IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16598

Discrimination against Women in Hiring

Adam Osman
Jamin D. Speer
Andrew Weaver

NOVEMBER 2023

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16598

Discrimination against Women in Hiring

Adam Osman
University of Illinois at Urbana-Champaign and IZA

Jamin D. Speer
University of Memphis and IZA

Andrew Weaver
University of Illinois at Urbana-Champaign

NOVEMBER 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Discrimination against Women in Hiring*

We study discrimination in hiring and its associated outcomes for the discriminators using a unique survey of Egyptian businesses. Discrimination against women is widespread and overt: about half (51%) of establishments directly admit that they prefer to hire men. The share varies widely across industries, from 60% in retail to only 16% in IT. Using a list randomization technique, we can rule out that discrimination against women is heavily stigmatized in Egypt, meaning that establishments are willing to admit it openly. We then provide novel suggestive evidence showing that discriminating against women is associated the employment of lower-quality workers. We also provide guidance on the use of the list randomization technique and how to interpret it in settings with limited stigma.

JEL Classification: J16, J71, C83, O12

Keywords: discrimination, list randomization, Egypt

Corresponding author:
Adam Osman
Department of Economics
University of Illinois at Urbana-Champaign
109 David Kinley Hall, 1407 W.Gregory
Urbana, IL61801
USA

E-mail: aosman@illinois.edu

* The authors thank Sarah Wahby, Abdelrahman Nagy, Mohamed Yousef, and Mohsen Nagy from J-PAL; Ismail Habrouk, Nora Abou El Seoud, Farah Osman, and Theo Wilson from EFE; Iris Boutros, Arjan De Haan, Rania Fazah and Flaubert Mbiekop from IDRC; Mohamed El Komi and Ebaa El Sayed from AUC; Ali Abdullah and Alia El-Mahdi from EgyPols; and the participants in the study. UIUC IRB Approval #16766, Memphis IRB Approval #4166. We are thankful for financial support from the International Development Research Centre. Authors retained full intellectual freedom to design, analyze and interpret results. All errors are our own.

1 Introduction

Gender discrimination is a topic of major concern for reasons related to both equity and potential economic impact. Numerous studies have demonstrated that gender influences labor market outcomes, albeit in complex ways (Altonji and Blank, 1999; Baert, 2018; Blau and Kahn, 2017; Goldin, 2014; Neumark, 2018). While the existence of discrimination in many contexts has been well-documented, far less is known about exactly how widespread and overt it is, the characteristics of employers that discriminate against women, or the associated employment outcomes at the employer level (Duflo, 2012; Kabeer and Natali, 2013; Kawaguchi, 2007). In particular, do gender-biased employers show signs of lower productivity or higher costs?

These issues are particularly important in a developing country context where discrimination is common (Jayachandran, 2015). There exists macroeconomic evidence that gender discrimination holds back economic growth (Klasen and Lamanna, 2009); however, these results are mixed (Baliamoune-Lutz and McGillivray, 2007; Kabeer and Natali, 2013). Evidence on discriminatory attitudes and behavior among individual employers in such a context, and the potential consequences of that discrimination, is difficult to come by (Demirgüç-Kunt et al., 2013; Knight and Sabot, 1982).

In this paper, we study discriminatory attitudes towards women in hiring decisions and the outcomes associated with those attitudes for employers. Using a detailed, original survey of Egyptian businesses along with a list randomization methodology, we provide evidence both on how willing businesses are to admit discriminatory attitudes and how stigmatized discrimination against women is in Egypt. We also provide guidance on how research designs using list randomization interact with choices about sample size and list construction. In addition, our survey collects data on hiring needs, including detailed data on the types of skills demanded, as well as data on hiring outcomes. We use these data to look at the correlates of discrimination by connecting establishments' hiring outcomes to their biases. Rather than looking at the macro-level consequences of discrimination, we are able to look at the characteristics and employment dynamics associated with discrimination for individual employers.

We find that discrimination against women in hiring is widespread, overt, and is associated with indicators of lower productivity for the employer. Over half (51%)

of establishments directly admit preferring to hire men over women. This share is highest in retail (60%) and lowest in information technology (17%).[1] To study whether these reports are honest representations of beliefs, we use a list randomization method, which allows establishments to admit discriminatory preferences in an anonymized way (Chuang et al., 2021; Karlan and Zinman, 2012; Tsuchiya et al., 2007). This approach gives an estimate of discriminatory attitudes that is statistically similar to the direct question method and allows us to rule out that discrimination against women is heavily stigmatized in Egypt, despite the large standard errors it produces. The willingness of establishments to admit directly to these attitudes enables us to interpret our direct survey responses as accurate measures of discrimination at the establishment level.[2]

We then turn to suggestive evidence on the outcomes associated with discrimination against women for the employers. Even descriptive evidence on this question is hard to come by, because it is typically difficult to accurately identify which firms discriminate. Because establishments in our survey are seemingly open about their discriminatory attitudes, however, we can look at how these biases correlate with hiring outcomes, even after including a host of control variables. While we cannot claim to identify causal relationships, this evidence is still novel, and the correlated fact patterns that we establish regarding gender discrimination and employer characteristics can guide future causal research designs.

We find that discriminatory attitudes toward hiring women in Egypt are associated with reduced productivity. Discriminating establishments end up with lower-quality workers than the non-discriminators; their new hires take about 1.4 weeks, or 24%, longer to reach an acceptable level of productivity, even when controlling for a large set of key factors, including skill requirements, educational profiles, and job quality. This correlation with lower productivity is particularly pronounced among informal, single-site establishments (a 36% longer time to productivity for those who discriminate), which are common in Egypt and other developing countries. While establishments

[1] This result is similar to Kuhn and Shen (2013), who find that gender discrimination in Chinese job ads is less prevalent as skill requirements increase.

[2] While our focus here is on discrimination against women, we also asked establishments about discrimination by socioeconomic status (SES), again using the list randomization method in combination with direct questions. In contrast to the gender results, we find that, relative to the indirect measures of bias, establishments are reluctant to directly admit putting more weight on SES than on workers' skills. We present these results in the Appendix.

that are larger in size discriminate less frequently, they also see a larger productivity disadvantage when they do discriminate (62% longer time to productivity). Although our data do not allow us to establish whether gender discrimination leads to these low-productivity outcomes or whether low-productivity employers are simply more likely to discriminate, the fact that we do not find any significant correlations between gender discrimination and positive labor-market outcomes tends to support the idea that gender discrimination in this context is taste-based rather than statistical (see Section 2 below).

We make three key contributions to the literature. Our primary contribution is to document widespread discriminatory attitudes - and a willingness to admit to these attitudes - among the most common global employer: small and medium-sized enterprises in a developing country. Discrimination is notoriously difficult to detect accurately in observational data. Methods such as the Kitagawa-Oaxaca-Blinder decomposition (Blinder, 1973; Kitagawa, 1955; Oaxaca, 1973) may provide bounds on the extent of discrimination, but are limited by what variables can be directly observed. The most common way of assessing the "true" prevalence of discrimination in hiring is through correspondence studies, in which (typically) fake resumes are submitted to companies with random changes in race, gender, or other characteristics (Baert, 2018; Bertrand and Mullainathan, 2004; Deming et al., 2016; Kline et al., 2021; Kroft et al., 2013). In the context of developing countries, a correspondence study is less feasible because many smaller, informal employers do not use either resumes or job websites in their hiring processes.[3]

Using text from online job ads, Helleseter et al. (2020) and Kuhn and Shen (2013) find that it is common in Mexico and China for employers to explicitly state a preference for male or female workers, depending on the type of job. Our results suggest that relying on online job ads would undercount the true extent of discrimination in our context, because these biased attitudes are more prevalent among smaller, informal establishments, who are the least likely to advertise jobs online. Our surveys allow us to reach employers who would not be reached through other methods.

Second, we provide guidance to researchers regarding the use of list randomization

[3] Banerjee et al. (2009), Siddique (2011), and Arceo-Gomez and Campos-Vazquez (2014) are exceptions, but their studies are limited to jobs advertised online.

to measure potentially stigmatized phenomena. Similar to Tsuchiya et al. (2007), our results indicate that sample size for the indirect question in the list randomization technique has an important effect on standard errors and the precision of estimates. Likewise, our results also imply that there is a trade-off between thoroughly "concealing" the potentially stigmatized statement by using a larger number of non-key statements and the size of the standard errors in the estimate of stigma or discrimination. We provide a detailed discussion of these results and how they can guide other researchers using this method.

Third, we provide evidence about gender discrimination at the establishment level, allowing us to demonstrate detailed conditional correlations between employers' gender preferences and productivity-relevant outcomes. Other work has attempted to use gender diversity as a proxy for discrimination, with conflicting results (Azmat and Boring, 2020; Kawaguchi, 2007; Liu et al., 2010; Tsou and Yang, 2019; Weber and Zulehner, 2014). Using observed gender diversity as a proxy limits the ability for researchers to directly study discrimination because firm gender composition could be due to an inability to hire women as opposed to an unwillingness to hire women. While there is a literature on how discrimination affects economic growth and macro outcomes (e.g., Esteve-Volart, 2000), it is difficult to find evidence on the characteristics of individual discriminating employers.

Our survey allows us to directly study the hiring outcomes associated with discrimination against women because of Egyptian employers' seeming honesty about their gender biases. Because we can reject heavy stigmatization, and thus have confidence that the direct response to the gender discrimination question is not an extreme underestimate, we can examine how employer admissions of discrimination correlate with their hiring outcomes. While our results are only suggestive and not causal impacts of discrimination, we also collect detailed data on the skills these employers demand when hiring workers. By controlling for these factors in our analysis, we increase the likelihood that we are picking up effects related to discriminatory attitudes rather than those driven by differing skill demands.

The paper proceeds as follows. Section 2 discusses our survey and the methods we use to detect discrimination and its consequences. Section 3 documents discrimination against women and studies the related outcomes for the discriminating employers.

Section 4 concludes.

2 Local Context, Our Survey and Techniques

Our study takes place in Egypt, a middle-income country with a PPP-adjusted GDP per-capita of about $12,000. Unemployment in Egypt has been persistently high for years, hovering around 11-12% from 2016 to 2019. Female labor force participation in Egypt is only 15% as of 2021, among the lowest in the world and even with Iran (ILO, 2016). While women in Egypt do not face the same kind of legal restrictions as in some Arab countries, labor market and other outcomes remain poor. The United Nations' 2019 Gender Inequality Index ranked Egypt 108th out of 162 countries for gender equality.

There are likely multiple factors that contribute to the poor labor market outcomes and opportunities for women in Egypt, but we focus on discriminatory attitudes in the labor market. While the Egyptian constitution says that "The state commits to achieving equality between women and men in all civil, political, economic, social, and cultural rights", and discrimination in wages by sex is illegal, enforcement actions associated with these behaviors is uncommon. In our field work interviewing Egyptian policymakers, businesses, and NGOs, we heard widespread anecdotal evidence that firms either prefer to hire men or simply refuse to hire women. These impressions are backed up by survey evidence: in 2017, a study by UN Women reported that 87% of Egyptian men – and 77% of Egyptian women – believe that women's primary role is to be housewives (El Feki and Barker, 2017). While these shares are lower in urban areas, where our survey was conducted, they remain high across Egypt. The United Nations' 2021-22 Gender Social Norms Index found that Egypt ranked last in the world based on the percentage of people biased against women in their views on economic opportunity and empowerment (93.6%). These findings show that discriminatory attitudes are widespread in Egypt, but it remains to be seen how such attitudes play out in the labor-market actions of Egyptian employers.

Types of Discrimination

In the course of our field research, labor market actors and analysts put forth several explanations for gender discrimination in Egypt. Some claimed that discrimination stemmed from traditional ideas about proper gender roles, while others emphasized a belief that men are more reliable workers than women. In other words, some assertions implied taste-based discrimination, while others were potentially consistent with statistical discrimination (Arrow, 1973; Phelps, 1972).

It is worth discussing what we can and cannot learn regarding these different types of discrimination from our study, as this is an important issue for public policy. While we will not be able to distinguish with certainty, we will provide some suggestive evidence on which type of discrimination is more common. The relationship between taste-based discrimination and worker productivity is theoretically ambiguous. We can divide the predicted correlations into a causal and a non-causal scenario. To the extent that gender discrimination has causal effects on the output and productivity of employers, the key issue is whether the number of jobs in non-discriminating employers exceeds the number of job-seekers who are discriminated against (Becker, 1957). If there are a sufficient number of non-discriminating positions, then segregation will occur, but there will be no correlation between taste-based discrimination and productivity. If, however, there are not sufficient positions, then taste-based discrimination will, on average, result in reduced productivity as discriminating employers prioritize non-productive attributes (such as gender) over marginal productivity (Arrow, 1973).[4]

In a non-causal taste-based scenario, it might be the case that productivity is primarily determined by other factors, but that employers with a certain level of productivity tend to discriminate more or less than others. If low-productivity establishments discriminate more, then discriminatory employer attitudes will be correlated with lower productivity. Alternatively, if high-productivity establishments discriminate more – perhaps because industry rents shield the employer from economic consequences (Hellerstein et al., 2002) – then taste-based discrimination will be correlated with higher productivity.

[4]There are also important impacts on job seekers, who may benefit from knowing which firms discriminate so that they can avoid spending time and effort applying to the discriminators (Black, 1995; Pager and Pedulla, 2015).

We can now compare these cases with statistical discrimination, which exists when employers engage in discrimination for rational, productivity-enhancing reasons (i.e., one group has more relevant skills than another group on average). In the statistical case, discriminatory attitudes should be correlated with higher productivity. The exception to this outcome would be a case in which it is primarily low-productivity employers who engage in statistical gender discrimination, and the productivity boost from this action is not sufficient to allow these low-productivity establishments to "catch up" to their high productivity counterparts.

Note that although we cannot arrive at any hard-and-fast conclusions by examining the cross-sectional correlations in our data, we can nevertheless use these patterns to set some tentative boundaries on the nature of gender discrimination. In the event of a correlation between discriminatory attitudes and positive labor market outcomes, such as higher productivity, we cannot distinguish between statistical discrimination and taste-based discrimination (whether causal or non-causal). However, if significant correlations exist between discriminatory attitudes and negative labor-market outcomes (such as lower productivity), these correlations point towards one of two scenarios. On the one hand, they are broadly consistent with a taste-based model of discrimination (causal or non-causal). On the other hand, if statistical discrimination is at play, then such discrimination is likely concentrated among low-productivity employers while at the same time yielding comparatively small productivity benefits. In our empirical specifications, we control for establishment size, formality and firm size, as well as other factors (such as the education profile of the workforce) that distinguish establishment productivity types and minimize selection. Although our results regarding discrimination type are not definitive, we argue that our findings suggest that gender discrimination in Egypt is either taste-based in origin (causal or non-causal) or is associated with negative selection and limited productivity benefits in the statistical case.

Knowing these boundaries is potentially important for policy. If employers are willing to hire women but are using gender as a statistical shortcut to obtain more skilled or productive workers, then it is necessary to focus policy attention on disparities in education and training (which, of course, may stem from discriminatory attitudes in other institutions). Alternatively, if employers are indulging their taste-based prejudice by rejecting qualified female candidates, policy interventions must ultimately seek to change employer behaviors. Our results indicate that the approaches that would have the greatest likelihood of addressing the types of gender discrimination that exist in the Egyptian labor market would be broad anti-discrimination policy efforts, perhaps coupled with education/training initiatives targeted at the low end of the labor market.

The Survey

To study these issues, we designed and conducted a survey of Egyptian businesses from four different industries in 2017-18. We surveyed a total of 1,076 establishments from the retail, information technology (IT), hotels, and restaurant sectors.[5] The sample was weighted using Egypt's 2018 enterprise census as a guide to ensure that the distribution of firm size by industry was nationally representative. In the first step of our approach, the survey firm identified locations that include well-known industry-specific clusters of establishments and aimed to survey the universe of establishments in that area. This strategy was implemented because we did not have access to a detailed list of all establishments in the country and so could not implement a random sampling methodology. We then produced sample weights that ensure that our sample lines up with the distribution of retail enterprises in Egypt using the census based employee sized groupings (5-9 employees, 10-24, 25-49, 50-99, 100-249 and 250+). The current sampling method likely leads to oversampling somewhat more capable and productive establishments, as they have been able to survive in competitive clusters.

The surveys were conducted via interviews with human resources or hiring managers at each establishment. Upon arrival, the interviewer asked to speak with a manager who has authority over personnel and hiring matters.[6] Each survey took approximately one hour and contained about 80 questions.[7]

[5]We interviewed an additional 108 establishments who ended up having fewer than 5 employees. We do not include them in the analysis because the enterprise census does not provide weights for these smaller businesses. The retail portion of the survey was also used for the analysis in Osman et al. (2022), where we study network-based hiring.

[6]In larger establishments, there may not be a single person who fits this description, and it is possible that the views of the person interviewed do not represent the views of the entire establishment.

[7]In our experience, success in implementing surveys is very context specific. It requires a great deal of localization, both with the choice and wording of the questions, as well as in the choice and training of the surveyors. An essential element is in choosing surveyors who are honest, hardworking

The sample is spread out geographically over Egypt. The greater Cairo area represents the largest population center in Egypt, and establishments there account for about 45% of our sample. Establishments in the greater Alexandria region, the second largest city in Egypt, account for 33% of our sample. The remaining 22% come from the southern population centers of Al-Minya and Assuit.

List Randomization

In addition to gathering large amounts of data on each establishment's characteristics and practices, we used a list randomization technique to elicit a measure of the share of establishments that discriminate. List randomization (also known as the item count or list experiment technique, introduced in Raghavarao and Federer (1979) and Miller (1984)) is a survey technique used to ask sensitive questions that might not elicit true responses if asked directly. In this technique, a portion of the sample, chosen randomly, was shown a short list of "non-key" statements that have nothing to do with the sensitive behavior we are interested in (discrimination against women). These include things like, for example, "My household owns an air conditioner." This portion of the sample is simply asked how many of the statements are true (not which ones are true).

The remainder of the sample was given the same list of statements, but with one added. The added statement is the "key" or "sensitive" statement – in our case, "When hiring a new employee, I prefer to hire men over women."[8] The respondents were also asked how many of these statements are true. Because the two groups are chosen randomly, one can interpret the difference in the mean number of true statements between the two groups as the proportion of the entire sample that agrees with the sensitive statement. Several studies have established that list randomization gives higher (and presumably more accurate) estimates of the prevalence of socially unacceptable

and charismatic, the last of which is hard to measure. We have also found that having local surveyors (ideally from the same neighborhood of the enterprises) helps because the basis of all of these interactions is in engendering trust between the surveyor and the respondent, and people often more likely to trust a neighbor than they are a stranger.

[8] We do not say anything about if the men and women who would be hired have the same skills or productivity. The benefit of this general statement is that it is a more natural question, and it allows us to collect *both* taste-based and statistical discrimination from the employers in the sample. The full list of statements we use is in Appendix Table A5.

or stigmatized behaviors (Holbrook and Krosnick, 2010; LaBrie and Earleywine, 2000; Tsuchiya et al., 2007), while it gives similar or even lower estimates for behaviors that are socially acceptable (Tsuchiya et al., 2007).[9]

For the half of the sample that was not given the sensitive statement about discrimination in the list randomization, we directly ask the sensitive question.[10] The results from these responses give us an idea of whether the discriminatory behavior is socially stigmatized or not. If discrimination against women is socially acceptable, for example, the estimates from direct elicitation and the list randomization should be similar. If discrimination is socially stigmatized, then list randomization should give a higher estimate than the direct question.[11] There are several important research design choices regarding list randomization that go beyond typical survey design guidelines regarding balanced treatment and control groups. We discuss these items in detail below.

Establishment Characteristics

We are interested in which establishment characteristics are associated with discriminatory attitudes. We are also interested in the outcomes associated with discriminating against women, which we can only measure accurately if establishments are honest about their attitudes. Our survey is highly detailed, with data on establishment-level labor-market outcomes, operational characteristics, and skill requirements. In this paper, we focus on three key hiring outcomes: average time needed to hire workers (a measure of the difficulty in finding workers), terminations (involuntary turnover), and average time it takes a worker to reach acceptable productivity. We interpret the latter item as a measure of the quality or initial productivity of the worker being hired, as all questions relate to the same type of entry-level job (within industry). In our regression analysis, we will also use several other establishment characteristics that come from

[9]For a detailed discussion of the benefits and potential limitations of list randomization, see Karlan and Zinman (2012) and Chuang et al. (2021).

[10]Based on our field work, this is not a context in which the announcement of discrimination (as opposed to the actual action of discrimination) is culturally encouraged (unlike some areas in Taliban-controlled Afghanistan, for example). As a result, we do not believe employers would claim to discriminate but not actually discriminate.

[11]Tsuchiya et al. (2007) find that list randomization gives a higher estimate than direct elicitation for stigmatized behavior (shoplifting) but not for more acceptable behavior (blood donation). LaBrie and Earleywine (2000) have a similar finding for unprotected sex (stigmatized) and drinking alcohol (acceptable) among undergraduate students.

the survey, including establishment size, use of formal vs. informal labor, whether the establishment is part of a larger firm (e.g., a multi-site firm), average pay, etc.

Because assertions of discrimination are often confounded by the differing skill demands across jobs, we also collected data on skill demands. We have detailed data on the skills demanded by each establishment for entry-level jobs. Within each of the four industries, we used a pilot survey to determine the most common entry-level job. Then, in our main survey, we asked about the importance of a variety of skills in that entry-level job.[12] This allows us to distinguish between establishments that demand "higher-level" or more specialized skills from those that do not.

Ultimately, our survey provides estimates of the share of establishments who admit discrimination (via the direct question) and reveals how stigmatized those attitudes might be (via the list randomization). At the same time, the rich set of establishment characteristics and the skills they demand increase confidence that any relationship we observe between discrimination and outcomes is not driven by skill requirements or other key observables.

3 Discrimination Against Women

Documenting Discrimination

We first document discriminatory attitudes against women in hiring using the direct question about bias and the list randomization method outlined in the previous section. The results are shown in Table 1. About half of the sample (541 establishments, randomly chosen) was asked directly, "When hiring a new employee, do you prefer to hire men over women?" The top row of Table 1 shows that over half (51.4%) of establishments asked this question admit to preferring men over women in hiring decisions.[13] The share of establishments that admit discrimination against women varies considerably by industry. It is highest in retail (60.1%) and lowest in IT (16.4%). While the standard errors are too large to draw firm conclusions about how the four

[12]The jobs we ask employers about, as identified by our pilot survey, are: salesperson (retail), junior programmer (IT), waiter (restaurants), and front desk/reception workers (hotels).

[13]All results in this table are weighted by establishment size to match the Egyptian census of establishments. Results are robust to not using establishment weights.

industries rank, IT seems to stand out from the others for its low rate of admitting to discriminatory attitudes.

The second line of the table gives estimates of the share of establishments that prefer men over women from the list randomization. Here the 541 establishments that were asked directly were given a list of 4 non-key statements without the statement about gender discrimination, and the remaining 535 establishments were given the same list with the key statement about gender discrimination added to it. To estimate the share that discriminate based on the list randomization questions, we regress the number of statements the establishment agreed with on a dummy variable for being in the group that had a fifth statement about gender discrimination.

The estimate from the list randomization shows that 41% of establishments prefer men over women, but the confidence interval is large (from 21% to 61%). This overlaps with the point estimate from the direct question, meaning that the two methods are giving a statistically similar finding. When we explicitly test of equality of the two measures, we get a p-value of 0.31. If we take the upper end of the confidence interval for the list randomization method (61%), this gives us an upper bound on both the estimated level of discrimination and on how stigmatized discrimination against women is. Since the direct question gives an estimate of 51%, even if the true share of discriminators is the upper bound of 61%, this would mean that about 80% of actual discriminating establishments are admitting it openly. In other words, we can rule out that discrimination against women is heavily stigmatized in Egypt. This is perhaps unsurprising given some of the statistics and rankings for Egypt from the United Nations that we shared earlier, but it is striking to hear it from those in a position of hiring employees.[14]

There are two other noteworthy aspects of our list randomization findings. One is that the point estimate for list randomization is actually lower than from the direct question. In theory, this should not happen; if the behavior is not stigmatized, then the list randomization should give an estimate similar to the direct question. However,

[14] Explicit preferences for one sex in hiring for certain positions are not terribly surprising. In studies using online job ads in Mexico and China, Helleseter et al. (2020) find high rates of gender preferences among employers using online jobs ads in Mexico and China, with their estimates ranging from 11 to 72 percent of job ads listing some preference for males or females. Many of these prefer females for certain jobs, though, so these are just not measures of discrimination against one sex.

Table 1: Discrimination by Gender

Panel A: Percent of establishments that prefer men over women	Overall	Retail	IT	Hotels	Restaurants	
Direct question	51.4%	60.1%	16.4%	43.7%	55.5%	
	[45.8, 56.9]	[52.1, 68.0]	[5.7, 27.2]	[20.4, 66.9]	[47.1, 63.9]	
Observations	541	218	61	78	184	
List randomization	41.0%	57.4%	25.9%	40.2%	27.0%	
	[21.3, 60.6]	[30.5, 84.4]	[-11.7, 63.5]	[-31.3, 111.7]	[-6.4, 60.4]	
Observations	1,076	442	123	164	347	
p-value for Direct vs. List	0.31	0.85	0.63	0.93	0.09	

Panel B: Predictors of admitting a preference for men over women	(1)	(2)	(3)	(4)	(5)	(6)
Part of larger firm?	0.062					0.095
	(0.052)					(0.060)
Number of employees		-0.004*				-0.006***
		(0.002)				(0.002)
Pays above competitors			-0.100			-0.170**
			(0.089)			(0.086)
Pays below competitors			0.084			0.065
			(0.141)			(0.132)
% workers w/institute/college ed				0.106		0.129*
				(0.072)		(0.066)
% workers formal				0.001		0.001
				(0.001)		(0.001)
% workers female					-0.768***	-0.983***
					(0.162)	(0.151)
Industry fixed effects?	Y	Y	Y	Y	Y	Y
Observations	541	541	541	535	541	514
R-squared	0.077	0.076	0.080	0.085	0.149	0.211

Panel A note: The 95% confidence interval is given in brackets below each figure. The direct question is the percentage of establishments answering yes to the question, "When hiring a new employee, do you prefer men over women?" The list randomization figure is the difference in mean number of true statements reported by the treatment group and the control group. Robust standard errors are used. Panel B note: Each column reports results from an OLS regression of admitting a preference for men over women. Robust standard errors are shown in parentheses. All regressions include industry fixed effects and a constant term. All estimates are weighted by establishment size to match the size distribution in the Egyptian census of businesses. We exclude all establishments with fewer than 5 employees due to a lack of data about them in the Egyptian census of establishments. The pays above/pays below variables are missing for all of the retail sector, we replace them with a constant which is made irrelevant by the industry fixed effects.*p<0.10 **p<0.05 ***p<0.01.

Tsuchiya et al. (2007) get a similar result to ours when they use list randomization to study donating blood, which is presumably not stigmatized. They show that this method can give lower estimates than the direct question particularly when there are several other non-key statements. Respondents may simply undercount as they try to add up the statements that are true, especially when the list of non-key statements is longer. This likely helps explain why the estimate for restaurants is marginally statistically different from the direct measure.

It is also important to discuss the trade-off between precision and bias. Our list randomization results are useful in that they allow us to rule out the possibility that gender discrimination is heavily stigmatized. However, the large standard errors limit our ability to precisely estimate the incidence of such discrimination. Because the list randomization method is based on the difference across two variables, the variance of this difference will be mechanically greater than the variance of a direct question (which is only based on one variable).[15] This is true even in cases, such as ours, where the direct question is only posed to half the sample. Thus, while the benefit of list randomization is that it can decrease bias in responses, it will mechanically have a higher variance. The implication is that list randomization will be most useful in situations in which the researcher's primary consideration is bias due to an activity being stigmatized, and in which the researcher is willing to sacrifice some level of precision.

Tsuchiya et al. (2007) provide some guidance on strategies that researchers can take to try limit the increase in variance from using a list randomization strategy. First is to greatly increase the sample size of the list randomization relative to the direct response question. A second strategy relates to the list of non-key items. Tsuchiya et al. (2007) show that a longer list of non-key items leads to a larger standard errors. There is an additional trade-off here. A short list of non-key items makes it more difficult to "hide" the stigmatized behavior for the respondent, which may discourage honest answers. However, a longer list creates more variance in the answers to the non-key items and thus makes the standard errors larger.

A further issue with list randomization methods (including our implementation) is highlighted by Chuang et al. (2021). The assumption that makes the method valid is that respondents will answer the question of "how many items are true" from the list honestly, even if they do not answer the direct question honestly. Chuang et al. (2021) note that this may not be true if respondents either are confused by the method or fear that their true responses to the sensitive statement can somehow be backed out

[15] The standard error on the "direct question" estimate of the proportion of employer's discriminating against women is the square root of $p(1-p)/N$, where p is the proportion answering yes. The standard error on the list randomization however would be the square root of $(2k+1)p(1-p)/(N)$, if we held the sample size constant and assumed the proportion of people answering yes to the k non-key questions was also p and the answers were independent. This is why the standard error from the list randomization is mechanically larger than the direct question.

by the researcher. They provide a test in which the sensitive statement is included in two separate lists, with different non-key statements, and each is given randomly to opposite subsets of the sample. This allows the researcher to calculate the "truth" from list randomization twice (once using each list of items), and those two estimates can be compared to see if they are the same. We did not include this test in our study, but we would advise other researchers to do so.

Taking all of this into account, list randomization is still useful in our case for ruling out large degrees of stigmatization. Discrimination against women is widespread in the Egyptian labor market. Estimates from both methods are consistent with about half of establishments preferring to hire men over women. The establishments that discriminate against women are clearly not shy about it. The lack of stigmatization allows us to proceed to an analysis that examines what employer characteristics and labor-market outcomes are associated with discrimination.

Correlates of Discrimination

For the 541 establishments that were directly asked about gender preference, we can look at what characteristics are correlated with the admission of a hiring preference for men over women. In Panel B of Table 1, we report results from OLS regressions of admitting discrimination on several establishment characteristics. All regressions also include industry fixed effects and a constant term. It is difficult to predict discrimination using most of these variables. The only variable that is strongly related to discrimination is the outcome one would expect to result from discrimination: the share of workers in the establishment that are female. When entered separately, this variable is the most significant of all the variables and has by far the largest R^2. In the combined regression, larger establishments and those that pay above-market wages are less likely to discriminate, but formality and being part of a larger firm do not predict this phenomenon. With all variables and industry fixed effects included, the R^2 is 0.211; without industry fixed effects, this falls to 0.170, and without the female share or industry effects, it is only 0.058. In other words, other than industry and the outcome we would expect to be related to discrimination, these variables have little explanatory power.

While our focus here is on gender, we also used the list randomization method to ask about discrimination by socioeconomic status (SES). We present these results in the Appendix. We find that while only one-third of establishments admit to putting more weight on SES than on workers' skills, the list randomization gives an estimate of 54%, and the two are significantly different. This suggests that discrimination by SES is stigmatized to some degree, while discrimination against women - which establishments admitted freely - is likely not. This contrast between SES and gender also shows that we have sufficient power to detect differences between the direct question and the list randomization in a case where the behavior in question is stigmatized. However, the confidence intervals for the list randomization are again quite large, given that we have the same sample size and number of non-key items as in the gender study. This limits our ability to say much about the "true" share of discriminators by SES.

Outcomes Associated with Gender Discrimination

Because Egyptian establishments openly admit to discriminatory attitudes against women, we are able to look at the employment characteristics associated with discrimination at the micro level. Specifically, we explore the correlation of discrimination against women with hiring patterns and difficulties on the part of establishments. As discussed above, if the discrimination is statistical rather than taste-based - that is, based on accurate and rational expectations of gender-based productivity differentials in the presence of imperfect information - then we would expect to see some type of positive correlation with an advantageous labor-market outcome. (Bohren et al., 2019; Guryan and Charles, 2013). Taste-based discrimination is theoretically consistent with multiple different predicted cross-sectional correlations, but a significant correlation with disadvantageous labor-market outcomes is primarily consistent with taste-based rather than statistical discrimination. Alternatively, if statistical discrimination is at play despite the disadvantageous correlation, then the implication would be that the discrimination is concentrated among low-productivity employers who receive limited benefits from discrimination.

This evidence is valuable even without the ability to establish causality. There is very little data on whether discriminatory attitudes are associated with positive or negative economic characteristics at the employer level, and our results can at least tell us which types of discrimination (along with their associated characteristics) are more likely.

In our empirical work, we analyze three basic hiring outcomes: average time to make a hire, average time it takes a new hire to reach "acceptable productivity", and employee terminations (involuntary turnover). The first is a measure of how easily the establishment can find workers. The second and third are related to the quality of those new hires.[16] Time to hire and time to productivity are measured in weeks. We model all of these items as count data using negative binomial models, estimating versions of the following base equation for establishment i in industry j:

$$Outcome_i = \alpha_i + GenderDisc_i\beta_1 + SkillReq_i\beta_2 + X_i\beta_3 + Industry_j + \epsilon_i \qquad (1)$$

All regressions are weighted based on establishment industry-size frequencies in the Egyptian enterprise census. We use Huber-White standard errors. The vector X_i contains various organizational characteristics, including detailed skill requirements that allow us to control for job complexity *within* occupation. In our full specifications, we include interactions between gender discrimination and three organizational attributes: (1) an indicator for multi-site operations and formal labor practices (vs. single-site or informal), (2) high and low levels of educational attainment among the establishment's current workforce, and (3) an indicator for above-median establishment-level employment. These interactions are important as the hiring frictions and productivity attributes that are associated with gender bias likely show heterogeneity across employer types. Multi-site formal organizations tend to have more resources and higher job quality. Consequently, they may incur lower market penalties for biased behavior as they are able to hire the "cream" of even a more limited pool of male-only workers.

When estimating effects associated with formality/job-quality, it is important to also control for establishment-level employment size. Although formality and employment size are positively correlated, they are very different phenomena. Indeed, more than a quarter of our sample consists of informal or single-site employers who neverthe-

[16]See Osterman and Weaver (2013) and Osman et al. (2022) for more on using these measures.

less have above-median levels of employment at the establishment level. Finally, our education interactions test whether discriminating employers with higher skill requirements show a different pattern of labor-market outcomes (Kuhn and Shen, 2013). We also include industry fixed effects as our target industries vary widely in both hiring practices and gender representation.

The results are shown in Table 2.[17] All coefficients are marginal effects. In the first specification in column (1), which measures the relationship between gender discrimination and time-to-hire, we can see that in a simple model that only controls for employment size and growth, discriminatory attitudes toward female job candidates are associated with an insignificant half-week reduction in time-to-hire.[18] A full model with controls for skill requirements, organizational characteristics, and industry effects also shows no significant correlation. However, the interaction effects do reveal a correlated negative outcome: employers who exhibit gender bias and whose workforce is above the 75th percentile of standardized institute/college-level education experience a 1.7 week longer recruitment time relative to biased employers with less-educated workers ($p<0.05$).[19] Other than employers seeking very highly educated workers, the results thus generally indicate that Egyptian employers with discriminatory attitudes toward women do not show signs of an obvious penalty in terms of extended hiring times.[20]

The termination results (columns 3 and 4) show somewhat more nuanced results.[21] Although a simple model shows a significant correlation between gender discrimination and greater terminations (thus theoretically implying either a match-quality penalty

[17]We omit reporting some of the covariates from columns 2, 4, and 6, but the full results are found in Appendix Table A3.

[18]Our time to hire measure is based on the time it takes for the employer to make an offer, and does not take into account how long it takes for the employee to accept and start the job, since this is what the employer can best control. The exact wording of the question is: "Based on your recent experience (i.e., past year or two), from when you start searching, how long does it typically take to identify a candidate for a [employee type] job to whom you make an offer?"

[19]Results are similar when comparing other groups by education, such as the 80th vs. 20th or 90th vs. 10th percentiles.

[20]About 22% of the sample did not respond to the question regarding the number of weeks it takes to hire. In Appendix Table A4 we regress a binary for non-response to this question on a wide ranging list of variables. We find that non-response is associated with establishments who pay a bit better, have fewer employees, are less likely to provide training and are less likely to discriminate. Importantly there is no difference in non-response based on the establishment's educational needs, which is the primary margin for which we see differences in our analysis of this metric. This result makes us less concerned that differential attrition is driving our results on time to hire.

[21]The exact question asked was "How many [employee type] were fired in the past twelve months?".

Table 2: Gender Discrimination and Hiring Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	**Wks to Hire**		**Terminations**		**Wks to Productivity**	
Bias against female candidates	-0.482	-0.187	0.446**	0.172	0.587	1.312**
	(0.513)	(0.458)	(0.212)	(0.222)	(0.528)	(0.553)
Above-median estab. employment	-0.207	-0.104	-0.247	-0.275	1.755***	1.608**
	(0.424)	(0.383)	(0.163)	(0.185)	(0.555)	(0.638)
Formal, part of larger org.		1.077**		0.167		0.890
		(0.509)		(0.226)		(0.699)
Institute/college ed. (std. proportion)		0.217		-0.094		0.326
		(0.297)		(0.095)		(0.275)
Unemployment rate		-0.217**		-0.057		-0.110
		(0.090)		(0.047)		(0.105)
Marginal interaction effects of employer gender discrimination:						
Formal, part of larger org.		-0.993		-0.398		-0.001
		(0.794)		(0.410)		(1.149)
Other type		0.071		0.477**		1.744***
		(0.521)		(0.236)		(0.618)
Difference in marginal effects		-1.064		-0.875*		-1.745
		(0.896)		(0.458)		(1.270)
Institute/college ed.–75th pctile		0.400		0.076		1.195**
		(0.561)		(0.224)		(0.609)
Institute/college ed.–25th pctile		-1.311**		0.393		1.513**
		(0.521)		(0.285)		(0.770)
Difference in marginal effects		1.711***		-0.317		-0.318
		(0.645)		(0.263)		(0.798)
Above-median estab. emp.: Yes		-0.830		0.116		3.716***
		(0.506)		(0.251)		(1.119)
Above-median estab. emp.: No		0.091		0.171		0.404
		(0.574)		(0.288)		(0.602)
Difference in marginal effects		-0.921		-0.056		3.313***
		(0.675)		(0.359)		(1.204)
Observations	443	423	537	514	541	516
Pseudo R-squared	0.001	0.033	0.012	0.085	0.005	0.027

Notes: Each column reports results from negative binomial regressions using establishment employment weights. The regressions in columns 2, 4, and 6 also include industry fixed effects; dummies for requiring advanced writing skills, English speaking, and computer usage; whether the establishment has higher technology and higher or lower wages than competitors; whether they use owner connections in hiring; whether they provide training; whether they have shifted to internal hiring in recent years; whether they experience frequent poaching of workers; and the unemployment rate in their governorate. All estimates are weighted by establishment size to match the size distribution in the Egyptian census of businesses. We exclude all establishments with fewer than 5 employees due to a lack of data about them in the Egyptian census of establishments. The pays above/pays below variables are missing for all of the retail sector, we replace them with a constant which is made irrelevant by the industry fixed effects. The full results are available in Table A3. Robust standard errors in parentheses. *p<0.10 **p<0.05 ***p<0.01

for discrimination or a non-causal correlation between discrimination and pre-existing employment volatility), this effect becomes smaller and insignificant in the presence of organizational controls and industry effects. However, the insignificant marginal effect of discrimination in the full model masks heterogeneous effects. Informal or single-site establishments (hereafter single/informal) that discriminate experience a significant increase in annual terminations equal to an additional half of an employee (0.477) per year. By contrast, formal establishments that are part of a multi-site organization (hereafter multi/formal) and that discriminate by gender experience insignificantly fewer terminations. The difference between these two marginal effects is significant at the 90% level and equates to about one additional termination per year for single/informal establishments. One interpretation of this result is that multi/formal establishments may be able to indulge a taste for discrimination by relying on their superior resources, greater male application rates, and access to better hiring networks (Hellerstein et al., 2002; Osman et al., 2022). Even from a non-causal perspective, it is worth noting that the labor-market outcomes that are associated with gender discrimination are not uniform, but rather vary by employer type.

The most pronounced direct effects associated with gender discrimination involve worker quality or productivity, as measured by time to acceptable productivity for new hires.[22] In the full specification (column 6), establishments that admit discriminatory attitudes toward women experience an average of 1.3-week delay in time-to-productivity (a 24% increase over non-discriminators in our estimation sample), significant at the 5% level. As with terminations, the interaction results show that this marginal effect is most concentrated among single/informal establishments. The marginal effect for single/informal employers is large and significant (1.7-week delay, a 36% increase over the mean for non-discriminators in that group), although the difference in marginal effects between formal and informal employers is not precisely estimated. It is important to emphasize that this result is net of skill and education controls and that single/informal employers have lower skill and hiring requirements (Osman et al., 2022). Thus this result runs counter to the obvious unmeasured skill bias.

The interaction effects also indicate, perhaps surprisingly, that the productivity

[22] The exact question asked was "After you hire a typical [employee type], how many weeks or months on average does it take a typical person to achieve an acceptable level of productivity?"

penalty for gender discrimination is more pronounced among establishments with above-median employment. Larger establishments discriminate less frequently (see Table 1), but the ones that do appear to have signs of lower productivity. The above-median-size establishments that discriminate face a nearly 4-week delay–a 62% increase over non-discriminators in that group–compared with no delay for the smaller establishments, and the difference is highly significant. It is also worth noting that establishments at both high and low points on the education distribution experience similar gender-bias-related productivity penalties, thus revealing a nonlinear pattern that is worthy of further investigation with other datasets. Based on our results, it appears that the differential frictions associated with seeking highly educated workers but excluding female candidates are most clearly correlated with worker quality rather than longer times to hire.

These effects reveal a number of interesting patterns. First, formality and employer size are very different phenomena, and they have different relationships with gender discrimination. While multi-site/formal employers tend to have larger establishment-level employment, these results imply that it may be the formality and its associated job quality that allow discriminating employers to hire high-productivity male workers and avoid a diminution of average productivity. Net of formality, larger employers are not able to use attributes of size (e.g., potential better matching due to economies of scale in hiring) to compensate for the negative productivity effects of bias against a class of (female) candidates. Indeed, larger establishments incur one of the largest penalties in our estimations. It could be that the business model of larger establishments, which are significantly more likely to show positive employment growth than smaller establishments, depends more heavily on throughput or uninterrupted growth (and thus makes these employers more willing to compromise on productivity). The results also show that the educational profile of an establishment's workforce is relevant as well, although the potential penalty shows up in recruitment time. While none of these results are necessarily causal, we stress that they are not obvious. The fact patterns uncovered here can guide future causal research designs regarding gender discrimination.

Overall, our results indicate that discrimination against women in hiring is associated with a negative profile of employment characteristics for some establishments.

Because our empirical work does not turn up any significant correlations with advantageous labor-market outcomes, the results are more consistent with either taste-based or heavily selected and limited-benefit statistical discrimination. Across all employers, the primary correlated negative outcome is lower worker quality rather than increased hiring times.[23] Discriminating single/informal employers show signs of lower match quality in the form of greater terminations, while establishments that employ highly educated workers do experience hiring delays. The negative correlations with disadvantageous outcomes are greatly attenuated for multi-site/formal employers. These effects are net of detailed controls for skill demands, including writing, English language requirements, and level of computer usage (standardized within each industry). Thus the effects that we estimate are not simply reflecting the fact that male jobs tend to involve higher skill requirements. The results are also robust to the inclusion of the percentage of workers who are female at each establishment, thus indicating that the productivity variable is not simply picking up frictions associated with integrating women and men in the workplace.

As a caveat, we note that due to the large standard errors on our list randomization results Table 1, we cannot rule out that a small share of employers are denying discriminatory attitudes when they in fact hold these biases. If so, this would mean that our results on the outcomes associated of discrimination are attenuated, since they rely on the assumption that businesses are answering the discrimination question honestly.

4 Discussion

Using a detailed survey of Egyptian establishments across four industries, we have shown compelling evidence of widespread discrimination against women in hiring. Roughly half of establishments directly admit to preferring to hire men over women. List randomization results rule out the possibility that gender discrimination is heavily stigmatized in Egypt.

Because of establishments' seeming willingness to admit to discriminatory attitudes,

[23] Another possibility, which we do not have the data to explore, is that the discrimination is taste-based, but based on the preferences of customers rather than employers. In this case, the discriminating employer would still be hiring lower-productivity workers, as we find, but they would not suffer lower profits.

we can look at the outcomes associated with discrimination at the micro level. Gender discrimination is associated with lower productivity for establishments. Although we cannot establish causality between discrimination and labor-market outcomes, this result is more consistent with employer-taste-based discrimination than statistical discrimination (even if all relationships are non-causal). Alternatively, if statistical discrimination is a relevant factor, the results imply that the practice is concentrated among low-productivity employers and yields little productivity benefit. While most discriminators do not take longer to find workers, our regression results suggest that they do end up with lower-quality workers than non-discriminators; their workers take 24% longer to reach acceptable productivity, even after controlling for detailed skill requirements and a large set of other factors.

These negative outcomes are concentrated among the informal, single-site establishments, which are common in developing countries. Those that discriminate among this group see a 36% rise in their workers' time to reach acceptable productivity. These enterprises are difficult to study by other means. Alternative methods, such as correspondence studies, typically rely on online job advertisements, a method rarely used by these employers. Our survey not only allows us to reach these establishments, but also to study their behavior and associated labor-market outcomes.

Overall, our results provide some of the clearest evidence yet of widespread discriminatory attitudes and practices in the labor market of a developing country. While gender bias is associated with negative employment dynamics, these adverse correlations are not evenly distributed. Our results indicate that market forces are unlikely to pressure formal, multi-site discriminators into eliminating the discriminatory practices, as these employers appear to be shielded from productivity-related penalties. Single-site/informal employers do show signs consistent with penalties, but they appear willing to make this trade-off. Alternatively, if the relationships are non-causal, then no adverse market signal will undo pre-existing labor-market patterns. To the extent that gender discrimination in this context is taste-based, public policies and labor-market interventions designed to lessen gender bias will likely need to focus on changing employer behavior rather than simply attempting to achieve gender parity via increased investments in female human capital. To the extent that some level of statistical discrimination is present, it is concentrated among low-productivity employers and yields little productivity boost. Any education or training efforts to address such a dynamic would best be targeted at the lower end of the labor market.

References

Altonji, J. G. and R. M. Blank (1999): "Race and Gender in the Labor Market," in *Handbook of Labor Economics, Volume 3*, ed. by O. Ashenfelter and D. Card, Elsevier Science.

Arceo-Gomez, E. O. and R. M. Campos-Vazquez (2014): "Race and Marriage in the Labor Market: A Discrimination Correspondence study in a Developing Country," *American Economic Review*, 104, 376–80.

Arrow, K. J. (1973): "The Theory of Discrimination," in *Discrimination in Labor Markets*, ed. by O. Ashenfelter and A. Rees, Princeton University Press, 3–34.

Azmat, G. and A. Boring (2020): "Gender diversity in firms," *Oxford Review of Economic Policy*, 36, 760–782.

Baert, S. (2018): "Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005," *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, 63–77.

Baliamoune-Lutz, M. and M. McGillivray (2007): "Gender inequality and growth in Sub-Saharan Africa and Arab countries," *International Centre for Economic Research Working Papers*, 25, 1–27.

Banerjee, A., M. Bertrand, S. Datta, and S. Mullainathan (2009): "Labor market discrimination in Delhi: Evidence from a field experiment," *Journal of comparative Economics*, 37, 14–27.

Becker, G. (1957): *The Economics of Discrimination*, University of Chicago Press.

Bertrand, M. and S. Mullainathan (2004): "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 94, 991–1013.

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

Blau, F. D. and L. M. Kahn (2017): "The gender wage gap: Extent, trends, and explanations," *Journal of economic literature*, 55, 789–865.

Blinder, A. S. (1973): "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 436–455.

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2019): "Inaccurate statistical discrimination: An identification problem," Tech. rep., National Bureau of Economic Research.

Chuang, E., P. Dupas, E. Huillery, and J. Seban (2021): "Sex, lies, and measurement: Consistency tests for indirect response survey methods," *Journal of Development Economics*, 148, 102582.

Deming, D. J., N. Yuchtman, A. Abulafi, C. Goldin, and L. F. Katz (2016): "The Value of Postsecondary Credentials in the Labor Market: An Experimental Study," *American Economic Review*, 106, 778–806.

Demirgüç-Kunt, A., L. F. Klapper, and D. Singer (2013): "Financial Inclusion and Legal Discrimination Against Women: Evidence from Developing Countries," World Bank Policy Research Working Paper 6416.

Duflo, E. (2012): "Women empowerment and economic development," *Journal of Economic literature*, 50, 1051–79.

El Feki, S., H. B. and G. Barker (2017): "Understanding Masculinities: Results from the International Men and Gender Equality Study in the Middle East and North Africa," Tech. rep., UN Women and Promundo-US.

Esteve-Volart, B. (2000): "Sex Discrimination and Growth," IMF Working Paper No. 00/84.

Goldin, C. (2014): "A grand gender convergence: Its last chapter," *American Economic Review*, 104, 1091–1119.

Guryan, J. and K. K. Charles (2013): "Taste-based or statistical discrimination: the economics of discrimination returns to its roots," *The Economic Journal*, 123, F417–F432.

Hellerstein, J. K., D. Neumark, and K. R. Troske (2002): "Market Forces and Sex Discrimination," *Journal of Human Resources*, 37, 353–380.

Helleseter, M. D., P. Kuhn, and K. Shen (2020): "The Age Twist in Employers' Gender Requests Evidence from Four Job Boards," *Journal of Human Resources*, 55, 428–469.

Holbrook, A. L. and J. A. Krosnick (2010): "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique," *Public Opinion Quarterly*, 74, 37–67.

ILO (2016): "Employent by Sex and Age – ILO Modelled Estimates," ILOSTAT. Accessed 02-14-2020, https://ilostat.ilo.org/data.

Jayachandran, S. (2015): "The Roots of Gender Inequality in Developing Countries," *Annual Review of Economics*, 7, 63–88.

Kabeer, N. and L. Natali (2013): "Gender equality and economic growth: Is there a win-win?" *IDS Working Papers*, 2013, 1–58.

Karlan, D. S. and J. Zinman (2012): "List Randomization for Sensitive Behavior: An Application for Measuring Use of Loan Proceeds," *Journal of Development Economics*, 98, 71–75.

Kawaguchi, D. (2007): "A market test for sex discrimination: Evidence from Japanese firm-level panel data," *International Journal of Industrial Organization*, 25, 441–460.

Kitagawa, E. M. (1955): "Components of a Difference Between Two Rates," *Journal of the American Statistical Association*, 50, 1168–1194.

Klasen, S. and F. Lamanna (2009): "The impact of gender inequality in education and employment on economic growth: new evidence for a panel of countries," *Feminist economics*, 15, 91–132.

Kline, P. M., E. K. Rose, and C. R. Walters (2021): "Systemic Discrimination Among Large US Employers," NBER Working Paper 29053.

Knight, J. B. and R. H. Sabot (1982): "Labor Market Discrimination in a Poor Urban Economy," *The Journal of Development Studies*, 19, 67–87.

Kroft, K., F. Lange, and M. J. Notowidigdo (2013): "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 128, 1123–1167.

Kuhn, P. and K. Shen (2013): "Gender discrimination in job ads: Evidence from china," *The Quarterly Journal of Economics*, 128, 287–336.

LaBrie, J. W. and M. Earleywine (2000): "Sexual Risk Behaviors and Alcohol: Higher Base Rates RevealedUsing the Unmatched-Count Technique," *Journal of Sex Research*, 37, 321–326.

Liu, J.-T., M.-W. Tsou, and P. Wang (2010): "Workforce composition and firm productivity: evidence from Taiwan," *Economic Inquiry*, 48, 1032–1047.

Miller, J. (1984): *A New Survey Technique for Studying Deviant Behavior*, George Washington University.

Mowafi, M., Z. Khadr, I. Kawachi, S. Subramanian, A. Hill, and G. G. Bennett (2014): "Socioeconomic Status and Obesity in Cairo, Egypt: A Heavy Burden for All," *Journal of Epidemiology and Global Health*, 4, 13–21.

Neumark, D. (2018): "Experimental research on labor market discrimination," *Journal of Economic Literature*, 56, 799–866.

Oaxaca, R. (1973): "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 693–709.

Osman, A., J. D. Speer, and A. Weaver (2022): "Connections, Referrals, and Hiring Outcomes: Evidence from an Egyptian Establishment Survey," *Journal of Economic Behavior & Organization*, 204, 342–355.

Osterman, P. and A. Weaver (2013): "Skills and Skill Gaps in Manufacturing," in *Production in the Innovation Economy*, ed. by R. M. Locke and R. L. Wellhausen, MIT Press.

Pager, D. and D. S. Pedulla (2015): "Race, self-selection, and the job search process," *American Journal of Sociology*, 120, 1005–1054.

Phelps, E. S. (1972): "The statistical theory of racism and sexism," *The american economic review*, 62, 659–661.

Raghavarao, D. and W. T. Federer (1979): "Block total response as an alternative to the randomized response method in surveys," *Journal of the Royal Statistical Society: Series B (Methodological)*, 41, 40–45.

Siddique, Z. (2011): "Evidence on Caste Based Discrimination," *Labour Economics*, 18, S146–S159.

Tsou, M.-W. and C.-H. Yang (2019): "Does gender structure affect firm productivity? Evidence from China," *China Economic Review*, 55, 19–36.

Tsuchiya, T., Y. Hirai, and S. Ono (2007): "A Study of the Properties of the Item Count Technique," *Public Opinion Quarterly*, 71, 253–272.

Weber, A. and C. Zulehner (2014): "Competition and gender prejudice: Are discriminatory employers doomed to fail?" *Journal of the European Economic Association*, 12, 492–521.

Appendix: Discrimination by Socioeconomic Background

While the focus of our paper is discrimination by gender, we were also told told by our Egyptian partners that discrimination by socioeconomic status (SES) or family background is important. Therefore, we also included questions about this discrimination both because of its interest to our partners and to test and validate our list randomization method. SES is difficult to define in Egypt, as there are no major ethnic or linguistic minorities that would automatically be seen as of lower background. Anecdotally, it could be that language accents and being from a rural area are part of how SES is seen by employers. Other studies that look at SES in Egypt define SES measures like education and parental education (Mowafi et al., 2014).

We perform the same analysis for SES as we did for gender, comparing the "direct question" and list randomization methods. Table A1 shows the results. Here the question and list statement are, "When hiring a new employee, do you put more weight on their socioeconomic background than on their technical skills?" About one-third of establishments asked this question admit to this behavior directly, with the share similar across industries.

Using the list randomization, however, the figure is 54%, economically and statistically significantly higher than implied by the direct question. Discrimination by SES is widespread, about as prevalent as discrimination by gender. The key difference is that it seems to be more stigmatized than preferring men over women, with the list method giving a significantly higher estimate than the direct question. The fact that we find this large difference for SES gives us confidence that our list randomization has enough power to detect true differences. This gives us increased confidence in our gender results.

Using the list randomization, we see a wide variance by industry that mirrors that by gender: half of retail establishments and most hotels and restaurants put more weight on SES, while very few IT establishments do. The standard errors are again large, but the pattern is striking. So while the direct question implied that discrimination was similar across industries, the list randomization method is more in line with the gender discrimination results, where the IT industry appears to have the lowest rates

Table A1: Discrimination by Socioeconomic Background

	Percent of establishments that put more weight on SES than skills				
	Overall	Retail	IT	Hotels	Restaurants
Direct question	32.3%	35.6%	33.4%	15.7%	33.5%
	[27.2, 37.5]	[28.0, 43.3]	[19.2, 47.7]	[-1.4, 32.8]	[25.4, 41.5]
n	525	229	58	69	169
List randomization	53.8%	49.1%	1.92%	75.2%	65.2%
	[36.9, 70.6]	[22.2, 75.9]	[-40.8, 44.6]	[9.7, 140.7]	[39.8, 90.5]
n	1,072	441	123	164	344
p-value for Direct vs. List	0.02	0.35	0.17	0.08	0.02

The 95% confidence interval is given in brackets below each figure. The direct question is the percentage of establishments answering yes to the question, "When hiring a new employee, do you put more weight on socioeconomic background than on their technical skills?" The list randomization figure is the difference in mean number of true statements reported by the treatment group and the control group. Robust standard errors are used. We exclude all establishments with fewer than 5 employees due to a lack of data about them in the Egyptian census of establishments.

of discrimination.

Note that, because establishments are not all being honest here, we are unable to convincingly analyze the determinants of being a discriminating establishment (as we did in Table 1, Panel B for gender), nor are we able to analyze the consequences of discrimination by SES (as we did in Table 2 for gender).

Appendix: Additional Tables

Table A2: Gender Discrimination and Hiring Outcomes (Uniform Sample)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	**Wks to Hire**		**Terminations**		**Wks to Productivity**	
Bias against female candidates	-0.461	-0.179	0.519*	0.143	0.310	1.377**
	(0.534)	(0.458)	(0.267)	(0.253)	(0.654)	(0.652)
Above-median estab. employment	-0.199	-0.109	-0.353*	-0.199	1.880***	1.700**
	(0.441)	(0.384)	(0.209)	(0.191)	(0.674)	(0.735)
Formal, part of larger org.		1.081**		0.259		0.880
		(0.511)		(0.267)		(0.713)
Institute/college ed. (std. proportion)		0.210		-0.100		0.500*
		(0.298)		(0.117)		(0.302)
Unemployment rate		-0.217**		-0.068		-0.177
		(0.090)		(0.071)		(0.146)
Marginal interaction effects of employer gender discrimination:						
Formal, part of larger org.		-0.993		-0.494		0.379
		(0.796)		(0.472)		(1.170)
Other type		0.083		0.448*		1.768**
		(0.521)		(0.263)		(0.740)
Difference in marginal effects		-1.075		-0.942*		-1.389
		(0.900)		(0.489)		(1.298)
Institute/college ed.–75th pctile		0.403		0.115		1.541**
		(0.561)		(0.259)		(0.738)
Institute/college ed.–25th pctile		-1.300**		0.184		0.983
		(0.523)		(0.354)		(0.810)
Difference in marginal effects		1.703***		-0.069		0.558
		(0.647)		(0.323)		(0.847)
Above-median estab. emp.: Yes		-0.837*		0.064		3.659***
		(0.507)		(0.294)		(1.272)
Above-median estab. emp.: No		0.106		0.163		0.347
		(0.576)		(0.327)		(0.692)
Difference in marginal effects		-0.943		-0.100		3.311***
		(0.678)		(0.411)		(1.369)
Observations	421	421	421	421	421	421
Pseudo R-squared	0.001	0.032	0.014	0.076	0.005	0.030

Notes: This table repeats the analysis in Table 2 of the main paper but holds the sample constant across regressions (dropping any observations with any missing values). Robust standard errors in parentheses. *p<0.10 **p<0.05 ***p<0.01

Table A3: Gender Discrimination and Hiring Outcomes (All Controls)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	**Wks to Hire**		**Terminations**		**Wks to Productivity**	
Bias against female candidates	-0.482	-0.187	0.446**	0.172	0.587	1.312**
	(0.513)	(0.458)	(0.212)	(0.222)	(0.528)	(0.553)
Above-median estab. employment	-0.207	-0.104	-0.247	-0.275	1.755***	1.608**
	(0.424)	(0.383)	(0.163)	(0.185)	(0.555)	(0.638)
Formal, part of larger org.		1.077**		0.167		0.890
		(0.509)		(0.226)		(0.699)
Institute/college ed. (std. proportion)		0.217		-0.094		0.326
		(0.297)		(0.095)		(0.275)
Unemployment rate		-0.217**		-0.057		-0.110
		(0.090)		(0.047)		(0.105)
Advanced writing		0.307		0.429		0.458
		(0.616)		(0.375)		(0.659)
Computer usage (standardized)		-0.124		0.359**		-0.361
		(0.229)		(0.145)		(0.262)
English speaking		0.118		0.053		0.862
		(0.513)		(0.218)		(0.586)
High level of technology		-0.450		-0.369*		0.026
		(0.478)		(0.209)		(0.641)
Provide formal/informal training		-0.106		0.131		0.234
		(0.529)		(0.195)		(0.596)
Shift to inside hiring		-0.429		0.082		-0.157
		(0.477)		(0.200)		(0.553)
IT industry		1.210		-1.129***		3.601***
		(0.985)		(0.326)		(1.265)
Hotel industry		1.485		-0.742*		0.803
		(1.161)		(0.410)		(1.043)
Restaurant industry		0.377		-0.729**		-0.499
		(0.637)		(0.290)		(0.553)
Wage above competitors		-1.277***		0.955**		-0.081
		(0.392)		(0.409)		(0.605)
Wage below competitors		-1.057*		-0.177		-1.807*
		(0.549)		(0.448)		(1.018)
Experience frequent poaching		-0.510		2.083*		-1.006
		(0.411)		(1.087)		(0.703)
Marginal interaction effects of employer gender discrimination:						
Formal, part of larger org.		-0.993		-0.398		-0.001
		(0.794)		(0.410)		(1.149)
Other type		0.071		0.477**		1.744***
		(0.521)		(0.236)		(0.618)
Difference in marginal effects		-1.064		-0.875*		-1.745
		(0.896)		(0.458)		(1.270)
Institute/college ed.–75th pctile		0.400		0.076		1.195**
		(0.561)		(0.224)		(0.609)
Institute/college ed.–25th pctile		-1.311**		0.393		1.513**
		(0.521)		(0.285)		(0.770)
Difference in marginal effects		1.711***		-0.317		-0.318
		(0.645)		(0.263)		(0.798)
Above-median estab. emp.: Yes		-0.830		0.116		3.716***
		(0.506)		(0.251)		(1.119)
Above-median estab. emp.: No		0.091		0.171		0.404
		(0.574)		(0.288)		(0.602)
Difference in marginal effects		-0.921		-0.056		3.313***
		(0.675)		(0.359)		(1.204)
Observations	443	423	537	514	541	516
Pseudo R-squared	0.001	0.033	0.012	0.085	0.005	0.027

Notes: Each column reports results from negative binomial regressions using establishment employment weights. Robust standard errors reported in parentheses. All regressions include industry fixed effects. All estimates are weighted by establishment size to match the size distribution in the Egyptian census of businesses. We exclude all establishments with fewer than 5 employees due to a lack of data about them in the Egyptian census of establishments. The pays above/pays below variables are missing for all of the retail sector, we replace them with a constant which is made irrelevant by the industry fixed effects. *p<0.10 **p<0.05 ***p<0.01.

Table A4: Non-response on Primary Outcomes

Outcome: Non-response = 1	(1) Wks to Hire	(2) Terminations	(3) Wks to Productivity
Formal, part of larger org.	-0.012	-0.008	-0.012**
	(0.033)	(0.007)	(0.005)
Education level	-0.018	-0.002	0.000
	(0.016)	(0.002)	(0.002)
Advanced writing required	-0.006	0.006	-0.005
	(0.040)	(0.014)	(0.008)
Computer usage required	-0.007	-0.001	-0.002
	(0.015)	(0.003)	(0.003)
English speaking required	-0.040	-0.005	0.003
	(0.035)	(0.009)	(0.005)
High level of technology	-0.003	0.003	0.012
	(0.034)	(0.010)	(0.008)
Provides training	-0.175***	0.014**	0.013**
	(0.030)	(0.007)	(0.006)
Wages above competitors	0.073**	0.012	0.005
	(0.037)	(0.013)	(0.008)
Wages below competitors	0.080	0.000	0.003
	(0.091)	(0.011)	(0.010)
Unemployment Rate	0.001	0.000	0.000
	(0.006)	(0.001)	(0.000)
Bias against female candidates	-0.066**	0.009	0.003
	(0.030)	(0.007)	(0.006)
Number of Employees	-0.093***	-0.003	-0.002
	(0.030)	(0.006)	(0.005)
Shift to internal hiring	0.044	0.004	0.007
	(0.030)	(0.007)	(0.005)
Experience frequent poaching	-0.059	0.019	0.024**
	(0.036)	(0.012)	(0.012)
N	541	541	541
R-squared	0.382	0.845	0.915

Notes: Each row reports the results of a regression of a binary variable that is equal to one if the establishment did not response to the question related to the outcome variable in the column, on the variables reported in each row. All regressions include industry fixed effects. Robust standard errors reported in parentheses. The pays above/pays below variables are missing for all of the retail sector, we replace them with a constant which is made irrelevant by the industry fixed effects.*p<0.10 **p<0.05 ***p<0.01.

Table A5: List Randomization Statements

Discrimination by gender	
"Control" group	"Treatment" group
I have 2 or more sisters in my family.	I have 2 or more sisters in my family.
Someone in my family has started a new business in the past year.	Someone in my family has started a new business in the past year.
I was born in this city.	I was born in this city.
My household owns a computer.	My household owns a computer.
	When hiring a new employee, I prefer to hire men over women.