)	(0.015)	(0.021)	(0.017)	(0.018)	(0.014)
High Income Country	0.042*	0.042	0.067**	0.034	0.021	0.044
	(0.016)	(0.043)	(0.019)	(0.055)	(0.016)	(0.049)
Diversity Training*High Income Country	0.021	0.040*	0.022	0.045*	0.024	0.040*
	(0.023)	(0.020)	(0.027)	(0.023)	(0.022)	(0.019)
Effectiveness Training*High Income Country	-0.010	-0.001	-0.006	0.004	0.002	0.009
	(0.024)	(0.020)	(0.027)	(0.023)	(0.023)	(0.020)
Full Controls Included		Y		Y		Y
Observations	10,433	10,433	10,433	10,433	10,433	10,433
R-squared	0.004	0.103	0.007	0.104	0.002	0.091

	% Non-Nationals Shortlisted		Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.017	0.012	0.020	0.013	0.016*	0.009
	(0.012)	(0.010)	(0.014)	(0.011)	(0.008)	(0.008)
Effectiveness Training	0.016	0.005	0.015	0.002	0.011	-0.002
	(0.013)	(0.009)	(0.015)	(0.010)	(0.008)	(0.007)
High Income Country	0.234**	0.189**	0.271**	0.172*	0.163**	0.159**
	(0.016)	(0.067)	(0.018)	(0.073)	(0.013)	(0.057)
Diversity Training*High Income Country	0.024	0.025	0.027	0.029	0.020	0.025
	(0.025)	(0.021)	(0.028)	(0.023)	(0.021)	(0.018)
Effectiveness Training*High Income Country	0.013	0.019	0.019	0.025	0.010	0.019
	(0.025)	(0.019)	(0.028)	(0.022)	(0.019)	(0.017)
Full Controls Included		Y		Y		Y
Observations	7,759	7,759	7,759	7,759	7,759	7,759
R-squared	0.114	0.253	0.121	0.261	0.072	0.162

	% Female Non-Nationals Shortlisted		Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	-0.000	0.003	-0.001	0.004	-0.001	-0.002
	(0.004)	(0.004)	(0.007)	(0.006)	(0.003)	(0.003)
Effectiveness Training	0.001	0.003	-0.002	-0.002	0.005	0.004
	(0.004)	(0.004)	(0.006)	(0.006)	(0.003)	(0.003)
High Income Country	0.066**	0.039	0.102**	0.021	0.046**	0.035
	(0.008)	(0.028)	(0.011)	(0.042)	(0.007)	(0.033)
Diversity Training*High Income Country	0.026*	0.025*	0.037*	0.034*	0.026*	0.027**
	(0.012)	(0.011)	(0.017)	(0.015)	(0.011)	(0.010)
Effectiveness Training*High Income Country	-0.004	-0.006	0.005	0.004	-0.002	-0.003
	(0.011)	(0.010)	(0.016)	(0.014)	(0.010)	(0.009)
Full Controls Included		Y		Y		Y
Observations	7,549	7,549	7,549	7,549	7,549	7,549
R-squared	0.038	0.109	0.048	0.130	0.024	0.070

Notes: Ordinary Least Squares (OLS) estimations using an interaction between treatment conditions and a binary variable if the requisition country is a high-income country according to World Bank country-level income classifications. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S5. OLS Regressions Predicting Hiring Behavior Using an Interaction between Treatment Conditions and Gender Equality of Requisition Country

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.004	0.009	0.002	0.009	0.003	0.008
	(0.017)	(0.015)	(0.020)	(0.017)	(0.016)	(0.014)
Effectiveness Training	0.006	0.004	0.005	0.001	0.002	0.002
	(0.019)	(0.016)	(0.021)	(0.018)	(0.017)	(0.014)
High Gender Equality	0.064**	0.074	0.092**	0.046	0.052**	0.033
	(0.016)	(0.044)	(0.019)	(0.063)	(0.016)	(0.045)
Diversity Training*High Gender Equality	0.020	0.028	0.029	0.035	0.006	0.013
	(0.023)	(0.020)	(0.026)	(0.023)	(0.022)	(0.019)
Effectiveness Training*High Gender Equality	0.009	0.021	0.016	0.031	0.010	0.015
	(0.024)	(0.021)	(0.027)	(0.024)	(0.023)	(0.020)
Full Controls Included		Y		Y		Y
Observations	10,402	10,402	10,402	10,402	10,402	10,402
R-squared	0.009	0.103	0.015	0.104	0.006	0.090

	% Non-Nationals Shortlisted		Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.001	0.002	0.002	0.002	0.003	0.001
	(0.011)	(0.010)	(0.013)	(0.011)	(0.007)	(0.008)
Effectiveness Training	0.005	-0.000	0.005	-0.003	0.003	-0.003
	(0.012)	(0.010)	(0.014)	(0.011)	(0.008)	(0.007)
High Gender Equality	0.209**	-0.175*	0.249**	-0.192*	0.147**	-0.056
	(0.016)	(0.072)	(0.019)	(0.079)	(0.013)	(0.059)
Diversity Training*High Gender Equality	0.051*	0.036	0.057*	0.040	0.041*	0.034
	(0.025)	(0.020)	(0.027)	(0.023)	(0.021)	(0.018)
Effectiveness Training*High Gender Equality	0.014	0.025	0.015	0.029	0.011	0.018
	(0.025)	(0.019)	(0.028)	(0.022)	(0.019)	(0.016)
Full Controls Included		Y		Y		Y
Observations	7,738	7,738	7,738	7,738	7,738	7,738
R-squared	0.098	0.252	0.107	0.261	0.064	0.160

	% Female Non-Nationals Shortlisted		Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	-0.003	0.001	-0.006	0.000	0.001	0.001
	(0.004)	(0.004)	(0.006)	(0.006)	(0.003)	(0.003)
Effectiveness Training	-0.002	-0.002	-0.007	-0.008	0.002	0.002
	(0.005)	(0.004)	(0.006)	(0.006)	(0.004)	(0.003)
High Gender Equality	0.060**	0.039	0.096**	0.021	0.041**	0.015
	(0.007)	(0.024)	(0.011)	(0.047)	(0.006)	(0.031)
Diversity Training*High Gender Equality	0.029**	0.024*	0.042**	0.036*	0.021*	0.020*
	(0.011)	(0.010)	(0.016)	(0.015)	(0.010)	(0.010)
Effectiveness Training*High Gender Equality	-0.002	0.003	0.005	0.014	-0.001	0.001
	(0.010)	(0.009)	(0.015)	(0.014)	(0.009)	(0.009)
Full Controls Included		Y		Y		Y
Observations	7,528	7,528	7,528	7,528	7,528	7,528
R-squared	0.034	0.110	0.044	0.131	0.019	0.069

Notes: Ordinary Least Squares (OLS) estimations using an interaction between treatment conditions and a binary variable if the requisition country is a country with above median gender equality according to the global gender gap index (2023) of the World Economic Forum. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S6. OLS Regressions Predicting Hiring Behavior Using an Interaction between Treatment Conditions and Headquarters Country

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.005	0.016	0.008	0.019	-0.002	0.008
	(0.013)	(0.011)	(0.015)	(0.012)	(0.012)	(0.010)
Effectiveness Training	0.015	0.019	0.020	0.022	0.012	0.014
	(0.013)	(0.011)	(0.015)	(0.013)	(0.013)	(0.011)
Headquarters	0.038*		0.061*		0.023	
	(0.019)		(0.024)		(0.019)	
Diversity Training*Headquarters	0.059*	0.065*	0.067	0.072*	0.049	0.053
	(0.029)	(0.027)	(0.035)	(0.032)	(0.029)	(0.027)
Effectiveness Training*Headquarters	-0.020	-0.015	-0.023	-0.015	-0.020	-0.015
	(0.030)	(0.027)	(0.036)	(0.033)	(0.029)	(0.027)
Full Controls Included		Y		Y		Y
Observations	10,433	10,433	10,433	10,433	10,433	10,433
R-squared	0.004	0.102	0.005	0.104	0.002	0.091

	% Non-Nationals Shortlisted		Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.016	0.007	0.021	0.011	0.014	0.006
	(0.016)	(0.011)	(0.018)	(0.012)	(0.013)	(0.009)
Effectiveness Training	0.018	0.007	0.020	0.007	0.014	0.005
	(0.015)	(0.010)	(0.017)	(0.012)	(0.011)	(0.009)
Headquarters	0.167**		0.203**		0.117**	
	(0.028)		(0.032)		(0.025)	
Diversity Training*Headquarters	0.082*	0.113**	0.077	0.103*	0.072	0.099**
	(0.040)	(0.038)	(0.045)	(0.042)	(0.039)	(0.036)
Effectiveness Training*Headquarters	0.029	0.046	0.028	0.045	0.008	0.019
	(0.042)	(0.038)	(0.047)	(0.043)	(0.036)	(0.033)
Full Controls Included		Y		Y		Y
Observations	7,759	7,759	7,759	7,759	7,759	7,759
R-squared	0.041	0.252	0.044	0.260	0.027	0.162

	% Female Non-Nationals Shortlisted		Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.005	0.005	0.008	0.008	0.004	0.002
	(0.006)	(0.005)	(0.009)	(0.007)	(0.005)	(0.005)
Effectiveness Training	0.003	0.003	0.003	0.002	0.007	0.005
	(0.006)	(0.005)	(0.008)	(0.007)	(0.005)	(0.005)
Headquarters	0.068**		0.114**		0.041**	
	(0.015)		(0.021)		(0.013)	
Diversity Training*Headquarters	0.054*	0.066**	0.073*	0.083**	0.055*	0.067**
	(0.023)	(0.022)	(0.032)	(0.031)	(0.023)	(0.022)
Effectiveness Training*Headquarters	-0.021	-0.020	-0.017	-0.013	-0.020	-0.019
	(0.021)	(0.019)	(0.031)	(0.029)	(0.018)	(0.017)
Full Controls Included		Y		Y		Y
Observations	7,549	7,549	7,549	7,549	7,549	7,549
R-squared	0.026	0.112	0.035	0.133	0.016	0.073

Notes: Ordinary Least Squares (OLS) estimations using an interaction between treatment conditions and headquarters country. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. In columns 2, 4, and 6, as a result of the inclusion of requisition country FE, the headquarters coefficient is dropped. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S7. OLS Regressions Predicting Shortlisting and Hiring in Headquarters Country

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.064*	0.084**	0.075*	0.088**	0.048	0.058*
	(0.026)	(0.026)	(0.032)	(0.031)	(0.026)	(0.026)
Effectiveness Training	-0.004	0.003	-0.002	0.009	-0.007	-0.006
	(0.027)	(0.025)	(0.033)	(0.031)	(0.026)	(0.025)
Control Mean	0.228		0.292		0.185	
Full Controls Included		Y		Y		Y
Observations	1,648	1,648	1,648	1,648	1,648	1,648
R-squared	0.006	0.122	0.006	0.110	0.004	0.096

	% Non-Nationals Shortlisted		Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.098**	0.130**	0.098*	0.120**	0.086*	0.106**
	(0.037)	(0.037)	(0.041)	(0.041)	(0.037)	(0.037)
Effectiveness Training	0.047	0.041	0.048	0.047	0.023	-0.004
	(0.039)	(0.037)	(0.044)	(0.041)	(0.035)	(0.033)
Control Mean	0.307		0.371		0.205	
Full Controls Included		Y		Y		Y
Observations	1,173	1,173	1,173	1,173	1,173	1,173
R-squared	0.008	0.110	0.007	0.107	0.007	0.084

	% Female Non-Nationals Shortlisted		Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.059**	0.079**	0.081*	0.098**	0.059**	0.072**
	(0.023)	(0.022)	(0.032)	(0.031)	(0.022)	(0.022)
Effectiveness Training	-0.018	-0.027	-0.014	-0.015	-0.013	-0.027
	(0.021)	(0.018)	(0.030)	(0.028)	(0.018)	(0.017)
Control Mean	0.102		0.169		0.063	
Full Controls Included		Y		Y		Y
Observations	1,162	1,162	1,162	1,162	1,162	1,162
R-squared	0.014	0.135	0.011	0.125	0.014	0.093

Notes: Ordinary Least Squares (OLS) estimations. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, and requisition unit. Since we analyze only the headquarters country, we are not controlling for requisition country FE. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S8. OLS Regressions Predicting Shortlisting and Hiring of Global South Non-Nationals in Headquarters Country

	% Global South Non-Nationals Shortlisted		Global South Non-National Shortlisted (0/1)		Global South Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.056	0.070*	0.071	0.086*	0.057	0.065*
	(0.032)	(0.033)	(0.040)	(0.040)	(0.030)	(0.031)
Effectiveness Training	0.003	0.001	0.010	0.018	-0.003	-0.010
	(0.031)	(0.031)	(0.038)	(0.038)	(0.027)	(0.028)
Control Mean	0.202		0.267		0.124	
Full Controls Included		Y		Y		Y
Observations	1,173	1,170	1,173	1,170	1,173	1,170
R-squared	0.005	0.084	0.005	0.094	0.006	0.067

	% Female Global South Non-Nationals Shortlisted		Female Global South Non-National Shortlisted (0/1)		Female Global South Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.034	0.044*	0.039	0.049	0.034*	0.036*
	(0.019)	(0.019)	(0.029)	(0.027)	(0.017)	(0.018)
Effectiveness Training	-0.022	-0.025	-0.022	-0.015	-0.016	-0.024
	(0.015)	(0.015)	(0.026)	(0.026)	(0.012)	(0.014)
Control Mean	0.072		0.130		0.039	
Full Controls Included		Y		Y		Y
Observations	1,162	1,159	1,162	1,159	1,162	1,159
R-squared	0.010	0.102	0.005	0.116	0.010	0.057

Notes: Ordinary Least Squares (OLS) estimations. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S9. OLS Regressions Predicting Hiring Behavior Using an Interaction between Treatment Conditions and Hiring Manager Gender

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.009	0.024*	0.011	0.029*	-0.001	0.011
	(0.013)	(0.011)	(0.015)	(0.012)	(0.012)	(0.010)
Effectiveness Training	0.007	0.017	0.008	0.017	0.002	0.008
	(0.013)	(0.011)	(0.016)	(0.013)	(0.013)	(0.010)
Female Manager	0.064**	0.016	0.066**	0.011	0.059**	0.017
	(0.020)	(0.019)	(0.023)	(0.022)	(0.020)	(0.019)
Diversity Training*Female Manager	0.023	0.008	0.032	0.006	0.035	0.025
	(0.032)	(0.028)	(0.037)	(0.032)	(0.032)	(0.028)
Effectiveness Training*Female Manager	0.035	-0.002	0.054	0.017	0.048	0.014
	(0.031)	(0.027)	(0.035)	(0.032)	(0.031)	(0.027)
Full Controls Included		Y		Y		Y
Observations	10,433	10,433	10,433	10,433	10,433	10,433
R-squared	0.008	0.101	0.008	0.103	0.008	0.089

	% Non-Nationals Shortlisted		Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.025	0.020	0.028	0.023	0.023	0.019
	(0.018)	(0.012)	(0.020)	(0.014)	(0.015)	(0.011)
Effectiveness Training	0.013	0.005	0.012	0.002	0.011	0.003
	(0.016)	(0.012)	(0.019)	(0.013)	(0.013)	(0.010)
Female Manager	0.018	-0.024	0.025	-0.031	0.018	-0.018
	(0.022)	(0.018)	(0.025)	(0.020)	(0.019)	(0.016)
Diversity Training*Female Manager	0.009	0.021	0.007	0.020	0.001	0.015
	(0.035)	(0.027)	(0.039)	(0.030)	(0.029)	(0.024)
Effectiveness Training*Female Manager	0.030	0.052*	0.041	0.072*	0.012	0.029
	(0.036)	(0.025)	(0.042)	(0.028)	(0.029)	(0.023)
Full Controls Included		Y		Y		Y
Observations	7,759	7,759	7,759	7,759	7,759	7,759
R-squared	0.002	0.251	0.003	0.260	0.002	0.160

	% Female Non-Nationals Shortlisted		Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.013	0.017**	0.016	0.021*	0.011	0.012*
	(0.007)	(0.006)	(0.011)	(0.009)	(0.006)	(0.005)
Effectiveness Training	-0.007	-0.005	-0.011	-0.010	-0.001	-0.002
	(0.006)	(0.006)	(0.009)	(0.008)	(0.005)	(0.005)
Female Manager	0.023	0.008	0.029	0.001	0.020	0.009
	(0.012)	(0.011)	(0.017)	(0.015)	(0.011)	(0.010)
Diversity Training*Female Manager	-0.005	-0.007	0.005	0.000	0.005	0.005
	(0.018)	(0.016)	(0.025)	(0.022)	(0.017)	(0.016)
Effectiveness Training*Female Manager	0.034	0.028	0.058**	0.058**	0.025	0.024
	(0.020)	(0.015)	(0.028)	(0.022)	(0.018)	(0.015)
Full Controls Included		Y		Y		Y
Observations	7,549	7,549	7,549	7,549	7,549	7,549
R-squared	0.006	0.108	0.007	0.131	0.005	0.068

Notes: Ordinary Least Squares (OLS) estimations using an interaction between treatment conditions and hiring manager gender. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S10. OLS Regressions Predicting Hiring Behavior Using an Interaction between Treatment Conditions and Seniority of Hiring Manager

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.009	0.024*	0.011	0.029*	-0.001	0.011
	(0.013)	(0.011)	(0.015)	(0.012)	(0.012)	(0.010)
Effectiveness Training	0.007	0.017	0.008	0.017	0.002	0.008
	(0.013)	(0.011)	(0.016)	(0.013)	(0.013)	(0.010)
Senior Manager	0.064**	0.014	0.066**	0.010	0.059**	0.016
	(0.020)	(0.019)	(0.023)	(0.022)	(0.020)	(0.019)
Diversity Training*Senior Manager	0.023	0.007	0.032	0.006	0.035	0.024
	(0.032)	(0.028)	(0.037)	(0.032)	(0.032)	(0.028)
Effectiveness Training*Senior Manager	0.035	-0.000	0.054	0.018	0.048	0.016
	(0.031)	(0.027)	(0.035)	(0.032)	(0.031)	(0.027)
Full Controls Included		Y		Y		Y
Observations	10,433	10,433	10,433	10,433	10,433	10,433
R-squared	0.008	0.102	0.008	0.103	0.008	0.091

	% Non-Nationals Shortlisted		Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.025	0.021	0.028	0.023	0.023	0.018
	(0.018)	(0.012)	(0.020)	(0.014)	(0.015)	(0.011)
Effectiveness Training	0.013	0.005	0.012	0.002	0.011	0.003
	(0.016)	(0.012)	(0.019)	(0.013)	(0.013)	(0.010)
Senior Manager	0.018	-0.023	0.025	-0.030	0.018	-0.017
	(0.022)	(0.018)	(0.025)	(0.020)	(0.019)	(0.016)
Diversity Training*Senior Manager	0.009	0.020	0.007	0.019	0.001	0.014
	(0.035)	(0.027)	(0.039)	(0.030)	(0.029)	(0.024)
Effectiveness Training*Senior Manager	0.030	0.051*	0.041	0.071*	0.012	0.029
	(0.036)	(0.025)	(0.042)	(0.028)	(0.029)	(0.023)
Full Controls Included		Y		Y		Y
Observations	7,759	7,759	7,759	7,759	7,759	7,759
R-squared	0.002	0.251	0.003	0.260	0.002	0.160

	% Female Non-Nationals Shortlisted		Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.013	0.017**	0.016	0.021*	0.011	0.012*
	(0.007)	(0.006)	(0.011)	(0.009)	(0.006)	(0.005)
Effectiveness Training	-0.007	-0.005	-0.011	-0.010	-0.001	-0.002
	(0.006)	(0.006)	(0.009)	(0.008)	(0.005)	(0.005)
Senior Manager	0.023	0.007	0.029	0.001	0.020	0.009
	(0.012)	(0.011)	(0.017)	(0.015)	(0.011)	(0.010)
Diversity Training*Senior Manager	-0.005	-0.007	0.005	-0.000	0.005	0.005
	(0.018)	(0.016)	(0.025)	(0.022)	(0.017)	(0.016)
Effectiveness Training*Senior Manager	0.034	0.028	0.058*	0.057*	0.025	0.024
	(0.020)	(0.015)	(0.028)	(0.022)	(0.018)	(0.015)
Full Controls Included		Y		Y		Y
Observations	7,549	7,549	7,549	7,549	7,549	7,549
R-squared	0.006	0.109	0.007	0.131	0.005	0.068

Notes: Ordinary Least Squares (OLS) estimations using an interaction between treatment conditions and a binary variable if the hiring manager has above median seniority (defined based on job levels). Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S11. OLS Regressions Predicting Hiring Behavior Using an Interaction between Treatment Conditions and Team Gender Diversity of Hiring Manager

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.007	0.016	0.018	0.026	-0.004	0.005
	(0.013)	(0.011)	(0.016)	(0.014)	(0.012)	(0.011)
Effectiveness Training	-0.011	-0.010	-0.003	-0.004	-0.026*	-0.027*
	(0.013)	(0.012)	(0.016)	(0.014)	(0.013)	(0.011)
Gender-Diverse Team	0.154**	0.092**	0.179**	0.110**	0.126**	0.068**
	(0.016)	(0.015)	(0.018)	(0.017)	(0.015)	(0.014)
Diversity Training*Gender-Diverse Team	0.012	0.015	-0.004	0.004	0.018	0.017
	(0.023)	(0.020)	(0.027)	(0.023)	(0.022)	(0.020)
Effectiveness Training*Gender-Diverse Team	0.027	0.039	0.019	0.032	0.052*	0.062**
	(0.022)	(0.020)	(0.025)	(0.023)	(0.021)	(0.019)
Full Controls Included		Y		Y		Y
Observations	10,163	10,163	10,163	10,163	10,163	10,163
R-squared	0.048	0.120	0.045	0.119	0.040	0.105

Notes: Ordinary Least Squares (OLS) estimations using an interaction between treatment conditions and a binary variable if the hiring manager has above median team gender-diversity (defined as having above median share of female direct reports). Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Equivalent tables for non-nationals and female non-nationals are not produced because the manager-level file used for this analysis only includes information about the number of female direct reports in the team, but not the number of non-national and female non-national direct reports in the team. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S12. OLS Regressions Predicting Hiring Behavior Using an Interaction between Treatment Conditions and Applicant Pool Diversity

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.001	0.006	0.008	0.014	-0.006	0.000
	(0.004)	(0.008)	(0.007)	(0.010)	(0.004)	(0.008)
Effectiveness Training	0.002	-0.000	0.006	0.002	0.001	0.001
	(0.005)	(0.008)	(0.007)	(0.009)	(0.005)	(0.008)
Diverse Applicant Pool	0.357**	0.342**	0.420**	0.398**	0.305**	0.296**
	(0.014)	(0.015)	(0.016)	(0.017)	(0.015)	(0.016)
Diversity Training*Diverse Applicant Pool	0.032	0.031	0.024	0.023	0.028	0.023
	(0.020)	(0.021)	(0.023)	(0.023)	(0.021)	(0.021)
Effectiveness Training*Diverse Applicant Pool	0.031	0.031	0.031	0.032	0.025	0.019
	(0.021)	(0.021)	(0.023)	(0.024)	(0.021)	(0.022)
Full Controls Included		Y		Y		Y
Observations	10,433	10,433	10,433	10,433	10,433	10,433
R-squared	0.246	0.303	0.255	0.306	0.184	0.241

	% Non-Nationals Shortlisted		Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	-0.001	-0.003	-0.004	-0.006	-0.004	-0.010
	(0.005)	(0.007)	(0.008)	(0.009)	(0.004)	(0.006)
Effectiveness Training	-0.005	-0.007	-0.010	-0.011	-0.006	-0.013*
	(0.004)	(0.007)	(0.007)	(0.009)	(0.004)	(0.006)
Diverse Applicant Pool	0.303**	0.212**	0.339**	0.224**	0.189**	0.132**
	(0.015)	(0.016)	(0.017)	(0.018)	(0.013)	(0.014)
Diversity Training*Diverse Applicant Pool	0.051*	0.056**	0.063*	0.067**	0.052*	0.063**
	(0.024)	(0.021)	(0.026)	(0.023)	(0.022)	(0.019)
Effectiveness Training*Diverse Applicant Pool	0.047*	0.043*	0.059*	0.051*	0.038*	0.043*
	(0.023)	(0.020)	(0.025)	(0.023)	(0.019)	(0.018)
Full Controls Included		Y		Y		Y
Observations	7,759	7,759	7,759	7,759	7,759	7,759
R-squared	0.212	0.305	0.214	0.308	0.116	0.193

	% Female Non-Nationals Shortlisted		Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	-0.000	-0.000	-0.000	0.002	0.000	-0.002
	(.)	(0.003)	(.)	(0.004)	(.)	(0.003)
Effectiveness Training	-0.000	-0.000	-0.000	0.000	0.000	-0.002
	(.)	(0.003)	(.)	(0.004)	(.)	(0.002)
Diverse Applicant Pool	0.090**	0.067**	0.147**	0.105**	0.057**	0.042**
	(0.008)	(0.008)	(0.011)	(0.011)	(0.007)	(0.007)
Diversity Training*Diverse Applicant Pool	0.024*	0.031*	0.034	0.038*	0.023*	0.029*
	(0.012)	(0.012)	(0.017)	(0.017)	(0.011)	(0.011)
Effectiveness Training*Diverse Applicant Pool	0.001	0.003	0.005	0.004	0.010	0.009
	(0.011)	(0.011)	(0.016)	(0.015)	(0.010)	(0.010)
Full Controls Included		Y		Y		Y
Observations	7,549	7,549	7,549	7,549	7,549	7,549
R-squared	0.068	0.136	0.089	0.162	0.037	0.084

Notes: Ordinary Least Squares (OLS) estimations using an interaction between treatment conditions and a binary variable (Diverse Applicant Pool) if the applicant pool for the job has above-median share of female (model 1 and 2), non-national (model 3 and 4) or female non-national applicants (model 5 and 6). A coefficient estimate cannot be produced in rare cases where there are too few observations of female non-nationals in the below-median diversity pool. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S13. OLS Regressions Predicting Hiring Behavior Using an Interaction between Treatment Conditions and Hiring Manager's First Requisition

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.033	0.039*	0.043*	0.051**	0.026	0.029
	(0.017)	(0.017)	(0.020)	(0.019)	(0.017)	(0.017)
Effectiveness Training	0.014	0.015	0.022	0.024	0.022	0.018
	(0.017)	(0.017)	(0.020)	(0.019)	(0.018)	(0.017)
First Requisition	-0.033*	-0.023	-0.055**	-0.033*	-0.023	-0.020
	(0.014)	(0.013)	(0.017)	(0.016)	(0.014)	(0.014)
Diversity Training*First Requisition	-0.029	-0.019	-0.038	-0.030	-0.030	-0.018
	(0.021)	(0.019)	(0.024)	(0.022)	(0.021)	(0.019)
Effectiveness Training*First Requisition	-0.004	0.002	-0.009	-0.005	-0.019	-0.009
	(0.021)	(0.019)	(0.025)	(0.023)	(0.021)	(0.020)
Full Controls Included		Y		Y		Y
Observations	10,433	10,433	10,433	10,433	10,433	10,433
R-squared	0.003	0.103	0.006	0.105	0.003	0.092

	% Non-Nationals Shortlisted		Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.013	0.012	0.022	0.022	0.023	0.020
	(0.020)	(0.018)	(0.022)	(0.020)	(0.018)	(0.017)
Effectiveness Training	0.005	0.004	0.010	0.009	0.008	0.004
	(0.020)	(0.017)	(0.022)	(0.019)	(0.018)	(0.017)
First Requisition	-0.103**	-0.049**	-0.116**	-0.049**	-0.068**	-0.033*
	(0.016)	(0.014)	(0.017)	(0.016)	(0.014)	(0.013)
Diversity Training*First Requisition	0.021	0.019	0.012	0.008	0.001	0.002
	(0.024)	(0.020)	(0.027)	(0.022)	(0.022)	(0.019)
Effectiveness Training*First Requisition	0.020	0.016	0.016	0.008	0.008	0.007
	(0.023)	(0.020)	(0.026)	(0.022)	(0.021)	(0.019)
Full Controls Included		Y		Y		Y
Observations	7,759	7,759	7,759	7,759	7,759	7,759
R-squared	0.014	0.252	0.016	0.261	0.010	0.161

	% Female Non-Nationals Shortlisted		Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.024*	0.026*	0.036*	0.040**	0.019	0.018
	(0.011)	(0.010)	(0.016)	(0.015)	(0.011)	(0.010)
Effectiveness Training	0.003	0.007	0.006	0.009	0.011	0.010
	(0.010)	(0.010)	(0.015)	(0.014)	(0.010)	(0.010)
First Requisition	-0.021*	-0.000	-0.040**	-0.007	-0.015*	-0.002
	(0.008)	(0.008)	(0.012)	(0.012)	(0.008)	(0.007)
Diversity Training*First Requisition	-0.017	-0.017	-0.029	-0.028	-0.011	-0.008
	(0.013)	(0.012)	(0.018)	(0.017)	(0.012)	(0.011)
Effectiveness Training*First Requisition	-0.008	-0.010	-0.011	-0.013	-0.012	-0.012
	(0.012)	(0.011)	(0.017)	(0.016)	(0.011)	(0.011)
Full Controls Included		Y		Y		Y
Observations	7,549	7,549	7,549	7,549	7,549	7,549
R-squared	0.006	0.109	0.010	0.131	0.004	0.068

Notes: Ordinary Least Squares (OLS) estimations using an interaction between treatment conditions and a binary variable if the requisition is manager's first requisition. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S14. Two Stage Least Squares (2SLS) Regressions Predicting Hiring Behavior Using Random Assignment of Treatment Condition as an Instrument for Video Play

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Diversity Training	0.079	0.157**	0.098	0.195**	0.030	0.097
	(0.073)	(0.061)	(0.086)	(0.070)	(0.068)	(0.058)
Full Controls Included		Y		Y		Y
Observations	7,014	7,014	7,014	7,014	7,014	7,014
Panel B						
Effectiveness Training	0.056	0.092	0.076	0.113	0.043	0.056
	(0.060)	(0.052)	(0.069)	(0.060)	(0.057)	(0.049)
Full Controls Included		Y		Y		Y
Observations	6,935	6,935	6,935	6,935	6,935	6,935

	% Non-Nationals Shortlisted		Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel C						
Diversity Training	0.150	0.157**	0.169	0.173*	0.134	0.131*
	(0.089)	(0.060)	(0.100)	(0.067)	(0.074)	(0.052)
Full Controls Included		Y		Y		Y
Observations	5,167	5,167	5,167	5,167	5,167	5,167
Panel D						
Effectiveness Training	0.082	0.058	0.088	0.067	0.059	0.032
	(0.067)	(0.049)	(0.077)	(0.055)	(0.052)	(0.042)
Full Controls Included		Y		Y		Y
Observations	5,170	5,170	5,170	5,170	5,170	5,170

	% Female Non-Nationals Shortlisted		Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel E						
Diversity Training	0.070	0.090**	0.096	0.129**	0.067*	0.071*
	(0.038)	(0.031)	(0.055)	(0.044)	(0.034)	(0.029)
Full Controls Included		Y		Y		Y
Observations	5,036	5,036	5,036	5,036	5,036	5,036
Panel F						
Effectiveness Training	-0.008	0.003	-0.007	0.009	0.014	0.012
	(0.027)	(0.024)	(0.039)	(0.035)	(0.024)	(0.022)
Full Controls Included		Y		Y		Y
Observations	5,014	5,014	5,014	5,014	5,014	5,014

Notes: Two Stage Least Squares (2SLS) estimations. We estimate the effect of one condition (the diversity training condition or the effectiveness training condition) against the control by dropping the other condition observations from the sample so as to be able to estimate each treatment effect separately (rather than an averaged effect of the two conditions combined). As mentioned in the text above, we caution against overinterpreting the analyses due to potential tracking data quality issues – that is, we cannot rule out the possibility that some managers may have watched the video but not tracked accordingly. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S15. Logistic Regressions Predicting Binary Shortlisting and Hiring Behavior

	Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)
Diversity Training	0.086	0.173*	0.035	0.118
	(0.074)	(0.068)	(0.080)	(0.075)
Effectiveness Training	0.083	0.121	0.062	0.092
	(0.075)	(0.070)	(0.082)	(0.076)
Full Controls Included		Y		Y
Observations	10,433	10,431	10,433	10,417

	Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(3)	(4)	(5)	(6)
Diversity Training	0.175	0.254**	0.225	0.288**
	(0.102)	(0.096)	(0.120)	(0.108)
Effectiveness Training	0.116	0.159	0.129	0.147
	(0.101)	(0.095)	(0.113)	(0.106)
Full Controls Included		Y		Y
Observations	7,759	7,681	7,759	7,655

	Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(3)	(4)	(5)	(6)
Diversity Training	0.227	0.365**	0.357*	0.467**
	(0.126)	(0.124)	(0.175)	(0.170)
Effectiveness Training	-0.023	0.032	0.104	0.121
	(0.129)	(0.129)	(0.181)	(0.179)
Full Controls Included		Y		Y
Observations	7,549	7,318	7,549	6,847

Notes: Logistic regression estimates. Variations in sample size occur for a given outcome variable due to dropped observations in logistics regressions when fixed effects are included. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Table S16. OLS Regressions Predicting Shortlisting and Hiring Behavior Conditional on Filled Requisitions (i.e. Hiring Someone for the Role)

	% Female Shortlisted		Female Shortlisted (0/1)		Female Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.015	0.026*	0.016	0.027	0.011	0.024
	(0.015)	(0.013)	(0.017)	(0.014)	(0.016)	(0.014)
Effectiveness Training	0.011	0.017	0.012	0.016	0.012	0.017
	(0.015)	(0.013)	(0.017)	(0.014)	(0.016)	(0.014)
Control Mean	0.251		0.294		0.255	
Full Controls Included		Y		Y		Y
Observations	6,764	6,764	6,764	6,764	6,764	6,764
R-squared	0.000	0.110	0.000	0.112	0.000	0.110

	% Non-Nationals Shortlisted		Non-National Shortlisted (0/1)		Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.042*	0.029*	0.042	0.025	0.044*	0.030*
	(0.021)	(0.014)	(0.023)	(0.016)	(0.022)	(0.015)
Effectiveness Training	0.013	0.004	0.011	-0.000	0.015	0.004
	(0.019)	(0.013)	(0.022)	(0.015)	(0.019)	(0.014)
Control Mean	0.196		0.231		0.189	
Full Controls Included		Y		Y		Y
Observations	4,428	4,428	4,428	4,428	4,428	4,428
R-squared	0.002	0.304	0.002	0.309	0.002	0.278

	% Female Non-Nationals Shortlisted		Female Non-National Shortlisted (0/1)		Female Non-National Hired (0/1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity Training	0.023*	0.019*	0.030*	0.023*	0.022*	0.017
	(0.010)	(0.008)	(0.013)	(0.011)	(0.010)	(0.009)
Effectiveness Training	-0.001	-0.005	-0.001	-0.005	0.003	-0.001
	(0.009)	(0.008)	(0.012)	(0.011)	(0.009)	(0.008)
Control Mean	0.054		0.082		0.052	
Full Controls Included		Y		Y		Y
Observations	4,229	4,229	4,229	4,229	4,229	4,229
R-squared	0.003	0.122	0.002	0.147	0.002	0.112

Notes: Ordinary Least Squares (OLS) estimations. Sample of requisitions that were filled. Full controls represent the preregistered specification and include manager gender, manager age, manager tenure and tenure squared, manager level, requisition level, requisition function, requisition unit, and requisition country. Robust standard errors are clustered at the manager level and are in parentheses. ** p<0.01, * p<0.05

Data Access Request Form: Behaviorally designed training leads to more diverse hiring

Thank you for your interest in our work.

To request access to the data used, each institution has to submit a data access request form and needs to arrange signed Data Sharing/Use and Non-Disclosure agreements for all researchers to be covered under the agreements. Researchers from the same institution can be named on the same data request form. Once completed, please email the form and all attachments to experiment@morethannow.co.uk.

Named researcher(s):	
Institution:	
Legal team email address:	
Date:	

I have obtained ethics approval (e.g. IRB, Ethics Board): ☐ Yes ☐ No
Please check here that you have attached the ethics approval: ☐ Attached

I have created a Data Management Plan (DMP): ☐ Yes ☐ No
Please check here that you have attached the DMP: ☐ Attached

Reason for seeking access to the data:

Please check here that you are willing to sign a Non-Disclosure Agreement with the field partner to protect their identity: ☐ Yes

Please check here that you are willing to set up a Data Sharing/Use Agreement with the field partner to ensure the data is securely handled: ☐ Yes

Upon submitting this request, MoreThanNow will share the request with the field partner. The field partner will review the request and make a decision as to whether they are willing to consider sharing their data with the named researchers. If they approve the request, the field partner will send Non-Disclosure and Data Sharing/Use Agreements with the legal contact listed above, which must be executed prior to data being shared. Ultimate decision authority on whether data requests are approved lies with the field partner.

Signatures for all researchers at the same institution:

Researcher 1 name: _____
Researcher 1 signature: _____

Researcher 2 name: _____
Researcher 2 signature: _____

[Add more if needed]