Journal of Labor Research (2022) 43:65–102
https://doi.org/10.1007/s12122-022-09331-4

Gender stereotypes in job advertisements: What do they imply for the gender salary gap?

Eva O. Arceo-Gomez[1] · **Raymundo M. Campos-Vazquez**[2] · **Raquel Y. Badillo**[3] · **Sergio Lopez-Araiza**[4]

Accepted: 22 March 2022 / Published online: 2 April 2022

Abstract

Gender stereotypes, the assumptions concerning appropriate social roles for men and women, permeate the labor market. Analyzing information from over 2.5 million job advertisements on three different employment search websites in Mexico, exploiting approximately 235,00 that are explicitly gender-targeted, we find evidence that advertisements seeking "communal" characteristics, stereotypically associated with women, specify lower salaries than those seeking "agentic" characteristics, stereotypically associated with men. Given the use of gender-targeted advertisements in Mexico, we use a random forest algorithm to predict whether non-targeted ads are in fact directed toward men or women, based on the language they use. We find that the non-targeted ads for which we predict gender show larger salary gaps (8–35 percent) than explicitly gender-targeted ads (0–13 percent). If women are segregated into occupations deemed appropriate for their gender, this pay gap between jobs requiring communal versus agentic characteristics translates into a gender pay gap in the labor market.

Keywords Gender stereotypes · Salary gap · Discrimination · Big data · Machine learning · Mexico

JEL Classification C52 · C53 · E24 · J64 · O54

✉ Raymundo M. Campos-Vazquez
rmcampos@colmex.mx

1 Universidad Iberoamericana, Departamento de Economía, Prolongación Paseo de la Reforma 880, Lomas de Santa Fe, 01219 Mexico City, Mexico

2 Centro de Estudios Económicos, El Colegio de México, Camino al Ajusco 20, Col. Pedregal de Santa Teresa, 10740 Mexico City, Mexico

3 Banco de México, Cinco de Mayo 2, 06000 Mexico City, Mexico

4 Instituto Tecnológico Autónomo de México, Mexico City, Mexico

Introduction

Although the global gender salary gap has declined somewhat over time, it has proven to be very persistent. According to the *Global Gender Gap Report* (World Economic Forum 2019), the salary gap between men and women doing similar work around the world has stalled at approximately 65 percent. Olivetti and Petrongolo (2016) explain that increasing gender convergence in human capital coupled with more stringent anti-discrimination measures leaves gender norms as the remaining explanation for the gap. Salary decompositions in various countries show that this portion of the gap is now much more important than the part explained by gender differences in education and experience.[1] The economics literature has responded with increased interest in gender stereotypes and their corresponding norms and roles. The most recent literature on economic penalties to motherhood in various countries also sheds light on gender stereotyping as an important source of disparities in the labor market.[2]

This paper analyzes how gender stereotypes affect the salary gap through a different mechanism. We propose that the gender stereotyping of jobs in the labor market opens salary gaps, given that firms assign different values to abilities associated with men and women. These associations are, of course, rooted in stereotypical views of how people should behave based on their gender. Ellemers (2018) explains that "assertiveness and performance are seen as indicators of greater agency in men, and warmth and care for others are viewed as signs of greater communality in women" (p. 277). We provide evidence that suggests that, even when they do not specify the desired gender of the prospective employee, implicit discrimination in job ads helps to explain the wage gap. The stereotyped language used to describe jobs on employment search websites may induce women to apply for jobs described with stereotypically feminine (communal) characteristics and men to apply for jobs with stereotypically masculine (agentic) characteristics (as shown, for example, in Born and Taris 2010; Gaucher et al. 2011; Flory et al. 2015). If stereotypically feminine and masculine characteristics are not valued equally in the labor market, this self-selection produces a salary gap.

Employment search websites provide a particularly rich set of big data regarding job descriptions, and this data has increasingly been used to study labor markets. The growing use of online job searches has provided a better understanding of the job search process (Faberman and Kudlyak 2016) and the matching process between vacancies and job seekers (Banfi et al. 2019; Kuhn and Shen 2013a). Other studies have used employment websites to analyze how job seekers respond to vacancies

[1] For the U.S., see Blau and Kahn (2017); for Mexico, see Arceo-Gómez and Campos-Vázquez (2014); for several countries in Latin America, see Hoyos and Ñopo (2010); for a world-wide study, see Ñopo, Daza, and Ramos (2012).

[2] See, for instance, for Chile, Berniell et al. (2019); for Denmark, Kleven, Landais, and Søgaard (2019a, b); for Denmark, Sweden, Germany, Austria, the U.S. and the U.K., Kleven et al. (2019a, b); for Mexico, Aguilar-Gomez, Arceo-Gomez, and De la Cruz Toledo (2019); for a comparison of heterosexual and lesbian couples in Norway, Eckhoff Andresen and Nix (2019); for Sweden, Angelov, Johansson, and Lindahl (2016); for the U.S. and the U.K., Kuziemko et al. (2018).

from distressed firms (Brown and Matsa 2016) and how changes in the demand for skills relate to tightness in labor markets (Hershbein and Kahn 2018; Modestino et al. 2019). There are also studies on the hiring preferences of firms (Kuhn and Shen 2015) and the effects of workers' job location preferences on unemployment rates (Marinescu and Rathelot 2018).

The present study focuses on salaries and gender discrimination, which previous studies have examined in several ways. Although earlier work has analyzed the effect of their wording, and some researchers have examined salary gaps in advertisements, we find no studies investigating both the relationship between the stereotypical wording of the ad and the salaries related to the stereotypes. Our paper aims to fill this gap in the literature.

The strand of literature most related to this work studies explicit discrimination in job postings. Kuhn and Shen (2013b) analyze explicit discrimination in the Chinese labor market through the use of ads directed specifically at men or women. They find that ads directed at women also tend to specify young applicants and criteria for height and attractiveness, whereas ads directed at men have older age requirements and rarely any for physical appearance. They also find that jobs with higher skill levels tend to be gender neutral (which, they argue, implies a negative targeting of skills to women) and attribute this phenomenon to the tight labor market for high-skilled workers. They do not, however, find any evidence of a salary gap in gender-targeted ads. Chowdhury et al. (2018) perform a similar analysis in India, and find a negative targeting of skills to women. As they dig deeper into the data, they find high occupational gender segregation, with women favored for low-status, low-salary jobs, while ads targeted to men specify higher salaries within the same occupations. They thus provide evidence of a gender salary gap.

Studies have also focused on other aspects of the salaries offered in job advertisements. Brenčič (2012) finds that the specification of salaries is strategic: ads for skilled jobs in the U.K. and Slovenia are less likely to specify compensation, possibly because firms want to be more selective for these jobs. Other researchers have studied how salaries vary either with the job title or with the qualifications specified in the ad. Marinescu and Wolthoff (2019) find that the job title explains up to 90 percent of the variation in advertised salaries in the U.S., but that only 20 percent of advertisements specify salary. They also observe that job titles explain more than 80 percent of the variation in applicants' education and experience. Finally, Deming and Kahn (2018) study the relationship between demand for skills and salaries specified. They find that ten specific skills explain about 12 percent of the variance in salaries across firms in the U.S., and that cognitive and social skills account for 5 percent of that variation.

Other characteristics of job advertisements might help or harm a firm's attempt to attract workers. For instance, Leibbrandt and List (2018) analyze the impact of having an equal employment opportunity (EEO) statement in a job advertisement. Their results show that minorities are less willing to apply to jobs with such statements, which survey evidence suggests could be the result of an unwillingness by jobseekers to be token hires. In a similar experiment, Flory et al. (2015) randomly vary the compensation scheme in job advertisements in order to gauge the effect of specifying a competitive compensation scheme on an advertisement's ability to attract female applicants. Corroborating the results of previous studies,[3] they find that women are less willing to apply for jobs with individual relative performance compensation schemes. To distinguish the effect of a competitive-compensation job from one stereotypically associated with men, they include jobs stereotypically associated with women, with variation in the wording of the ad to appeal to men. These jobs attract more women, as expected, and when language is added to appeal to men, the applicant pool almost balances.

We focus on Mexico, a developing country with more traditional gender norms than those in developed countries. Mexico's female labor force participation rate is 46 percent, lower than that of other countries in Latin America. Economic penalties for motherhood are quite large (Aguilar-Gomez et al. 2019), and women still carry over 70 percent of the burden of household work.[4] The gender salary gap is not very large, however (about 14 percent in 2018).[5] It is mostly unexplained across income percentiles, but particularly at the top and the bottom of the salary distribution (Arceo-Gómez and Campos-Vázquez 2014). Our working hypothesis is that gender stereotypes not directly associated with childbearing may help to explain part of the remaining gender salary gap.

Recent studies have begun to exploit online job data to explore gender discrimination in Mexico. Using advertisements from CompuTrabajo.com, Delgado Helleseter et al. (2018) find that even though gender-specific advertisements are evenly balanced between men and women, there is an age twist: ads targeted to women specify young women, and those targeted to men specify middle-aged men. There is also negative targeting of skill in Mexico, as Kuhn and Shen (2013b) found in China. Arceo-Gomez and Campos-Vazquez (2019a) combine analysis of explicit discrimination with a correspondence study to examine whether explicitly discriminatory ads are more biased with regard to marital status or race, as measured by callback rates. They find evidence of double discrimination against married women in ads targeted at women, and some discrimination towards dark-skinned women in explicitly discriminatory ads. However, there are no studies of the salaries specified, of gender salary gaps, or of stereotypes in the wording of job advertisements and their relationship with salary.

There is ample evidence that gender stereotypes are also reflected in the labor market, particularly in the types of jobs associated either with men or women. The main contribution of the current study is to relate stereotyped language in job advertisements to the gender gap in posted salaries. This stereotyped language leads to gender segregation in the labor market, since women and men both tend to apply to jobs stereotypically associated with their sex. Evidence for gender salary gaps in advertisements could thus translate into gender salary gaps in the labor market.

Our study tests whether advertisements whose language expresses stereotypes about women are associated with lower salaries than those expressing stereotypes

[3] See, for instance, Gneezy, Leonard, and List (2009), Niederle and Vesterlund (2007), and Niederle and Vesterlund (2011) for reviews of the literature on gender and competition.

[4] ADOECD Time Use Database. Retrieved on March 5, 2020 from https://stats.oecd.org/index.aspx?queryid=54757.

[5] OECD Gender Data. Retrieved on March 5, 2020 from http://www.oecd.org/gender/data/employment/.

about men. To do this, we first exploit the use of gender-targeted employment advertisements in Mexico. Using over 2.5 million job ads from three different employment search websites (OCC Mundial, Bumeran, and CompuTrabajo), we identify approximately 235,000 explicitly gender-targeted ads. With this sample, we analyze the frequency of the words used to describe jobs targeted to men relative to those targeted to women. Since only a small percentage of ads are gender-targeted (i.e., explicitly discriminatory ads), we use a random forest algorithm to classify non-gendered ads as directed toward men or women according to the language used. We call this set of ads with predicted gender-targeting "implicitly discriminatory" ads.[6] Finally, we estimate the gender salary gap for both explicitly and implicitly discriminatory ads.

Our findings show that ads with "communal" words (often related to women) specify salaries that are 13 percent less than those without such words. On the other hand, ads with "agentic" words (often related to men) offer salaries that are 8 percent higher than those without such words. Explicitly discriminatory ads show a gender salary gap of 13 percent in OCC Mundial; however, there is no substantial gap in Bumeran or CompuTrabajo.

Following Oster (2018), we train a random forest algorithm to identify whether non-gendered job ads are implicitly targeted at men or women. The algorithm shows a good in-sample prediction behavior, with false assignments reasonably low for all three websites. Using this model to identify the implicitly-gendered ads, we estimate that they have a wage gap of 35 percent in Bumeran, 21 percent in OCC Mundial, and 9 percent in CompuTrabajo. That is, ads that convey gender only indirectly, through stereotyped language, have greater gender salary gaps than ads that are explicitly gendered.

The rest of this paper is organized as follows. Section 2 describes our data collection and cleaning process. Section 3 presents our three main results. Finally, Sect. 4 discusses the results and offers some concluding remarks.

Data

The data was gathered by scraping job advertisements from three different employment search websites in Mexico: OCC Mundial, Bumeran, and CompuTrabajo. These three private-sector platforms have become essential tools for Mexican job seekers and companies recruiting workers, and they also have an important presence in other Latin American countries. Such websites have the advantage of describing the skills and characteristics required for certain jobs with great precision. They are not, however, a random sample of all vacancies in the country, so our results are not representative of the entire Mexican labor market.

Data collection from all three websites began in February 2018, and continued through January 2020, with a break in data collection from CompuTrabajo from May to December 2018 (see Table 1). The dataset includes a total of 2,638,754 job

[6] Table 5 shows examples of advertisements classified as explicit and implicit discriminatory ads.

Table 1 Descriptive Statistics

Category		OCC Mundial			Bumeran (Bum)			CompuTrabajo (CT)			p-value Ho: Male = Female		
		Total	Male	Female	Total	Male	Female	Total	Male	Female	OCC	Bum	CT
	Number of observations	1,560,716	58,489	55,760	555,592	22,386	24,364	522,446	41,330	32,957	-	-	-
	Start and end date	February 2018 to January 2020			February 2018 to January 2020			February 2018 to May 19, 2018 and December 2018 to January 2020					
Demographic Characteristics Specified	Specifies woman	0.04	0	1	0.04	0	1	0.06	0	1	-	-	-
	Specifies man	0.04	1	0	0.04	1	0	0.08	1	0	-	-	-
	Includes salary	0.99	1.00	1.00	0.59	0.68	0.65	0.38	0.47	0.45	0.05	0.00	0.00
	Mean real salary (MXN/mo.)	$12,597	$11,709	$10,362	$13,571	$8,866	$8,913	$8,439	$7,925	$7,859	0.00	0.62	0.23
	Specifies age	0.23	0.69	0.64	0.38	0.79	0.71	0.83	0.91	0.91	0.00	0.00	0.92
	Minimum age	24.67	25.46	24.18	22.66	23.81	23.62	22.81	23.52	22.99	0.00	0.00	0.00
	Maximum age	39.63	40.10	36.75	41.99	40.13	38.52	42.07	40.69	38.33	0.00	0.00	0.00
	Specifies location	0.99	1.00	0.99	0.93	0.95	0.95	0.87	0.90	0.86	0.00	0.10	0.00
	Specifies experience	0.72	0.84	0.83	0.79	0.87	0.83	0.79	0.84	0.83	0.00	0.00	0.00
	Specifies married	0.00	0.04	0.01	0.00	0.03	0.01	0.01	0.03	0.00	0.00	0.00	0.00
	Specifies single	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.00
	Specifies student	0.03	0.02	0.03	0.04	0.01	0.02	0.02	0.02	0.02	0.00	0.00	0.00
	Jr. High Diploma	0.11	0.14	0.03	0.10	0.24	0.06	0.14	0.30	0.08	0.00	0.00	0.00
	High school diploma	0.22	0.24	0.29	0.33	0.30	0.42	0.33	0.30	0.39	0.00	0.00	0.00
	Some school	0.08	0.09	0.16	0.11	0.11	0.15	0.10	0.09	0.14	0.00	0.00	0.00
	Tech. Cert	0.02	0.06	0.03	0.02	0.06	0.02	0.03	0.07	0.03	0.00	0.00	0.00

Table 1 (continued)

Category		OCC Mundial			Bumeran (Bum)			CompuTrabajo (CT)			*p*-value Ho: Male = Female		
		Total	Male	Female	Total	Male	Female	Total	Male	Female	OCC	Bum	CT
	Bachelor's degree	0.36	0.29	0.48	0.39	0.29	0.44	0.36	0.28	0.44	0.00	0.00	0.00
	Engineering degree	0.17	0.26	0.08	0.13	0.12	0.05	0.10	0.14	0.05	0.00	0.00	0.00
Communal	Commitment	0.10	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.03	0.12	0.00	0.00
	Punctual	0.07	0.10	0.11	0.09	0.14	0.12	0.08	0.11	0.09	0.00	0.00	0.00
	Honest	0.02	0.07	0.05	0.03	0.05	0.04	0.03	0.07	0.05	0.00	0.00	0.00
	Attentive	0.24	0.17	0.39	0.32	0.24	0.40	0.31	0.21	0.41	0.00	0.00	0.00
	Teamwork	0.22	0.17	0.15	0.11	0.12	0.10	0.09	0.09	0.09	0.00	0.00	0.43
	Helpful	0.07	0.09	0.13	0.08	0.08	0.11	0.08	0.09	0.11	0.00	0.00	0.00
	Courteous	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.01	0.03	0.00	0.00	0.00
	Enthusiasm	0.02	0.01	0.03	0.04	0.01	0.03	0.03	0.02	0.02	0.00	0.00	1.00
Agentic	Control	0.17	0.25	0.22	0.13	0.21	0.20	0.14	0.18	0.17	0.00	0.04	0.00
	Initiative	0.03	0.04	0.04	0.03	0.04	0.04	0.02	0.03	0.04	0.00	0.46	0.00
	Motivation	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.02	0.02	0.00
	Pressure	0.08	0.14	0.13	0.09	0.12	0.09	0.09	0.11	0.10	0.00	0.00	0.00
	Proactive	0.07	0.11	0.13	0.09	0.12	0.10	0.08	0.08	0.10	0.00	0.00	0.00
	Responsible	0.09	0.15	0.14	0.11	0.14	0.10	0.29	0.30	0.28	0.01	0.00	0.00
Appearance	Requests photograph	0.04	0.07	0.12	0.03	0.03	0.06	0.04	0.04	0.08	0.00	0.00	0.00
	Specifies good appearance	0.11	0.11	0.31	0.13	0.14	0.32	0.13	0.09	0.30	0.00	0.00	0.00
Language	English	0.17	0.11	0.15	0.18	0.07	0.10	0.11	0.06	0.11	0.00	0.00	0.00
	Common computer software	0.10	0.11	0.15	0.09	0.07	0.11	0.11	0.10	0.14	0.00	0.00	0.00
Customer service	Sales	0.33	0.32	0.39	0.43	0.30	0.47	0.42	0.30	0.44	0.00	0.00	0.00

Table 1 (continued)

Category		OCC Mundial			Bumeran (Bum)			CompuTrabajo (CT)			*p*-value Ho: Male = Female		
		Total	Male	Female	Total	Male	Female	Total	Male	Female	OCC	Bum	CT
	Customer	0.42	0.29	0.50	0.48	0.29	0.54	0.42	0.28	0.51	0.00	0.00	0.00
	Follow-up	0.16	0.15	0.23	0.17	0.12	0.21	0.15	0.10	0.20	0.00	0.00	0.00
Availability	Availability	0.03	0.04	0.04	0.03	0.04	0.03	0.04	0.05	0.04	0.60	0.00	0.00
	Travel	0.07	0.13	0.05	0.04	0.08	0.04	0.28	0.28	0.22	0.00	0.00	0.00
	Driver's license	0.02	0.09	0.02	0.02	0.07	0.01	0.03	0.08	0.01	0.00	0.00	0.00
Career	Growth	0.16	0.17	0.20	0.28	0.15	0.17	0.25	0.21	0.20	0.00	0.00	0.04
	Development	0.03	0.02	0.05	0.04	0.01	0.08	0.03	0.02	0.05	0.00	0.00	0.00
	Training	0.18	0.14	0.18	0.26	0.12	0.20	0.25	0.14	0.19	0.00	0.00	0.00
Benefits	Bonus	0.19	0.17	0.19	0.30	0.22	0.25	0.25	0.20	0.22	0.00	0.00	0.00
	Benefits	0.68	0.72	0.72	0.58	0.78	0.74	0.66	0.76	0.74	0.06	0.00	0.00
	Insurance	0.04	0.03	0.04	0.03	0.03	0.02	0.04	0.03	0.02	0.00	0.00	0.00
	Commissions	0.13	0.09	0.19	0.25	0.10	0.29	0.24	0.10	0.22	0.00	0.00	0.00
	Base salary	0.25	0.23	0.30	0.36	0.23	0.34	0.36	0.29	0.33	0.00	0.00	0.00

Data collection from CompuTrabajo was interrupted from May to December 2018 because of a technical issue. Statistics were calculated by the authors using data from OCC Mundial, Bumeran, and CompuTrabajo. Each word is a dummy variable that is 1 if the word appears in the job description and 0 otherwise. See Table S1 for the Spanish words used advertisements, with two restricted samples including explicitly gender-targeted ads: 113,081 seeking women and 122,205 seeking men.

To download and process the job advertisement data, we use different algorithms for each website, but the information in all three is alike, allowing us to build a unified database. Generally, each job advertisement has some defined fields (salary, age, company) and some text. To generate our variables of interest, we use text analysis to find keywords related to sociodemographic characteristics (gender, salary, age, location, experience, marital status, and required education), skills, and explicit discrimination. We identify a set of keywords for specific traits and skills required for the job: educational requirements, age, work experience, driver's license, and language requirements, as well as personal traits like responsibility, teamwork, and commitment. For a list of the Spanish terms with English translations, see Table S1 in the Supplementary Materials.

Descriptive statistics for data from the three websites and the samples are shown in Table 1.[7] Only about 10 percent of all job advertisements explicitly specify male or female applicants; CompuTrabajo has the highest proportion of gender-targeted ads (14.2 percent). Almost all job advertisements in OCC Mundial specify a salary (99 percent), while only about half of the ads in Bumeran (59 percent) and CompuTrabajo (38 percent) do so. Among those that include salary, the average monthly salary for all three websites is around $8000 MXN (approximately $400 USD). Gender-targeted ads specify lower salaries, which is probably due to the negative relationship between gender targeting and schooling described in Kuhn and Shen (2013a, b). The greatest gender salary gap is close to 13 percent, or $1350 MXN ($67.50 USD) monthly, on OCC Mundial.[8]

Some characteristics are more frequent in gender-targeted ads than in the ads as a whole. The most striking fact is that gender-targeted ads also discriminate by age. For example, in OCC Mundial, although only a quarter of the ads specify age, the proportion increases to 70 percent in gender-targeted ads, and the difference is similar in Bumeran and CompuTrabajo. Moreover, when gender is specified, women are usually required to be younger than men (the maximum age specified in ads in OCC Mundial is 36.7 for women vs. 40 for men, in Bumeran the figures are 38.1 vs. 39.9, and in CompuTrabajo they are 38.3 vs. 40.6). In general, job ads are restricted to relatively young applicants (approximately 32 years old), confirming the "age twist" found by Delgado Helleseter et al. (2018).

Other common specifications in gender-targeted ads are experience and marital status. For instance, in OCC Mundial, 73 percent of the ads require experience, but the figure is 84 percent for gender-targeted ads, evenly balanced by gender. In the sample as a whole, specification of marital status is not common, but it appears

[7] We also perform a classification by occupation using specific keywords in the job titles. However, only 80 percent of the ads could be classified. See Table S2 in the Supplementary Materials.

[8] The multiplicity of sites allows us to identify gender and labor dynamics with a representation of different labor markets. We report results separately to acknowledge the fact that a large number of our job ads come from OCC Mundial, and that gender targeting and salary transparency are different among the platforms.

more often in gender-targeted ads. Interestingly, in gender-targeted ads, it is more common to specify married men (4 percent) and single women (2 percent).

Some educational requirements are more common in ads specifying men than in those specifying women, including technical school and junior high school diplomas: in OCC Mundial these were requested in 14 percent of the ads directed at men, as compared with only 3 percent of those directed at women. Engineering degrees were specified in 26 percent of ads targeted at men, but only 8 percent of those targeted at women. Advertisements requiring only some college or bachelor's degrees are more frequent in ads for women than in those for men.

In addition to these sociodemographic data, we also search for specific skill requirements associated with gender stereotypes. We follow the classifications of Heckman and Kautz (2012), as shown in the first column of Table 1. The first two specifications are "communal" and "agentic" characteristics. Here, we follow a vast literature on non-cognitive skills that discusses returns to personal traits (Heckman and Kautz 2012) as well as a literature on the categorization of personality traits (Asch 1946; Saucier 2009; Spence et al. 1975). The classification distinguishes between communal content, which refers to relationships and social functioning, and agentic content, which refers to goal achievement and task functioning. These characteristics have been related to stereotypes in women and men (see, for example, Rudman and Phelan 2010; Spence et al. 1975). In general, traits such as independence, leadership, and decision-making are stereotypically associated with men, while those like understanding, warmth, and gentleness are associated with women (Abele and Wojciszke 2014).

We classify the traits specified in job ads as communal if they apply to enhancing group work and agentic if they apply to individual work or goal achievement. Examples of words referring to communal characteristics are *commitment*, *punctual*, *honest*, *attentive*, *teamwork*, *helpful*, *courteous*, and *enthusiasm*. All of these qualities enhance social interaction and relationships. Examples of words referring to agentic characteristics are *control*, *initiative*, *motivation*, *pressure*, *proactive*, and *responsible*.

Table 1 shows descriptive statistics for traits related to communal and agentic characteristics. On the one hand, some communal characteristics, such as commitment, are used equally in ads directed at both genders. Others, such as honesty and teamwork, are mainly found in ads targeted to men. Ads directed at women are twice as likely to specify attentiveness, courtesy, and enthusiasm than those directed at men. Agentic characteristics (for example, initiative, motivation, proactivity, and pressure) are specified equally in ads directed at men and at women.

We also analyze the requirement of appearance in job advertisements, following Kuhn and Shen (2013b), who find that Chinese employment search websites require attractiveness in job advertisements. This category includes use of the words *photograph* or *appearance*. Table 1 shows that on all three websites these characteristics are specified twice as often in ads targeted to women.

We classify other important skill requirements that are common in our dataset, following Deming and Kahn (2018). These categories include use of the words *language*, *software*, *customer service*, *availability*, *driver's license*, *career development*, and *benefits.* Ads on all three websites ask for specific abilities from women, for example, knowledge of English or of software like Word, Excel, or Windows.

Customer service terms like *client*, *follow-up*, and *sales* are also more frequent in ads directed at women. Terms referring to travel and time availability are more common in ads directed at men, which may be related to household gender roles and employers' expectations that women have less flexibility to travel or work extended hours.

We also find substantial differences in what employers offer apart from salary. Companies are described as good places to work, with opportunities for training and advancement slightly more frequent in ads targeted to women. Benefits such as bonus payments, commissions, and private medical insurance are described roughly equally for both genders. One striking difference is that ads targeted to women are twice as likely to specify compensation by commission.

Results

Job Advertisements Use Gender Stereotypes

The main aim of this section is to analyze the most frequent words in job advertisements for each website, using single-word analysis, bigram analysis, and category analysis. We analyze advertisements that explicitly specify men or women. First, stop words (highly common words like *on*, *a*, *some*, *like*, *the*) and diacritical marks were removed from the dataset. A count was made of unique words for each gender, using the root of the word and disregarding the gender inflection. For example, *honest* was counted as a single word regardless of whether it appeared in the masculine form *honesto* or the feminine form *honesta*.[9]

Because the total number of unique words is different for each gender (as in Arceo-Gomez and Campos-Vazquez 2019b),[10] we calculate a ratio for database b (OCC Mundial, Bumeran, or Computrabajo) as shown in Eq. (1): the number of mentions of word i as a proportion of the total number of unique words in ads directed at women, over the same proportion in ads directed at men. If the specifications of worker characteristics are the same for both genders, this ratio equals one. If it is larger than one, then word i is more common in ads directed at women than in those directed at men.

$$Ratio_{i,b} = \frac{\frac{\sum_{ad} 1(word=i)_{ad,b,female}}{\sum_{ad} Numberofwords_{ad,b,female}}}{\frac{\sum_{ad} 1(word=i)_{ad,b,male}}{\sum_{ad} Numberofwords_{ad,b,male}}} \tag{1}$$

[9] We do not exploit the gender inflection to classify announces by gender as the neutral gender in Spanish has the same structure as the masculine. We can only classify words as feminine that are specifically inflected for gender. There are also neutral words that have neither masculine nor feminine inflection (for example, receptionist [recepcionista], typist [capturista], masseuse [masajista], or analyst [analista]). In figure S11, we classify implicit discrimination for some occupations using gender inflection of nouns.

[10] This difference is not large. The mean word count of an ad directed at women is 130.75 words; for men it is 132.59 words (women: 128.83 in OCC Mundial, 117.68 in Bumeran, and 145.73 in CompuTrabajo; men: 129.43 in OCC Mundial, 120.32 in Bumeran, and 147.85 in CompuTrabajo.).

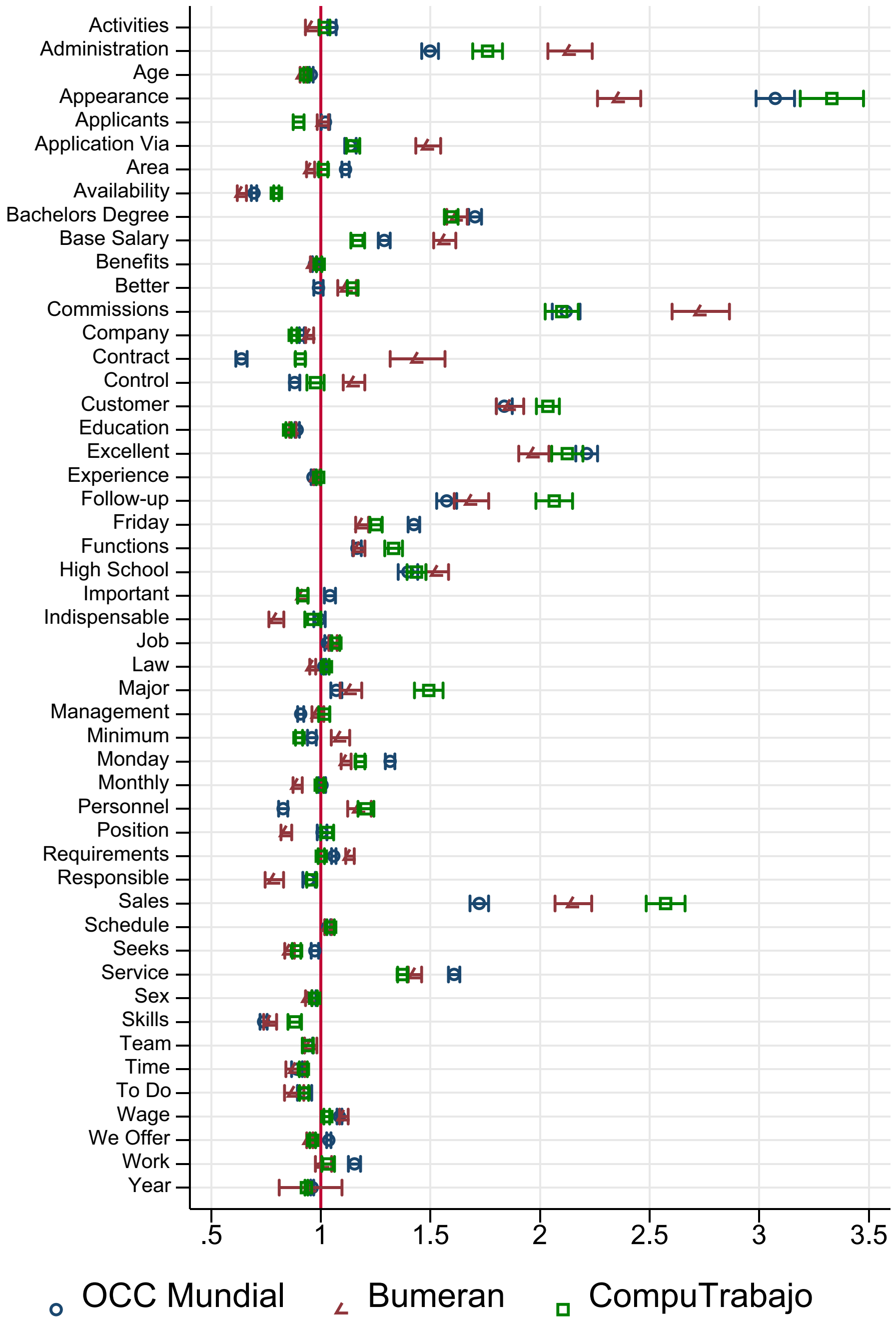


Fig. 1 Ratios of frequencies of most commonly used words. Note: Lines represent 95 percent confidence intervals calculated with 200 random bootstrap samples. Adjacent points indicate similar word roots with different endings. Each point represents the ratio between the proportions of word occurrence for each website. Stop words and punctuation are omitted. Sample restricted to gender-targeted ads. The original Spanish words are in Table S3

To obtain a ranking of the most frequent words overall, we count the number of unique words in the total set of advertisements directed at both genders. We then calculate the ratio of the frequency of each word with respect to the total sample. The frequency of the 50 words with the highest ratio represents 28.2–29.2 percent of the total frequency of words.[11]

Figure 1 shows these ratios for the top 50 words for each website. The words are presented in alphabetical order. These words can be divided into two categories: the first expressing basic characteristics of the job, such as location and desired sociodemographic characteristics of the applicant, and the second related to the specific skills sought. The most frequent words expressing basic characteristics are *year*, *experience*, *job*, *benefits*, *age*, *law*,[12] *company*, *sex*, *schedule*, *salary*, *requirements*, *education*, *Monday*, *monthly*, *Bachelor's degree*, *area*, *seeks*, *Friday*, *activities*, *functions*, *position*, *minimum*, *base salary*, *personnel*, *time*, *better*, *work*, *to do*, *contract*, *high school*, *major*, *we offer*, *applicants*, *indispensable*, and *commissions.* The most frequent words describing desired skills are *service*, *customer*, *availability*, *excellent*, *sales*, *management*, *team*, *follow-up, skills*, *control*, *appearance*, *administration*, and *responsible.*

The first set of these most frequent words describes the basic characteristics of the job opportunity. Some of them are about hours, location, and contact information (e.g., *Monday*, *Friday*, *schedule*, *time*, *area*, *applicants*, *company*), some are about the nature of the work (e.g., *position*, *work*, *job*, *to do*, *functions*, *activities*, *requirement*, *request*, *personnel*, *indispensable*). Other words are about the applicant's characteristics and qualifications (e.g., *high school*, *bachelor's degree*, *education*, *age*, *year*, *sex*, *major*, *experience*). Finally, there are words referring to the conditions of the employment, such as type and frequency of payment (e.g., *monthly*, *commissions*, *base salary*, *contract*, *salary*, *law*, *responsible*, *benefits*).

The following words describing the nature of the work are generally balanced between genders, with ratios close to one: *year*, *experience*, *work*, *benefits*, *age*, *law*, *company*, *sex*, *requirements*, *offer*, *salary*, *monthly*, *zone*, *seeks*, *applicant*, *position*, *area*, and *job*. However, advertisements are more specific regarding women's hours: the word *schedule* appears 1.1 times more often in ads directed at women, and the words *Monday* and *Friday* also appear more often (1.2–1.35 and 1.3–1.4 times). Advertisements also tend to be more explicit about women's education: *bachelor's degree* and *high school* are repeated nearly 1.5–1.7 more times in ads directed at women. These ads also tend to mention more details about salary: *base salary* and *commissions* are almost twice as frequent as in ads directed at men. However, *contract* is only 0.6–0.8 times as frequent in ads directed at men in OCC Mundial and CompuTrabajo.

Words about characteristics that appear in ads directed at both genders are *team* and *management.* Those mentioned in ads targeted to women include *appearance*,

[11] In OCC Mundial, these words represent 28.23 percent of the total, in Bumeran, 29.56 percent, and in CompuTrabajo, 29.20 percent.

[12] The word *ley* ("law") appears frequently in Mexican job advertisements in the phrase *prestaciones de ley* ("benefits as provided by law") or *prestaciones superiores a la ley* ("benefits superior to those required by law").

Fig. 2 Ratios of frequencies of most commonly used bigrams. Note: Lines represent 95 percent confidence intervals calculated with 200 random bootstrap samples. Adjacent points indicate similar word roots with different endings. Each point represents the ratio between the proportions of bigram occurrence for each website. Stop words and punctuation are omitted. Sample restricted to gender-targeted ads. The original Spanish words are in Tables S4 and S5

which is 2.2–3.4 times more frequent than in ads directed at men. Women are also asked to have more interaction with customers and perform more administrative tasks, as seen in the following words found more frequently in ads directed at them: *sales* (with a ratio of 1.6–2.6), *follow-up* (1.7–2.3), *service* (2.3–2.6), *customer* (1.7–2.1), and *administration* (1.4–2.2). Ads targeting men mention other words, like *skills* (0.75-0.9), *responsibility* (0.75-0.9), *time* (0.7-0.8), and *availability* (0.6-0.8).

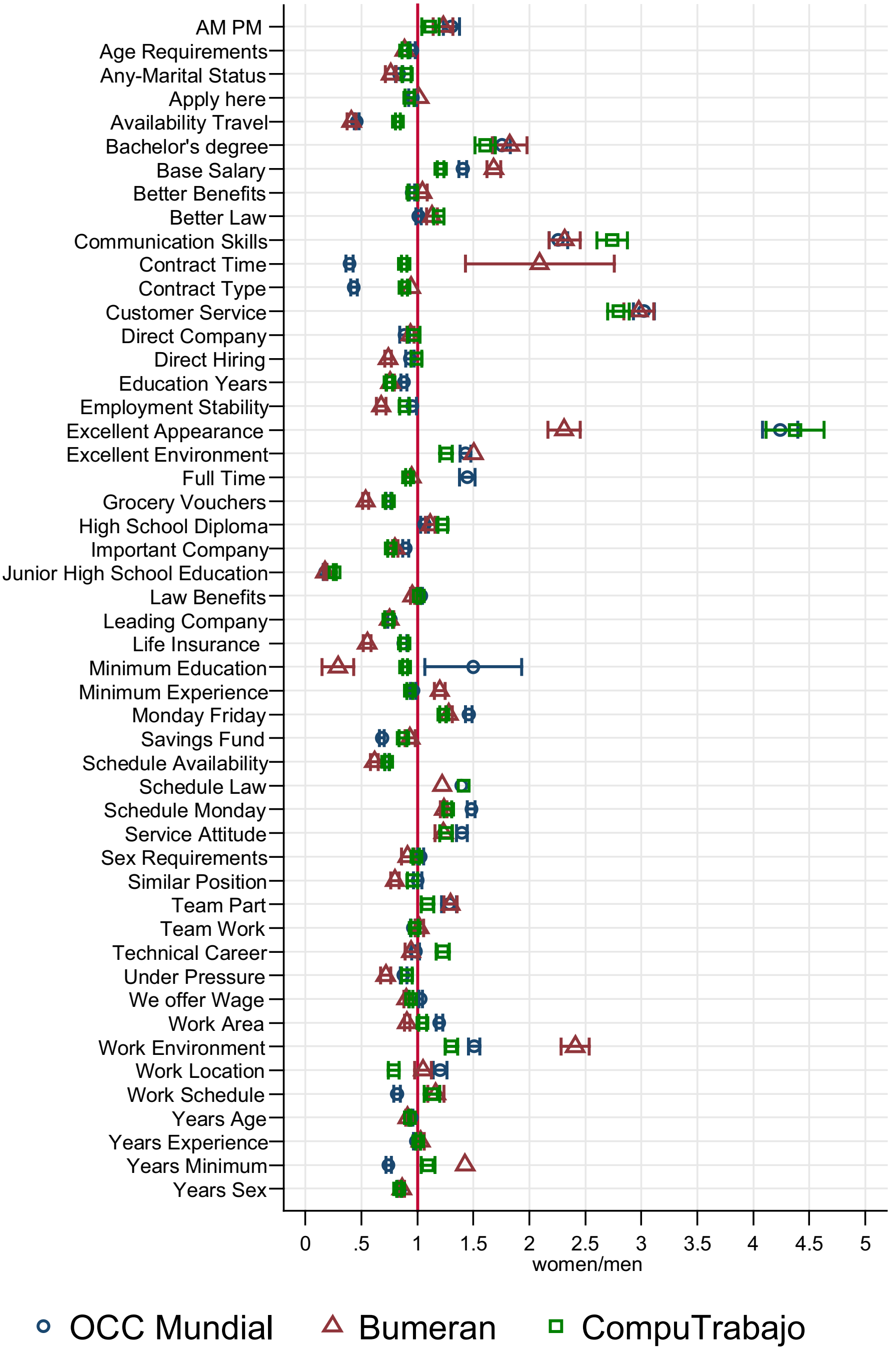
In general, advertisements tend to use nearly the same words to describe the basic characteristics of the job. However, some words related to education, hours, and salary are more frequent in ads directed at women. Words related to applicant characteristics and skills more frequently specify good appearance for women, and ads directed at them include more customer service-related terms like *sales*, *follow-up*, *service*, and *customer.* Words in this category directed at men include *availability*, *time*, *responsibility*, and *skills.*

In order to provide more context for the analysis, we calculate the same ratios for bigrams. Punctuation and stop words are removed and unique pairs of words are counted, independent of word order. The ratio between the proportions of bigrams are then calculated for each website. The results are shown in Fig. 2. Standard errors are larger here than in the single-word analysis. The bigrams are classified into three categories: description of general characteristics of the job, description of sociodemographic characteristics and skills sought in applicants, and description of salary and benefits.

Many of the bigrams are related to the basic job description, with some related to job conditions such as hours, contact information, and location. Some of these bigrams are *Monday-Friday*, *schedule-Monday*, *work-location*, *contract-time*, *full-time*, *apply here*, and *a.m.-p.m*. Words related to days and hours appear more often in advertisements directed at women. *Monday-Friday* has a ratio of 1.2–1.4, and the ratio for *schedule-Monday* is similar.

The second set of bigrams is related to applicants' sociodemographic characteristics (e.g., *years-age*, *years-experience*, *years-sex*, *age-requirements*, *sex-requirements*, *technical-degree*, *high school-education*, *junior high school-education*, *education-years*, *similar-position*, *years-minimum*, *any-marital status*, *and bachelor's-degree*). Almost all of these sociodemographic bigrams are equally balanced between men and women, including *years-age*, *years-experience*, *minimum-experience*, *age-requirements*, and *sex-requirements*. *Bachelor's-degree* is more frequent in ads targeted to women, with a ratio of 1.5–2. Finally, *any-marital status* is slightly dominant in advertisements targeted to men (with a ratio of 0.7-0.95), as is *junior high school-education* (0.2-0.4).

Bigrams that describe applicants' skills and characteristics, such as *team-work*, show similar frequency for both genders. Those that appear more often in ads


AM PM
Age Requirements
Any-Marital Status
Apply here
Availability Travel
Bachelor's degree
Base Salary
Better Benefits
Better Law
Communication Skills
Contract Time
Contract Type
Customer Service
Direct Company
Direct Hiring
Education Years
Employment Stability
Excellent Appearance
Excellent Environment
Full Time
Grocery Vouchers
High School Diploma
Important Company
Junior High School Education
Law Benefits
Leading Company
Life Insurance
Minimum Education
Minimum Experience
Monday Friday
Savings Fund
Schedule Availability
Schedule Law
Schedule Monday
Service Attitude
Sex Requirements
Similar Position
Team Part
Team Work
Technical Career
Under Pressure
We offer Wage
Work Area
Work Environment
Work Location
Work Schedule
Years Age
Years Experience
Years Minimum
Years Sex
0
.5
1
1.5
2
2.5
3
3.5
4
4.5
5
women/men
OCC Mundial
Bumeran
CompuTrabajo

targeted at women are *service-attitude* (1.2–1.5), *customer-service* (2.6–3.0), *communication-skills* (2.2–2.9), and *excellent-appearance* (2.2–4.7); those more frequent in ads targeted to men are *schedule-availability* (0.6-0.75), *availability-travel* (0.4-0.9), and *under-pressure* (0.7-0.95). These findings are similar to those of the single-word analysis.

Many bigrams are related to salary and benefits, such as *law-benefits*, *we offer-salary*, *base-salary*, *better-law*, *better-benefits*, *grocery-vouchers*, and *savings-fund* are equally frequent in ads targeted to men and women. An exception is *life-insurance*, which is more frequent in ads targeted to men (0.5-0.9). Another set of bigrams refers to work environment. This category includes *work-environment*, *excellent-environment* and *leading-company*. The first two bigrams are more frequent in ads directed at women, while *leading-company* is more frequent in ads directed at men.

The analysis of bigrams shows the same tendencies as the single-word analysis. However, it also offers new insight into the most specified characteristics in advertisements directed at men. One example is the explicitly-stated irrelevance of their marital status; another is the specification of workers with only a junior high school education. Bigrams also offer robustness to some results, including the finding that sales and customer service jobs are more often directed at women, and they confirm that women are discriminated against based on their physical appearance. Finally, the bigram analysis confirms that advertisements directed at men are more likely to specify time availability. The frequency of words could be related to the type of job advertised.

The third part of this descriptive analysis is a study of categories of words. Most of the repeated words and bigrams are grouped into different categories, including communal or agentic characteristics, appearance, language, software, customer service, availability, driver's license, career development, and benefits. Table 1 shows the specific arrangement of these categories of words; in addition to the most frequent words, other descriptive words are also included.

Figure 3 shows the gender ratios of the frequency of these categories of words. The presentation is similar to that of Figs. 1 and 2. Characteristics are categorized as communal if they refer to group work and agentic if they refer to individual work or achievement of individual goals (Abele and Wojciszke 2014). Words associated with communal characteristics include *commitment*, *motivation*, *punctual*, *honest*, *attentive*, *helpful*, *teamwork, enthusiasm and courteous*; those considered agentic include *pressure*, *control*, and *proactive*. Communal characteristics are more frequent in ads targeted to women, with a ratio of 1.25–1.5 for the three websites. Agentic characteristics are evenly balanced in OCC Mundial and CompuTrabajo, but in Bumeran are more frequent in ads directed at men.

Words in a number of categories are used more often in ads directed at women. Words in the Appearance category, which includes *appearance* and *photograph*, are used nearly 3.2 times as often, and words in the Language category, which includes the word *English*, are repeated 1.3–2.3 times as often. Customer Service words such as *sales*, *customer*, and *follow-up* are used 1.6–1.8 times as often, and words in the Career Development category are used 1.4–2.1 times as often. However, consistent with the findings of the single-word and bigram analyses, ads directed at men use words in

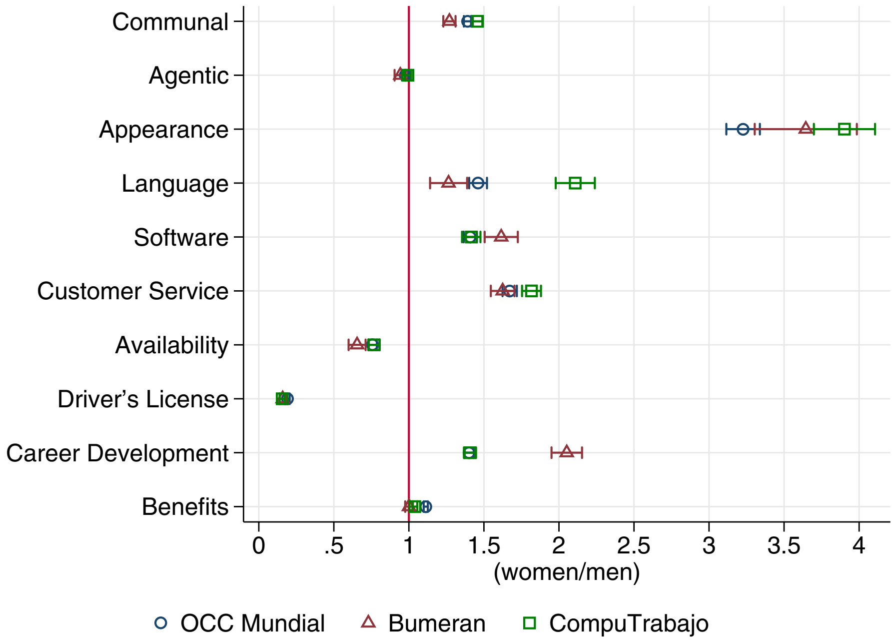


Fig. 3 Ratio of frequencies of most commonly used categories. Note: Lines represent 95 percent confidence intervals calculated with 200 random bootstrap samples. Adjacent points indicate similar word roots with different endings. Each point represents the ratio between the proportions of category occurrence for each website. Stop words and punctuation were omitted. Communal category includes the words *commitment*, *punctual*, *honest*, *attentive*, *teamwork*, *helpful*, and *courteous,* and *enthusiasm*; Agentic category includes the words *control*, *initiative*, *motivation*, *pressure*, *proactive*, *responsible*, *stress*; Appearance category includes the words *photograph* and *appearance*. The Language category includes the word *English*; Software category includes the words *Excel*, *Word*, and *Windows*; Customer Service category includes the words *sales*, *customer*, and *follow-up*. Availability category includes the words *time* and *travel*; Career category includes the words *growth*, *development*, and *training*; Benefits category includes the words *bonus*, *benefits*, *insurance*, *commissions*, and *base salary*. Sample restricted to gender-targeted ads

the Availability category, which refer to time availability, travel, and having a driver's license, more often than those directed at women: the ratio for this category is 0.6-0.9.

Overall, the analysis of categories indicates that ads directed at women specify more communal characteristics as well as those related to appearance, language, software, and customer interaction, while those directed at men more frequently specify time and travel availability (as measured by the requirement of a driver's license). Agentic characteristics are specified equally for both genders.

Specific Words Are Related to Salary

In this section, we analyze the relationship between words and the salary specified in job advertisements using three different estimations. The first calculates the effect of a word on the salary specified and its asymmetric effect in

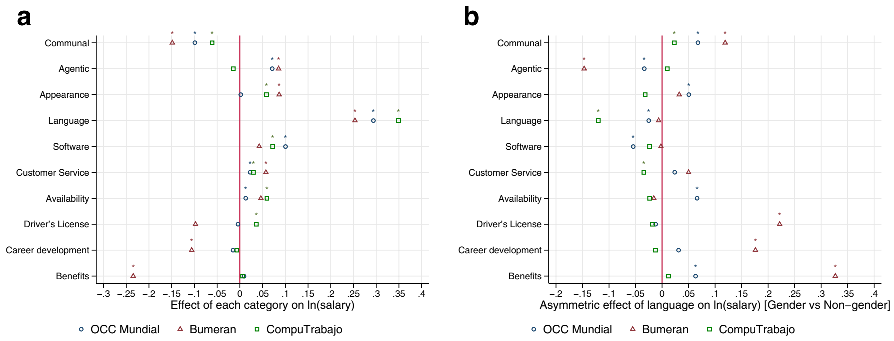


Fig. 4 Effect of categories on salary (targeted vs. non-targeted). Note: Regression includes all advertisements from the three websites in the period analyzed. The dependent variable is the ln of real salary (July 2018 = 100). Each category is a dummy variable that represents the appearance of at least one component word in the advertisement. In Panel A, each point represents the coefficient of the ln(salary) in this dummy. In Panel B, each point represents the coefficient of the ln(salary) in the interaction of a dummy variable for gender-targeted with the dummy for presence of the category. The regression controls for the presence of the category and specifies a gender, state, age (minimum and maximum), experience, and educational level. An asterisk (*) signifies that the regression is statistically significant at 5 percent; robust standard errors are clustered by location. Communal includes the words *commitment*, *punctual*, *honest*, *attentive*, *teamwork*, *helpful*, *courteous*, and *enthusiasm*; Agentic includes the words *control*, *initiative*, *motivation*, *pressure*, *proactive*, and *responsible*; Appearance includes the words *photograph* and *appearance*; Language includes the word *English*; Software includes the words *Excel*, *Word*, and *Windows*; Customer Service includes the words *sales*, *customer*, and *follow-up*; Availability includes the words *time* and *travel*; Career includes the words *growth*, *development*, and *training*; Benefits includes the words *bonus*, *benefits*, *insurance*, *commissions*, and *base salary*

gender-targeted and non-targeted ads. The second method is a LASSO estimation to analyze the extent to which different words explain the salary variation in ads directed at women and at men. The final estimation is an analysis of adjusted *R*-squared to determine which word or characteristic explains more of the variation.

The analysis of the asymmetric salary effect between gender targeted ads and non-targeted ones consists of the regression in Eq. (2) with (log) specified salary as a dependent variable and an interaction of each category with a gender-targeted dummy controlled by the presence of the word and the dummy:

$$\ln(salary)_{ad,base} = \alpha_0 + \alpha_1 d.GenderTargeted_{ad} + \sum_j \left(\delta_j skill_{j,ad}\right) + \sum_j \beta_j (skill_{j,ad} * d.GenderTargeted_{ad}) + \sum_i \begin{matrix}\gamma_i(X_{i,ad} \\ +d.X_{i,ad})\end{matrix} + u_{ad} \tag{2}$$

This regression is at the ad level. S*alary* is the real monthly salary specified, and *skill* is a matrix of dummies of categories presented in Table 1 (*communal*, *agentic*, *appearance*, *language*, *software*, *customer service*, *availability*, *license*, *career development*, and *benefits*). This variable is equal to 1 if a word included in the category is included in the job ad and 0 otherwise. $X_{i,base}$ is a matrix of demographic controls, which includes marital status, age, level of education, experience, and location. Finally, $d.X_{ad}$ is a matrix of dummy variables that denotes the presence of demographic characteristics in the ad, equal to 1 if the demographic characteristic is present and 0 otherwise.

We are interested in two effects of this regression. The first is the effect of a skill or characteristic on the salary specified among non-targeted ads (δ_j). The second is the coefficient of the interaction (β_j), which can be interpreted as the differentiated effect of a skill or characteristic on the gender-targeted salary compared to the non-gender-targeted salary. If there are no salary differences between targeted and non-targeted ads, we expect the interaction coefficient to equal zero. A positive coefficient implies an association between the word and a higher salary in gender-targeted ads; a negative coefficient implies an association with a higher salary in non-targeted ads.[13]

Figure 4 shows the effects of each category on salary and the asymmetric effects. The Communal category has a negative effect on the salary specified among non-gender-targeted ads, reducing it by 7–20 percent. However, almost half of this effect is offset in gender-targeted ads, where the presence of these words increases the salary by 2.5–14 percent. This characteristic is more frequent in ads targeted to women, so they are most affected by its presence. Ads including communal words depict an average wage gap of 13 percent with respect to ads that do not include those words (controlling for demographic characteristics).

The Agentic and Language categories behave similarly. Both have a positive effect: words in the Agentic category increase the salary by 11 percent (except in CompuTrabajo) and those in the Language category increase it by 27–37 percent among non-targeted ads. However, both categories are penalized if the ad is gender-targeted. The presence of Language category words in gender-targeted ads reduces the salary by 4–11 percent. The Language category is more frequent in ads directed at women, so they are penalized more than men. On average, ads with words in the Agentic category show an 8 percent higher salary than ads without those words, and ads with Language category words offer salaries that are approximately 32 percent higher (controlling for demographic characteristics).

The presence of words in the Appearance category has a positive effect of 5–7 percent on the salary, except in OCC Mundial. However, words in this category have a positive and significant effect in gender-targeted ads in OCC Mundial. This category, which is more common in ads directed at women, always has a positive effect for gender-targeted ads. The Customer Service category, which is also more frequent in ads directed at women, also has a positive effect on the salary. The Software and Availability categories both have a positive effect on the salary of 4–11 percent, regardless of whether the ad is gender-targeted or not. Finally, words in the Benefits category have no effect on salary, unless the ad is gender-targeted. The presence of these words increases the salary by 6–35 percent in gender-targeted ads.

[13] We also analyze this effect separated by sex in Figure S1. Women have lower compensation for agentic words than men. Driver's License has a positive effect on salary when it is mentioned in ads targeted at women, but not in ads targeted at men. Customer Service has a greater effect when it is mentioned in ads targeted at men.

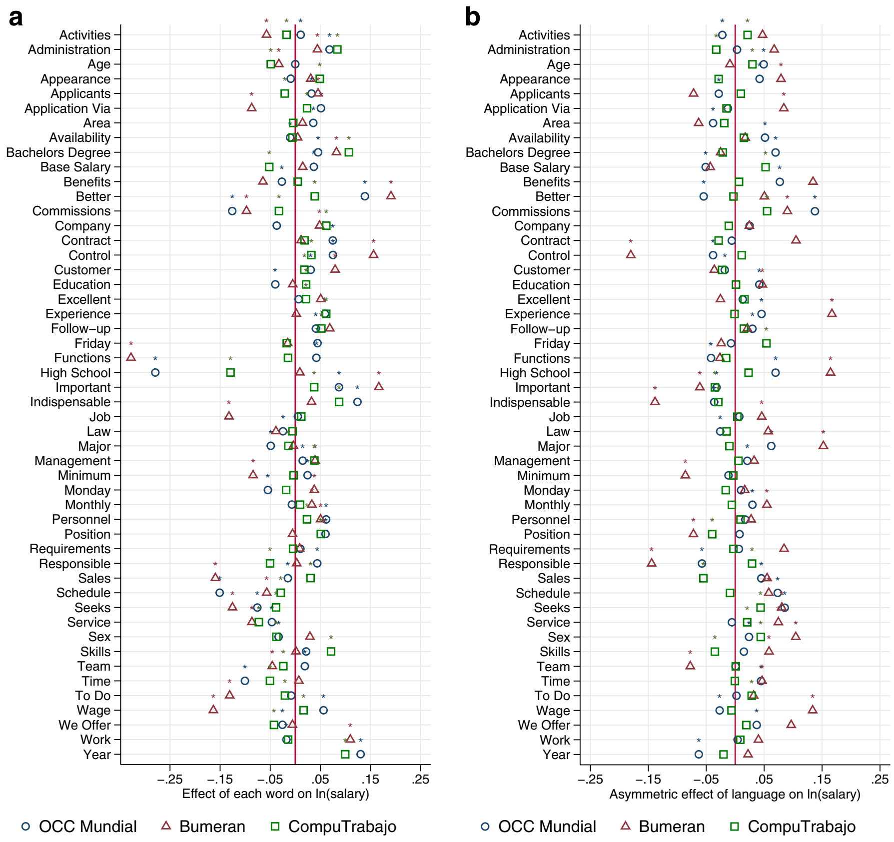


Fig. 5 Asymmetric effect of language on salary for 50 most frequently used words, in alphabetical order. Note: Figure includes all advertisements from the three websites in the period analyzed. The dependent variable is the ln of real salary (July 2018 = 100). Each category is a dummy variable that indicates whether the word appeared in the job advertisement. Each point represents the coefficient of the ln(salary) in the interaction of a dummy variable for gender-targeted with a dummy for presence of the word. The regression controls for location, age (minimum and maximum), experience, and educational level. Asterisk (*) signifies that the regression is statistically significant at 5 percent; robust standard errors are clustered by location

We carry out the same analysis with dummy variables for the most frequently used words described in the previous section. The dummy variable is equal to 1 if the word appears in the advertisement, and 0 otherwise. We then interact this dummy variable with one that indicates whether the ad is gender-targeted. This coefficient is shown in Fig. 5, which illustrates the effect of each word on salary and the differences in salary between gender-targeted and non-targeted ads for the most frequently used words.

The presence of the word *sex* has a robust positive effect on salary of 1–11 percent in gender-targeted advertisements. Advertisements for jobs with the same characteristics offer a higher salary when gender is specified, which could be interpreted as a salary premium if applicants meet the gender requirement. When there is explicit discrimination, as in gender-targeted ads, some agentic characteristics, such as those represented by the words *control* and *responsible*, have a salary penalty of 4–17 percent. Other characteristics are rewarded, such as those represented by the words *appearance* (4–7 percent) and *commissions* (4–14 percent). This result is similar to that found in the *Availability* category. Words such as *administration*, *bachelor's degree*, *better*, and *important* have a positive robust effect on offered salary, and words such as *commission* a robust negative effect of 4–14 percent.

In sum, some characteristics, such as communal ones, have a negative effect on the salaries specified. Others, such as appearance, have a positive effect that is reinforced when the job advertisement is gender-targeted. Gender-targeted advertisements specify different characteristics than those of the labor market in general. Specification of a gender can compensate for some salary penalties, such as those associated with communal characteristics, and penalize others, such as agentic characteristics and language.

The second method is a LASSO regression in the gender-targeted sample. This method helps to select the most important characteristics that explain the variation in salary for each gender. A penalization parameter (λ) for each characteristic expresses its importance. As we are interested in skills or characteristics, the regression does not penalize demographic characteristics (such as age, marital status, educational level, or location). The penalization parameter is selected by cross-validation of ten folds.[14] The post-LASSO regression estimates are an OLS regression with the non-zero coefficients from the LASSO regression.[15]

Table 2 shows the results. Communal characteristics have a negative effect both on men and on women; this result is robust for all three databases and both genders. Salaries are approximately 5.3–9.3 percent lower in advertisements that use words referring to communal characteristics. In the previous analysis of asymmetric effects, we found that communal characteristics have a negative effect on salary, which is offset in gender-targeted ads. In this analysis, we confirm that these characteristics carry a penalty for the salary specified, even in gender-targeted ads. Agentic characteristics have an ambiguous effect by gender and website. The previous analysis shows that the presence of these characteristics increases the salary, except in gender-targeted ads; this characteristic has no consistent effect on salary in gender-targeted ads.

The use of words in the Appearance and Software categories have a robust positive coefficient for both genders. The specified salary is 1.2–16.7 percent greater in ads that mention appearance or photographs, consistent with the previous analysis.

[14] $\left(\widehat{\alpha},\widehat{\beta},\widehat{\gamma}\right) = argmin(Ln\,salary_{ad} - \sum_j \lambda_j JobCharac_{ad} - \sum_k \gamma_k (X_{ad} - d.X_{ad}))^2$ subject to $\sum_j |\lambda_j| \leq t$. $Lnsalary$ is the log of monthly salary specified. $JobCharac$ is a matrix of dummies of words describing the job and applicants' characteristics, as in Table 1 (communal character, agentic character, appearance, language, software, customer services, availability, license, career development, benefits, and occupational classification). λ_j is the penalization term for each word. Finally, X_{ad} are sociodemographic characteristics that are not penalized, such as minimum age, maximum age, and location. $d.X_{ad}$ is a matrix of dummy variables that indicate the presence of sociodemographic characteristics in the ad. Dummies are equal to 1 if the characteristic is present in the ad and 0 otherwise.

[15] See OLS regression on all coefficients in Table S6. The results are very similar to the LASSO regression.

Table 2 LASSO Post-Estimation Results

	OCC Mundial		Bumeran		CompuTrabajo	
	Male	Female	Male	Female	Male	Female
Communal	-0.060	-0.056	-0.093	-0.053	-0.055	-0.055
	[0.004]	[0.004]	[0.01]	[0.008]	[0.005]	[0.006]
Agentic	0.075	0.025	-0.044	-0.026	0.019	
	[0.004]	[0.004]	[0.011]	[0.008]	[0.005]	
Appearance	0.012	0.043	0.167	0.092	0.010	0.050
	[0.006]	[0.004]	[0.013]	[0.008]	[0.008]	[0.007]
Language	0.321	0.266	0.392	0.329	0.278	0.280
	[0.008]	[0.006]	[0.02]	[0.014]	[0.02]	[0.013]
Software	0.048	0.025	0.097	0.047	0.025	0.085
	[0.006]	[0.005]	[0.015]	[0.01]	[0.009]	[0.009]
Customer Service	0.057	0.028	0.126	0.032	0.013	
	[0.004]	[0.004]	[0.01]	[0.009]	[0.005]	
Availability	0.094	0.139	0.071	0.113	0.059	0.056
	[0.006]	[0.007]	[0.015]	[0.019]	[0.009]	[0.011]
Driver's License	-0.073	0.099	0.028	0.081		0.109
	[0.006]	[0.017]	[0.017]	[0.043]		[0.032]
Career Development	0.025	0.029	0.114	0.116		-0.020
	[0.004]	[0.004]	[0.011]	[0.009]		[0.006]
Benefits	0.057	0.073	0.031	0.122	0.006	0.027
	[0.005]	[0.005]	[0.019]	[0.014]	[0.007]	[0.01]
N	58,464	55,748	15,265	15,899	19,349	14,775
Adj. R^2	0.430	0.303	0.314	0.306	0.384	0.329

Dependent variable is ln(salary). Coefficients presented are the result of OLS post-estimation of a LASSO regression. All regressions are controlled without penalty for the following characteristics specified in the advertisement: education, minimum and maximum age, marital status, and job location. Robust standard errors in bold. Communal includes the words *commitment*, *punctual*, *honest*, *attentive*, *teamwork*, *helpful*, *courteous*, and *enthusiasm*; Agentic includes *control*, *initiative*, *motivation*, *pressure*, *proactive*, and *responsible*; Appearance includes *photograph* and *appearance*; Language includes *English*; Software includes *Excel*, *Word*, and *Windows*; Customer Service includes *sales*, *customer*, and *follow-up;* Availability includes *time* and *travel*; Career includes *growth*, *development*, and *training*; Benefits includes *bonus*, *benefits*, *insurance*, *commissions*, and *base salary*

Language is associated with salary in all three websites: salary is 26.6–39.2 percent greater when English is mentioned. This estimate is among the largest effects of skill on salary and is also consistent with the previous analysis.

Customer service has a positive coefficient for men; the salary specified is 1.3–12.6 percent greater where follow-up, sales, or customers are mentioned in advertisements targeted to men. As seen in the previous analysis, these advertisements are less likely to use words referring to customer service. Availability has a positive coefficient for all three websites and for both genders: the use of words in this category is associated with salaries that are 5.6–11.3 percent higher. In advertisements directed at women, the mention of driver's licenses is associated with an 8.1–10.9 percent increase in salary but has no effect in advertisements directed at men.

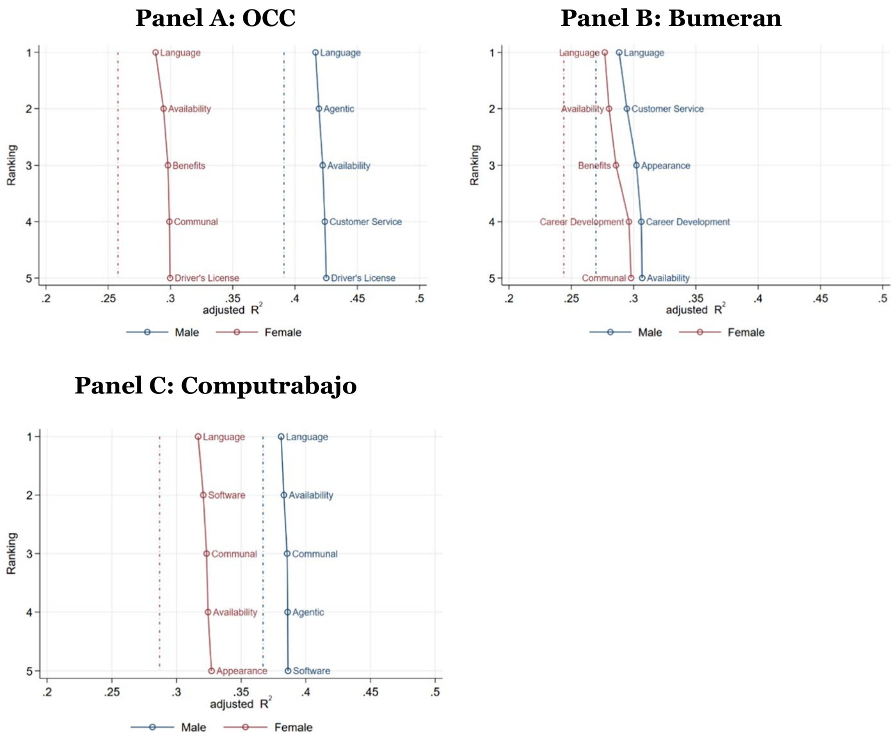


Fig. 6 Goodness-of-fit analysis. Note: Dependent variable is ln(salary). All regressions are controlled for sociodemographic characteristics, which include location, marital status, type of job, educational level, experience, and minimum and maximum age. Ranking explains the importance of the word in terms of its adjusted *R*-squared. Ranking 1 explains the most; ranking 5 the least. The dotted line shows the *R*-squared for demographic characteristics only

In sum, characteristics such as knowledge of English and computer software are valued with a wage premium for both genders. Employers seem to value characteristics such as appearance and availability only in female workers. Characteristics relating to customer service are important in advertisements directed at men. Communal characteristics are penalized in the job market.

Finally, an analysis of adjusted *R*-squared is performed to determine which characteristics explain most of the variance in the wage specified. The LASSO regression examines which factors have the greatest positive effect on gender-targeted job advertisements and which are the most significant for each gender. The objective of the adjusted *R*-squared analysis is to rank characteristics in order of importance in explaining the salary specified in the ads.

A LASSO regression is performed as above. The penalization parameter is varied to obtain the top five characteristics responsible for the variation in salary: Fig. 6 shows the order of importance. The y-axis of the graphs is the ranking, which is the number of variables selected by the regression after increasing the penalization parameter. The x-axis represents the goodness-of-fit in terms of adjusted *R*-squared.

We include a horizontal dotted line in the figure that shows the adjusted *R*-squared for demographic characteristics only. Using this dotted line, we can see that the initial goodness-of-fit is greater for men than for women, because demographic controls better capture the variation in salary for men. However, when other characteristics are included, the increase in adjusted *R*-squared is greater for women. This means that these specific characteristics explain slightly more of the variation in salary for women than for men. The inclusion in the regression of the five characteristics shown explains almost all the variation found in the optimal LASSO regression shown in Table 2. For example, the adjusted *R*-squared for men in OCC Mundial with five categories is 0.425, compared to 0.429 in the model with all characteristics (see Table 2).

The most important category for both genders on all three websites is Language. Speaking English is the key characteristic that explains most of the variation in salary. As this characteristic is more prevalent in advertisements directed at women, the benefit to them is greater. Another frequently used category for women is Appearance, which ranks among the top five on all three websites. This result is consistent with the LASSO analysis, which also found that Appearance was an important characteristic explaining the salary specified in ads directed at women; for men, Appearance is important only on one of the websites. Availability is another important characteristic on all three websites explaining salary variation in ads directed at women; for men, it is important only on two.

In sum, controlling for sociodemographic variables, one of the most important characteristics that explains the variation in salary is language, specifically English. Some of the characteristics most frequently mentioned for both genders are associated with higher salaries. Appearance, especially for women, explains more of the variation in salary than other characteristics.

Out-of-Sample Predictions

Approximately 10 percent of ads are gender-targeted. However, it is possible that other ads are targeting women using words expressing gender stereotypes. The goal of this analysis is to determine whether ads that do not explicitly target men or women can be classified as doing so. We use a random forest (RF) model[16] to find the most important words and characteristics in job advertisements to predict whether those ads are targeted at women or men (a similar approach in a different context is implemented in Oster 2018). We use the prediction model to define a threshold to determine which ads are targeted.

We divide the sample into ads that are explicitly targeted at men, ads that are explicitly targeted at women, and ads that are not explicitly gender-targeted. We use a randomly chosen three-fourths of the gender-targeted samples as training samples

[16] An additional analysis with a LASSO model is also included in the Supplementary Materials (Figure S8, Tables S11-S14).

Table 3 Random Forest Confusion Matrix

			Predicted Value			
			Male	Female	Missing	Total
Actual Value	OCC Mundial	Male	69.25	8.19	22.56	14,515
		Female	6.66	69.46	23.88	14,048
	Bumeran	Male	94.95	1.17	3.88	5,643
		Female	0.88	95.07	4.05	6,045
	CompuTrabajo	Male	78.82	4.79	16.39	10,217
		Female	6.30	72.10	21.60	8,355

Data include all ads that explicitly discriminate, including test and training samples

Table 4 Out-of-Sample Prediction (Non-Gender-Targeted Ads)

	Out-of-Sample Prediction			
	Male	Female	Missing	Total
OCC Mundial	24.95	30.52	44.53	1,446,467
Bumeran	11.10	38.28	50.62	508,842
CompuTrabajo	23.37	29.26	47.37	448,159

Data include all ads that do not explicitly discriminate

for the models, and then make predictions using the remaining fourth. Then we use the model to predict the implicitly targeted gender of the third group.[17] To predict the implicit gender target of the ad, we use dummy variables for selected words from Table 1 and the 50 most frequent words presented in Section I. We choose a random forest model because it can detect nonlinearities and interactions between variables (Chen et al. 2019). The final model is the one with the lowest prediction error for the test sample.

We use RF regression instead of RF classification to obtain a prediction parameter between 0 and 1. This allows us to obtain a mean probability of implicit gender targeting for each ad and thus select the desired threshold for classification. We obtain the gender-targeted ad prediction for the non-explicitly gender-targeted sample by defining thresholds to ensure that we do not take into account ads that are targeted to both genders: a prediction parameter of 0–0.33 is considered implicitly targeted to men, 0.34–0.66 a missing value (both genders), and 0.67–1 implicitly targeted to women.

Table 3 shows the RF confusion matrix for the explicitly discriminating test sample. This table is useful for analyzing false positives and false negatives

[17] This algorithm was implemented in Python with sklearn, and random state 12,345 with 200 trees, to create the ensemble. This model is an ensemble learning method composed of decision trees. An ensemble method combines predictions for each outcome of each tree. The outcome of random forest regression models is the mean prediction (regression) of the individual trees.

Table 5 Examples of Neutral and Explicitly and Implicitly Gender-Targeted Ads

Gender	Explicit /Implicit	Job Ad
Male	Explicit	• Position: **Engineer (temporary)** • **Education:** Software Technician Certificate (or related coursework) • Salary: $4,000 gross, monthly + **benefits** required by law + allowances • *Schedule: Monday*-Friday, 2–7 p.m • **Sex: male** • **Age:** 20–35 years old • **Marital status: any** • **Skills:** Windows and Microsoft office software, software and hardware maintenance. Respond to service requests when required. Network installation. **Minimum experience** 1 year on related projects. **Availability** • Requirements: enthusiastic, **responsible, committed**, disciplined, goal-oriented, pro-active, ability to work in a **team** • **To do or Main responsibilities:** installation and configuration of Lexmark printing equipment • Comply in a timely manner with the delivery of service orders, and document that each piece of equipment is maintained. Maintain reports, logs, and checklists, and manage the project assets assigned. Maintain the **minimum** necessary productivity of facilities so that the position is self-sustaining. Apply and follow general project policies
	Implicit	• Age: 30–50 years • Languages: English • Years of **experience**: 2 years • Academic preparation: industrial engineering • **Skills**: quality system, problem solving analysis, process control, control plan, FMEA, plastic injection, IATF 16,949 update certificate, certified internal auditor, ISO 14000, vda 6.3. Coordinate, carry out, and troubleshoot first-, second-, and third-party audits, maintain quality management system documentation

Table 5 (continued)

Gender	Explicit /Implicit	Job Ad
		• Requirements: goal-oriented, **leadership, initiative, decision making, responsibility,** and **commitment**, communication, **teamwork**, and ability to work **under pressure**, planning and execution, analytical skills, negotiation, conflict resolution, **availability**, *punctuality*
Female	Explicit	• [*Details and location of the company*]
		• Requirements: 2 years of experience in *sales*, preferably related to IT
		• Education: minimum *high school diploma* or technical certificate
		• *Sex: female* (preferably)
		• Age: 20–30 years
		• Experience preferred in computing or IT. Intermediate command of advanced office software. *Good appearance* (essential). *Communication skills*, **teamwork,** goal-oriented, **responsible**, *courteous*
		• Activities: prospecting new clients, client portfolio development of sales strategies (promotions and offers), *follow-up* with distributors, placing orders and follow-up, updating product catalogs
		• We offer attractive *base salary, commissions, benefits required by law*, opportunity for advancement
		• Resume with *photo* (essential)
	Implicit	• Financial advisor. We are seeking a talented person who seeks exponential growth and is passionate about their work to join our select group of financial advisers, with economic and professional development in the advice and marketing of banking products, insurance, and retirement accounts
		• We offer:
		◦Excellent income from unlimited commissions and bonuses, *training* and professional *development*, coaching and personalized career plan, client portfolio, national and international meetings, membership in a dynamic, specialized, and successful team
		• Requirements:
		◦Age 25 to 58 years, *good appearance*, experience (preferably in banking), and interested in sales; *proactive*, dynamic and hardworking, *helpful*, motivated, high earnings potential, college degree or *high school* diploma
		• Residing in or near…

Table 5 (continued)

Gender	Explicit /Implicit	Job Ad
		• If you meet the requirements send resume with a <u>*photograph*</u> to the emails above
		• Location: …
		• Contact: …
Neutral		Requirements:
		• Minimum education: high school diploma
		• **Minimum experience**: 1 year in commercial insurance for heavy equipment (trucks, tractors) required
		• Activities:
		◦Prospecting clients: 80% field, 20% office
		◦Consulting, quotes, and sale of insurance policies for tractors, trucks, forklifts
		◦Portfolio management and monitoring
		• Work location: …
		• <u>*Schedule: Monday*</u> to Friday, 9:00 a.m. to 7:00 p.m
		• Compensation:—$14,000 monthly gross plus commissions, benefits required by law
		Apply using this platform and send your current resume to the contact email, care of…, indicating the position and salary requirements

Identifying information has been redacted. Supplementary Material, Table S17 shows the original ads in Spanish. Male-related words are highlighted in bold and female-related words in italics and underlined and has two axes. One axis is the predicted value, obtained with the RF algorithm, and the other is the actual value, based on explicit gender targeting in an advertisement. False assignments (predicting targeting of men when the ad actually targets women, or vice versa) are less than 9 percent of the total observations for the three websites. The proposed algorithm is better at predicting ads implicitly targeted at women, with fewer false predictions than for ads targeted at men.

Table 4 analyzes the gender of ads not in the sample, that is, of non-gender-targeted ads. Approximately 50 percent of ads in this sample refer to both genders (the Missing category). Also, ads targeted at women are more identifiable. Conditional on being implicitly gender-targeted, more than half of the advertisements on all three websites can be identified as targeting women: 55 percent on OCC Mundial, 77.5 percent on Bumeran, and 55.6 percent on CompuTrabajo.

For a qualitative assessment of the differences between explicitly and implicitly gender-targeted advertisements, we offer some examples of each category in Table 5. The table highlights in bold the words that are usually associated with advertisements targeted at men, and in italics those targeted at women. In general, explicitly gender-targeted ads use similar words as their implicit counterparts. For example, ads targeted at men use words like *any marital status*, *availability*, *teamwork*, and *committed*. Those targeted at women use words like *good appearance*, *photograph*, *commissions*, *courteous*, and *helpful*. Neutral ads use very few gender-targeted words.

To identify the characteristics intended to target a specific gender, we analyze the out-of-sample descriptive statistics predicted by the random forest model. Table 6 shows the ratios of percentage of occurrence of each characteristic of women relative to men (complete descriptive statistics can be found in the Supplementary Material, Table S8); it has the same structure as Table 1. The ratios between models are similar.

In Table 6 the ratio of the characteristics targeting women is 1.25–3.45 times greater than that targeting men. The proportion of job advertisements specifying salary is also similar for women and men. Interestingly, the magnitudes of the ratios shown are generally in agreement with the previous analyses. When marital status is specified, there is a preference for married men and single women, but the number of ads with this specification is small. There is a difference in the level of education specified for women and men: ads targeted at women ask for students, high school graduates, college graduates, or those with unfinished education, but those targeted at men ask for junior high school graduates, those with technical certificates, or engineers. There are specific characteristics for each gender: women are asked to be attentive, helpful, and to have a good appearance; men are asked to be honest, to have the ability for teamwork, and to be available to travel. The type of compensation is also different: women are more frequently offered commissions.

To understand the differences between explicit and implicit targeting of ads, we repeat the analysis of ratios and asymmetric effects for the out-of-sample data. Figure 7 repeats the analysis of the ratio of frequencies of the most commonly used

Table 6 Out-of-Sample Ratio Analysis

		OCC	Bum	CT
Demographic Characteristics	Number of observations	1.22	3.45	1.25
	Includes salary	1.00	0.92	0.91
	Mean real salary	0.83	0.74	0.92
	Includes age	1.10	0.79	0.96
	Minimum age	0.90	0.91	0.95
	Maximum age	0.89	0.96	0.92
	Includes location	1.00	1.01	0.95
	Experience	1.10	1.01	0.95
	Married	0.04	0.17	0.02
	Single	2.52	7.44	3.79
	Student	4.47	1.80	2.64
	Junior high school	0.14	0.15	0.14
	High school	1.94	2.04	1.99
	Some school	3.19	2.43	2.84
	Technician	0.23	0.12	0.17
	Bachelor's degree	3.21	2.32	1.98
	Engineering degree	0.10	0.07	0.04
Communal	Commitment	0.42	1.30	1.19
	Punctual	0.93	1.25	0.98
	Honest	0.91	0.58	0.54
	Attentive	4.51	2.63	4.37
	Teamwork	0.56	1.03	0.76
	Helpful	2.08	1.80	1.37
	Courteous	3.19	1.20	2.98
	Enthusiasm	1.77	3.20	1.60
Agentic	Control	0.71	0.55	0.64
	Initiative	1.33	1.25	1.20
	Motivation	1.17	1.92	1.77
	Pressure	1.12	1.01	0.86
	Proactive	1.06	0.88	0.83
	Responsible	1.55	1.69	1.13
Appearance	Requests photograph	2.13	1.26	2.24
	Requests good appearance	6.92	7.01	11.69
Language	English	1.20	1.27	1.15
	Common computer software	1.77	1.18	1.12
Customer service	Sales	1.98	2.54	2.14
	Customer	1.99	2.61	3.39
	Follow-up	2.33	1.70	2.69
Availability	Availability	1.42	0.54	1.01
	Travel	0.18	0.23	0.26
	Driver's license	0.05	0.01	0.04
Career	Growth	2.23	1.42	1.16

Table 6 (continued)

		OCC	Bum	CT
	Development	2.73	3.26	2.65
	Training	2.29	2.18	1.60
Benefits	Bonus	1.69	1.91	1.71
	Benefits	0.94	0.90	0.96
	Insurance	1.27	0.32	1.64
	Commissions	7.64	4.54	4.31
	Base salary	2.21	1.24	1.36

Summary statistics calculated using data from OCC Mundial, Bumeran, and CompuTrabajo. Each word is represented by a dummy variable equal to 1 if the word appears in the job description or 0 otherwise. See Table S1 for the Spanish words used

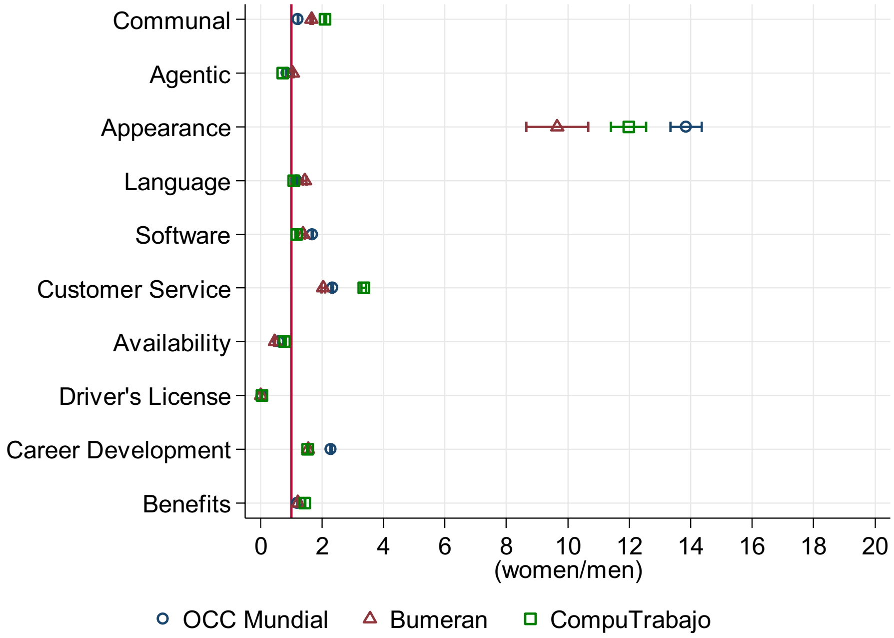


Fig. 7 Ratio of frequencies of most commonly used categories (Replication of Fig. 3 using out-of-sample prediction only). Note: Lines represent 95 percent confidence intervals calculated with 200 random bootstrap samples. Adjacent points indicate similar word roots with different endings. Each point represents the ratio between the proportions of category occurrence for each website. Stop words and punctuation are omitted. Communal category includes the words *commitment*, *punctual*, *honest*, *attentive*, *teamwork*, *helpful*, *courteous*, and *enthusiasm*; Agentic category includes the words *control*, *initiative*, *motivation*, *pressure*, *proactive*, and *responsible*; Appearance category includes the words *photograph* and *appearance*. The Language category includes the word *English*; Software category includes the words *Excel*, *Word*, and *Windows*; Customer Service category includes the words *sales*, *customer*, and *follow-up;* Availability category includes the words *time* and *travel*; Career category includes the words *growth*, *development*, and *training*; Benefits category includes the words *bonus*, *benefits*, *insurance*, *commissions*, and *base salary*

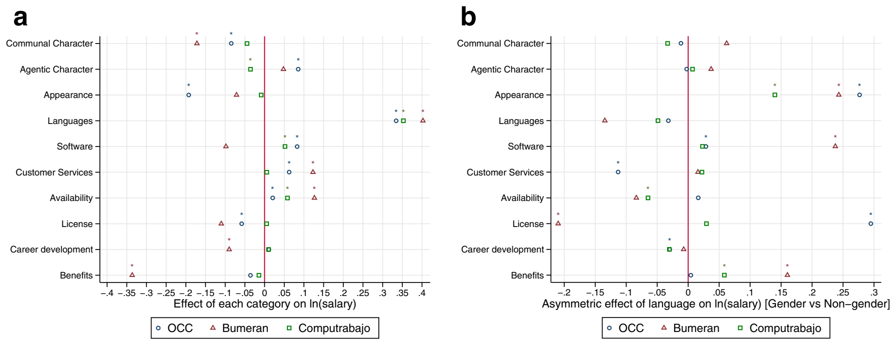


Fig. 8 Effect of categories on salary (Replication of Fig. 4 using out-of-sample prediction only). Note: Regression includes all non-gender-targeted advertisements from the three websites in the period analyzed. The dependent variable is the ln of real salary (July 2018 = 100). Each category is a dummy variable mean that represents the appearance of at least one component word in the advertisement. In Panel A, each point represents the coefficient of the ln(salary) in this dummy. In Panel B, each point represents the coefficient of the ln(salary) in the interaction of a dummy variable for gender-targeted with the dummy for presence of the category. The regression controls for the presence of the category and specifying a gender, state, age (minimum and maximum), experience, educational level. Asterisk (*) signifies that the regression is statistically significant at 5 percent; robust standard errors are clustered by location. Communal category includes the words *commitment*, *punctual*, *honest*, *attentive*, *teamwork*, *helpful*, *courteous*, and *enthusiasm*; Agentic category includes the words *control*, *initiative*, *motivation*, *pressure*, *proactive*, and *responsible*; Appearance includes the words *photograph* and *appearance*; Language includes the word *English*; Software includes the words *Excel*, *Word*, and *Windows*; Customer Service includes the words *sales*, *customer*, and *follow-up*; Availability includes the words *time* and *travel*; Career includes the words *growth*, *development*, and *training*; Benefits includes the words *bonus*, *benefits*, *Insurance*, *commissions*, and *base salary*

words (Sect. 3.1). The patterns observed in explicitly-targeted ads (in Fig. 3) are also seen in implicitly-targeted ads (Fig. 7). The Communal, Appearance, Customer Service, and Career Development categories are more frequent in ads targeted to women. Availability, Agentic, and Driver's License are more frequent in ads targeted to men. However, some categories, such as Appearance, Customer Service, and Career Development, show higher ratios than in the original sample.

Figure 8 shows the asymmetric effects and the effects of each category on salary, using observations from the sample prediction only. In general, the effects of each category on salary mimic those in the original sample. However, only a few asymmetric effects remain statistically different from zero in the three datasets. A striking example of this is the Appearance category, which shows a greater asymmetric effect for gender-targeted ads.

Surprisingly, even though the frequencies of some characteristics are similar to those in the analysis of explicit discrimination, the gender salary gap is greater in job ads that implicitly discriminate. The ratio of the mean salary of women to men is 0.74–0.92; in gender-targeted ads the ratio is 0.88–1.01.

Figure 9 shows the gender salary gap for explicitly gender-targeted job advertisements and the implicitly gender-targeted ads we identified using the random

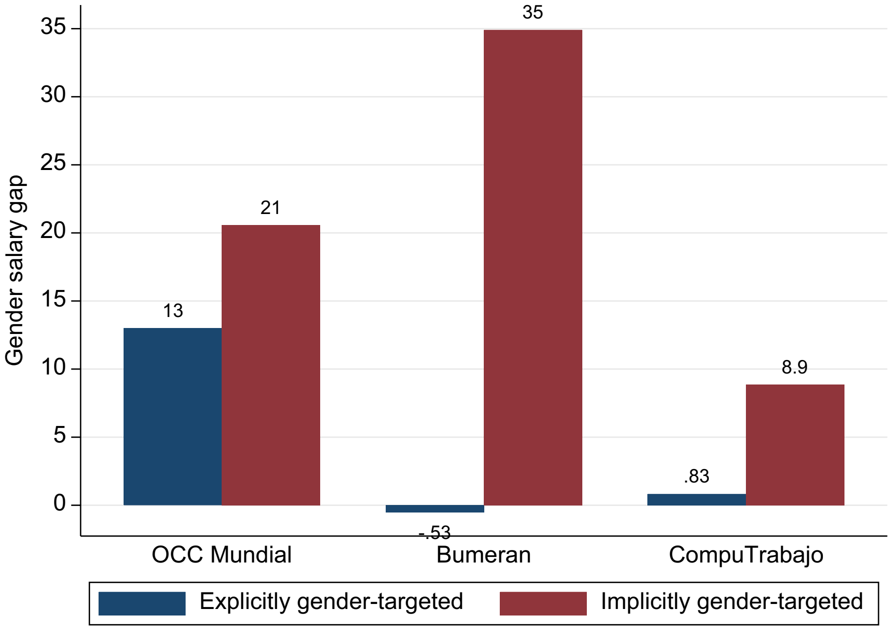


Fig. 9 Gender Salary Gap in Explicitly and Implicitly Gender-Targeted Job Advertisements. Notes: Gender salary gap is the difference between the mean salary in ads explicitly or implicitly targeted at women and the salary in ads explicitly or implicitly targeted at men. "Implicitly targeted" refers to prediction of targeted gender using the random forest algorithm

forest algorithm.'[18][19] This result is conditional to salary appearing in the ad and is a lower bound of the estimation. This gap could increase if we considered ads that do not include salary, since gender-discriminating companies could self-select to not including salary information. Moreover, we might expect the actual gap to be larger once applicants take the jobs, because women have been found to be less likely to negotiate compensation than men, especially when there is no explicit statement that salary is negotiable (Leibbrandt and List 2015).[20]

[18] In figure S5, we restrict the comparison to the ads classified by the algorithm as male or female. That is, we use only the ads with predicted probability less than 0.33 for men and greater than 0.66 for women in the gender-targeted and non-gender-targeted ads. The results are similar to those presented here. Moreover, Figures S9-S10 and Tables S15-S16 replicate the main figures and tables in the paper for the non-explicitly gender-targeted sample that is identified as implicitly targeted by the random forest algorithm.

[19] Figure S11 shows the salary gap with another definition of implicit gender-targeted ads. The figure exploits the spanish gender inflection of nouns in the announcement to classify implicit requests for women. The results show that the salary gap is greater in implicitly targeted ads in OCC Mundial and CompuTrabajo, but not in Bumeran.

[20] The heterogeneity in results can be driven by the heterogeneity in salary advertisement, since reporting salary is not random. In fact, Table S15 shows that one of the most important variables in predicting gender in job ads is the inclusion of salary: it is among the five most important random forest results. Further research is necessary to know how this self-selection in salary advertisement affects the gender salary gap.

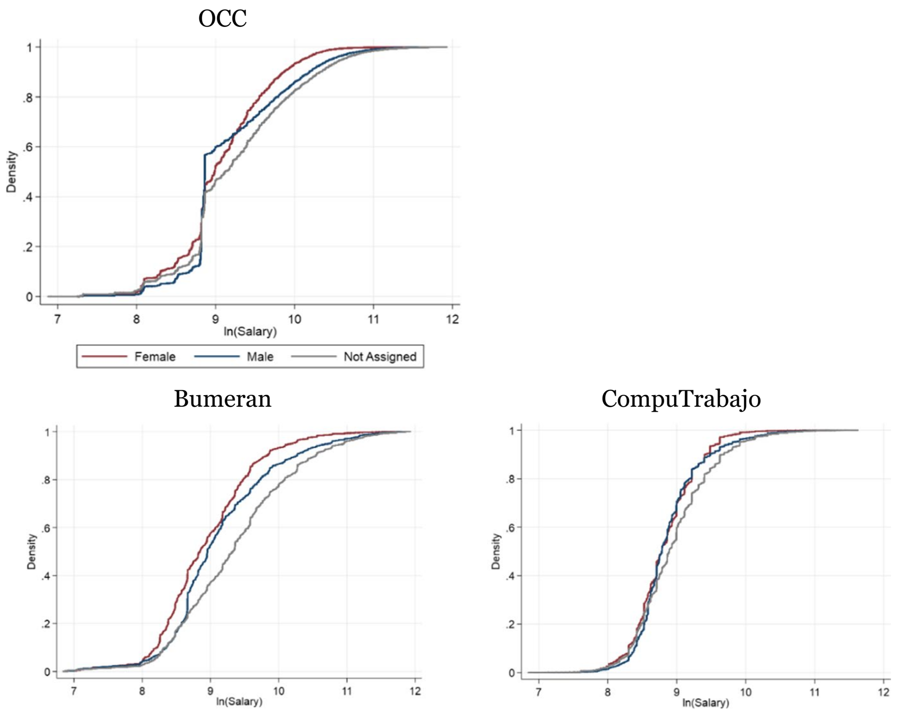


Fig. 10 Cumulative Density Functions of Salary by Predicted Gender Target, Using Random Forest Algorithm. Note: Sample includes advertisements predicted to be gender-targeted

Using our prediction of implicit gender targeting, we can calculate the gender salary distribution in non-targeted ads. Figure 10 includes the cumulative distribution function of the (ln) salary for the three samples of non-gendered-targeted ads. It shows that ads directed at women are more likely to specify a lower salary everywhere except in the middle of the distribution.

Explicitly gender-targeted ads specify characteristics associated with gender stereotypes. We examine words describing these characteristics to predict the implicit targeting of men or women in job ads that do not explicitly discriminate. This implicit targeting implies that employers are thinking about characteristics associated with gender stereotypes in describing the positions advertised. We find that implicit discrimination occurs in almost half of the job advertisements that do not explicitly discriminate, that this discrimination targets women over men, and that the salary gap in these advertisements is greater than that in explicitly discriminatory job ads.

4. Discussion

Explaining the gender wage gap through job advertisements is important for the identification of patterns that systematically pay female workers less than their male counterparts. Although explicit gender discrimination is present in the

Mexican context, it can only explain a small fraction of the gender wage gap: according to our results, the gender wage gap in explicitly discriminatory ads is at most 13 percent. We therefore argue that there is a more subtle mechanism through which potential employers implicitly signal the gender of their desired candidates.

In order to explore this mechanism, we infer the implicit gender discrimination on three different employment search websites in Mexico. We begin by studying the content of the gendered job advertisements, analyzing how the frequency of the most used words, bigrams, and categories changes according to the sex specified in the ad. We document various patterns: for example, when advertisements explicitly targeted at women tend to be more specific regarding the working hours, refer more often to the physical appearance of the candidate, and include communal or customer-service related terms.

We then use an OLS model to determine whether the words, bigrams, and categories already noted have a different effect on salary when they are used in gender-specific ads. The regression estimates indicate that some words and categories are associated with a wage penalty, while others are associated with a premium in gender-targeted ads. Thus, specifying gender does seem to be associated with other important cues that determine salary.

We then use a LASSO estimation to determine whether these words have different effects for men and women. We find important differences, such as the fact that characteristics related to appearance and ability are more valued in women, but we also find some characteristics valued for both genders, such as having English and computing skills. An R^2 analysis confirms that knowing English is the most important determinant of a higher salary for both genders.

Finally, we examine how these differences could affect the gender wage gap, by extending our analysis to the rest of the sample. We exploit the gender-differentiated patterns in the gender-specific job ads, using a random forest model that allows us to identify the non-explicitly gendered ads in our sample as directed to women, directed to men, or neutral. With this amplified sample, we find gender wage gaps of up to 35 percent. We argue that this implicit discrimination is an even more important factor in the gender wage gap in the Mexican labor market.

There are several limitations to our study that could be addressed in future research. The results presented here could be extended by using the random forest algorithm to examine how job seekers respond to implicitly-targeted ads as compared to those that are gender neutral. This analysis would serve as a robustness check on both the out-of-sample predicted values and on the mechanisms generally employed in this paper. We are also aware that our sample includes only online job advertisements. There is thus an obvious selection bias in our data, and our results are not representative of the Mexican labor market as a whole. Future research could shed light on how the skills and characteristics described in these ads are valued outside the online employment market, and how explicitly and implicitly gender-targeted ads contribute to a more general gender wage gap.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12122-022-09331-4.

Declarations

Conflict of interest The authors declare that they have no known financial interests or personal relationships that have influenced the work reported in this paper. Badillo states that the opinions expressed herein do not necessarily represent those of the Banco de Mexico or its Board. We thank an anonymous reviewer and the editor for their constructive comments, which helped us to to improve the manuscript.

References

Abele AE, Wojciszke B (2014) Communal and Agentic Content in Social Cognition. In A. Press (Ed.), Advances in Experimental Social Psychology (50th ed., pp. 195–255). https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Aguilar-Gomez S, Arceo-Gomez E, De la Cruz Toledo E (2019) Inside the Black Box of Child Penalties. In SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3497089

Andresen M, Nix E (2019) What Causes the Child Penalty? Evidence from Same Sex Couples and Policy Reforms. In Statistics Norway (No. 902). Retrieved from https://www.ssb.no/en/forskning/discussion-papers/_attachment/382466. Accessed June 2021

Angelov N, Johansson P, Lindahl E (2016) Parenthood and the Gender Gap in Pay. J Law Econ 34(3):545–579. https://doi.org/10.1086/684851

Arceo-Gómez EO, Campos-Vázquez RM (2014) Evolución de la brecha salarial de género en México. El Trimestre Económico, 81(323), 619–653. https://doi.org/10.20430/ete.v81i323.125

Arceo-Gomez EO, Campos-Vazquez RM (2019) Double Discrimination: Is Discrimination in Job Ads Accompanied by Discrimination in Callbacks? J Econ Race Policy 2(4):257–268. https://doi.org/10.1007/s41996-019-00031-3

Arceo-Gomez EO, Campos-Vazquez RM (2019b) Gender stereotypes: The case of MisProfesores.com in Mexico. Econ Educ Rev 72:55–65. https://doi.org/10.1016/j.econedurev.2019.05.007

Asch SE (1946) Forming impressions of personality. Psychol Sci Public Interest 41(3):258–290. https://doi.org/10.1037/h0055756

Banfi S, Choi S, Villena-Roldan B (2019) Sorting On-Line and On-Time. In SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3323068

Berniell L, Mata D De, Edo M, Marchionni M (2019) Gender Gaps in Labor Informality : The Motherhood Effect. *Documentos de Trabajo Del CEDLAS*, *no. 247*(Issue 247), 1–41. Retrieved from http://sedici.unlp.edu.ar/handle/10915/77486. Accessed June 2021

Blau FD, Kahn LM (2017) The Gender Wage Gap: Extent, Trends, and Explanations. J Econ Lit 55(3):789–865. https://doi.org/10.1257/jel.20160995

Born MP, Taris TW (2010) The Impact of the Wording of Employment Advertisements on Students' Inclination to Apply for a Job. J Soc Psychol 150(5):485–502. https://doi.org/10.1080/00224540903365422

Brenčič V (2012) Wage posting: evidence from job ads. Can J Econ 45(4):1529–1559. https://doi.org/10.1111/j.1540-5982.2012.01738.x

Brown J, Matsa DA (2016) Boarding a Sinking Ship? An Investigation of Job Applications to Distressed Firms. J Financ 71(2):507–550. https://doi.org/10.1111/jofi.12367

Chen J, de Hoogh K, Gulliver J, Hoffmann B, Hertel O, Ketzel M, Hoek G (2019) A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. Environ Int 130(1):104934. https://doi.org/10.1016/j.envint.2019.104934

Chowdhury AR, Areias AC, Imaizumi S, Nomura S, Yamauchi F (2018) Reflections of Employers' Gender Preferences in Job Ads in India: An Analysis of Online Job Portal Data. In Policy Research Working Paper 8379, World Bank. https://doi.org/10.1596/1813-9450-8379

Delgado Helleseter M, Kuhn P, Shen K (2018) The Age Twist in Employers' Gender Requests: Evidence from Four Job Boards. Journal of Human Resources, 0416–7836. https://doi.org/10.3368/jhr.55.3.0416-7836R2

Deming D, Kahn LB (2018) Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. J Law Econ 36(S1):337–369. https://doi.org/10.1086/694106

Ellemers N (2018) Gender Stereotypes. Annu Rev Psychol 69(1):275–298. https://doi.org/10.1146/annurev-psych-122216-011719
Faberman, J. R., and Kudlyak, M. (2016). What Does Online Job Search Tell Us about the Labor Market? *Economic Perspectives*, *40*(1), 15. Retrieved from https://www.chicagofed.org/publications/economic-perspectives/2016/1-faberman-kudlyak. Accessed June 2021
Flory J, Leibbrandt A, List J (2015) Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions. Rev Econ Stud 82(1):122–155. https://doi.org/10.1093/restud/rdu030
Gaucher D, Friesen J, Kay AC (2011) Evidence that gendered wording in job advertisements exists and sustains gender inequality. J Pers Soc Psychol 101(1):109–128. https://doi.org/10.1037/a0022530
Gneezy U, Leonard KL, List JA (2009) Gender Differences in Competition: Evidence From a Matrilineal and a Patriarchal Society. Econometrica 77(5):1637–1664. https://doi.org/10.3982/ecta6690
Heckman JJ, Kautz T (2012) Hard evidence on soft skills. Labour Econ 19(4):451–464. https://doi.org/10.1016/j.labeco.2012.05.014
Hershbein B, Kahn LB (2018) Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings. American Economic Review 108(7):1737–1772. https://doi.org/10.1257/aer.20161570
Hoyos, A., and Ñopo, H. (2010). Evolution of Gender Gaps in Latin America at the Turn of the Twentieth Century: An Addendum to "New Century, Old Disparities." IADB Working Paper 176, Inter-American Development Bank. https://doi.org/10.2139/ssrn.1817290
Kleven H, Landais C, Posch J, Steinhauer A, Zweimüller J (2019a) Child Penalties across Countries: Evidence and Explanations. AEA Papers and Proceedings 109:122–126. https://doi.org/10.1257/pandp.20191078
Kleven H, Landais C, Søgaard JE (2019b) Children and Gender Inequality: Evidence from Denmark. Am Econ J Appl Econ 11(4):181–209. https://doi.org/10.1257/app.20180010
Kuhn, P., and Shen, K. (2013a). Do Chinese Employers Avoid Hiring Overqualified Workers? Evidence from an Internet Job Board. In *Research in Labor Economics*: *Vol. 37*. *Labor Market Issues in China* (pp. 1–30). w
Kuhn P, Shen K (2013b) Gender Discrimination in Job Ads: Evidence from China. Q J Econ 128(1):287–336. https://doi.org/10.1093/qje/qjs046
Kuhn P, Shen K (2015) Do employers prefer migrant workers? Evidence from a Chinese job board. IZA Journal of Labor Economics 4(1):22. https://doi.org/10.1186/s40172-015-0038-0
Kuziemko I, Pan J, Shen J, Washington E (2018) The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood? (No. 24740). https://doi.org/10.3386/w24740
Leibbrandt A, List J (2018. Do Equal Employment Opportunity Statements Backfire? Evidence From A Natural Field Experiment On Job-Entry Decisions (No. 25035). https://doi.org/10.3386/w25035
Leibbrandt A, List JA (2015) Do Women Avoid Salary Negotiations? Evidence from a Large-Scale Natural Field Experiment. Manage Sci 61(9):2016–2024. https://doi.org/10.1287/mnsc.2014.1994
Marinescu I, Wolthoff R (2019) Opening the black box of the matching function: The power of words. Journal of Labor Economics. https://doi.org/10.1086/705903
Marinescu I, Rathelot R (2018) Mismatch Unemployment and the Geography of Job Search. Am Econ J Macroecon 10(3):42–70. https://doi.org/10.1257/mac.20160312
Modestino AS, Shoag D, Ballance J (2019) Upskilling: Do Employers Demand Greater Skill When Workers are Plentiful? The Review of Economics and Statistics, 1–46. https://doi.org/10.1162/rest_a_00835
Niederle M, Vesterlund L (2007) Do Women Shy Away From Competition? Do Men Compete Too Much? Q J Econ 122(3):1067–1101. https://doi.org/10.1162/qjec.122.3.1067
Niederle M, Vesterlund L (2011) Gender and Competition. Annu Rev Econom 3(1):601–630. https://doi.org/10.1146/annurev-economics-111809-125122
Ñopo H, Daza N, Ramos J (2012) Gender earning gaps around the world: a study of 64 countries. Int J Manpow 33(5):464–513. https://doi.org/10.1108/01437721211253164
Olivetti C, Petrongolo B (2016) The Evolution of Gender Gaps in Industrialized Countries. Annual Review of Economics 8(1):405–434. https://doi.org/10.1146/annurev-economics-080614-115329
Oster E (2018) Diabetes and Diet: Purchasing Behavior Change in Response to Health Information. Am Econ J Appl Econ 10(4):308–348
Rudman LA, Phelan JE (2010) The Effect of Priming Gender Roles on Women's Implicit Gender Beliefs and Career Aspirations. Social Psychology 41(3):192–202. https://doi.org/10.1027/1864-9335/a000027

Saucier G (2009) What Are the Most Important Dimensions of Personality?Evidence from Studies of Descriptors in Diverse Languages. Soc Pers Psychol Compass 3(4):620–637. https://doi.org/10.1111/j.1751-9004.2009.00188.x

Spence JT, Helmreich R, Stapp J (1975) Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. J Pers Soc Psychol 32(1):29–39. https://doi.org/10.1037/h0076857

World Economic Forum (2019) Global Gender Gap Report 2020: Insight Report. https://doi.org/10.1002/9781119085621.wbefs350