

# UC Irvine

## UC Irvine Previously Published Works

### Title

Does Ageist Language in Job Ads Predict Age Discrimination in Hiring?

### Permalink

<https://escholarship.org/uc/item/0sw052m9>

### Journal

Journal of Labor Economics, 40(3)

### ISSN

0734-306X

### Authors

Burn, Ian

Button, Patrick

Munguia Corella, Luis

et al.

### Publication Date

2022-07-01

### DOI

10.1086/717730

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



# HHS Public Access

Author manuscript

*J Labor Econ.* Author manuscript; available in PMC 2023 July 01.

Published in final edited form as:

*J Labor Econ.* 2022 July ; 40(3): 613–667. doi:10.1086/717730.

## Does Ageist Language in Job Ads Predict Age Discrimination in Hiring?

**Ian Burn,**

Department of Economics, University of Liverpool

**Patrick Button,**

Department of Economics, Tulane University

**Luis Munguia Corella,**

Department of Economics, University of California–Irvine

**David Neumark**

Department of Economics, University of California–Irvine

### Abstract

We study the relationships between ageist stereotypes – as reflected in the language used in job ads – and age discrimination in hiring, exploiting the text of job ads and differences in callbacks to older and younger job applicants from a resume (correspondence study) field experiment (Neumark, Burn, and Button, 2019). Our analysis uses computational linguistics and machine learning methods to examine, in a field-experiment setting, ageist stereotypes that might underlie age discrimination in hiring. In so doing, we develop methods and a framework for analyzing textual data, highlighting the usefulness of various computer science techniques for empirical economics research. We find evidence that language related to stereotypes of older workers sometimes predicts discrimination against older workers. For men, we find evidence that age stereotypes about all three categories we consider – health, personality, and skill – predict age discrimination, and for women, age stereotypes about personality predict age discrimination. In general, the evidence that age stereotypes predict age discrimination is much stronger for men, and our results for men are quite consistent with the industrial psychology literature on age stereotypes.

---

### Introduction

We develop and implement methods to explore the role of stereotypes in hiring discrimination using the text of job ads and apply these methods to evidence on age discrimination. We make two contributions. First, we develop techniques that leverage machine learning and textual analysis to analyze the text data in job ads from a large-scale field experiment on discrimination. Second, we use these techniques to produce evidence on which age-related stereotypes appearing in job ads are associated with an experimental measure of hiring discrimination against older workers – the first evidence we know of that establishes relationships between age-related stereotypes and actual employer behavior. This

---

analysis provides evidence on whether employers who use ageist language in their job ads also have less intent to hire older workers – as captured in our experimental results.

The hypothesis underlying this analysis is that employers might use stereotyped language in job ads to shape the applicant pool, discouraging older workers from applying and reducing the likelihood that age discrimination is detected. Alternatively, different jobs may have different requirements, which could be stated in job ads, and employers may hold stereotypes about older job applicants' abilities to meet these job requirements – for example, assuming that older workers are less able to do heavy lifting. In either case, we might expect that employers who use job-ad language expressing ageist stereotypes are more likely to engage in age discrimination in hiring, with the latter case representing pure statistical discrimination.<sup>1</sup>

Age discrimination is of great policy interest to the United States and other countries. Rapidly aging populations coupled with lower labor force participation of older individuals implies rising dependency ratios and strains on the finances of public programs targeted at older individuals, especially retirement and health care programs. As a result, there is an imperative to increase the employment of older individuals. The hiring of older individuals is likely an essential part of the solution. Nearly half of older workers move to “bridge” jobs or “partial retirement” jobs (Johnson, Kawachi, and Lewis, 2009) before transitioning to complete retirement, and many leave retirement to take jobs before retiring again (Maestas, 2010). Age discrimination may hinder the ability of older individuals to move into new jobs or re-enter the workforce.

The most credible evidence on discrimination in hiring comes from field experiments – more specifically, resume-correspondence studies (Fix and Struyk, 1993; Bertrand and Duflo, 2017; Gaddis, 2018; Neumark, 2018). These studies have been applied to discrimination based on race, ethnicity, gender, age, and other group membership (e.g., disability). Resume-correspondence studies of age discrimination create fictitious but realistic job applicants who are on average equivalent except for age, which is signaled through school graduation year(s). Researchers use these fictitious job applicants to apply for real job openings, and age discrimination in hiring is measured by comparing interview request rates (“callbacks”) between older and younger applicants. Previous resume-correspondence studies almost always point to substantial age discrimination in hiring.<sup>2</sup> Recently, we conducted a large-scale field experiment studying age discrimination in hiring, focusing on potential sources of bias in past studies. We found compelling evidence of age discrimination – especially against older women (Neumark et al., 2019, henceforth NBB).<sup>3</sup>

Our goal in the present paper is to advance the experimental literature on age discrimination in a direction that helps us understand what underlies age discrimination, delving inside

<sup>1</sup>We address the possibility that if employers influence the applicant pool strongly enough, they may not have to hire at lower rates from the older applicants who do apply.

<sup>2</sup>See: Bendick, Jackson, and Romero (1997); Bendick, Brown, and Wall (1999); Riach and Rich (2006, 2010); Lahey (2008); Baert et al. (2016); Farber, Silverman, and von Wachter (2017); Farber et al. (2019); Carlsson and Eriksson (2019); Neumark, Burn, and Button (2016, 2019); and Neumark et al. (2019).

<sup>3</sup>NBB provide an extensive discussion regarding the interpretation of resume-correspondence study findings as reflecting age discrimination. Here, we simply interpret the evidence this way, and refer readers to that paper for discussion of this issue.

the black box of why or how employers discriminate based on age. Specifically, we use the text of job ads from NBB to explore whether – and if so which – age stereotypes are associated with actual discrimination by employers. This inquiry is motivated by research in industrial psychology (and related areas), discussed in detail below, documenting that employers and others have negative stereotypes about older workers – such as lower ability to learn, less adaptability, worse interpersonal skills, less physical ability, lower productivity, worse technological skills and knowledge, and less creativity – all of which can deter their hiring. However, little is known about which stereotypes employers act on, if any, when making actual hiring decisions. The industrial psychology literature primarily uses surveys given to small samples of students or a general population about their attitudes concerning older individuals, but not necessarily in employment contexts, let alone the specific context of older workers seeking new jobs. Although some studies survey managers with hiring experience, in their actual roles as managers they may not use these stereotypes in making decisions. In addition, survey respondents may not honestly reveal discriminatory preferences, stereotypes, or values if they are socially undesirable (e.g., Barnett, 1998; and Krumpal, 2013).

For these reasons, in this paper we pursue evidence on the importance of age-related stereotypes for actual labor market behavior. We provide, to our knowledge, the first study that links age stereotypes to evidence on actual age discrimination in hiring.<sup>4</sup> We use the text from over 11,000 job advertisements from our field experiment, and we explore what job-ad language related to age stereotypes predicts age discrimination in hiring – as measured in the experiment by the younger applicant being called back but not the older applicant. For example, one stereotype of older workers is that they are not as good with technology (McCann and Keaton, 2013). Job ads could contain language related to this stereotype (e.g., “must be a technological native”). We can then ask whether job ads containing such language are less likely to result in callbacks for older job applicants.

We find evidence that language related to stereotypes of older workers often predicts discrimination against older workers, especially for men.<sup>5</sup> For men, we find evidence that age stereotypes about all three categories we consider – health, personality, and skill – predict age discrimination, and for women, age stereotypes about personality predict age discrimination. In general, the evidence is much stronger for men, and our results for men are quite consistent with the industrial psychology literature on age stereotypes.<sup>6</sup>

The machine learning and textual analysis methods we develop can be used to leverage text data in audit and correspondence studies on discrimination in labor markets and in other

---

<sup>4</sup>There are other studies that find a link, albeit less directly, between age discrimination in hiring and age stereotypes. Carlsson and Eriksson (2019) conduct a resume-correspondence study and ask employers about stereotypes, finding that employers in their survey think that older workers have lower ability to learn new tasks, are less flexible/adaptable, and have less ambition. But they do not directly link the hiring outcomes to these survey responses about stereotypes. van Borm, Burn, and Baert (2019) survey recruiters about their beliefs regarding the skills and abilities of hypothetical candidates in a vignette study. They find that employers view older workers as having worse skills, and that the perceived differences in skills explained more of the discrimination against older women than older men that is suggested by the vignette study.

<sup>5</sup>Of course, our methods do not speak to the role of stereotypes held by employers that are not reflected in job ads.

<sup>6</sup>As discussed later in the paper, there is some evidence from industrial psychology and related research of stereotypes that are favorable to older workers, and some that stereotypes that can either favor or disfavor them. We discuss the evidence on these stereotypes as well.

markets.<sup>7</sup> These methods build significantly upon earlier research leveraging text data from job ads, applying machine learning methods to the text analysis rather than just searching for key phrases.<sup>8</sup> Our approach allows researchers to analyze text data when the phrasing is complex, varied, and not always obvious (e.g., the numerous ways one could phrase “communication skills”), and can also be applied to a wider range of empirical research questions in economics. We constructed our approach to create an a priori classification of the language that can be developed independently of the analysis of the relationship between the coding of language and the outcomes of interest. Researchers doing future correspondence studies or other types of studies who wish to utilize the text of the ads or other sources of information could pre-register the application of and “output” from this method before collecting the data.

Our evidence on whether employers who use ageist language in their ads are less likely to hire older workers has implications for policy responses to reduce age discrimination. For example, job training, job coaching, or educational campaigns can focus on addressing the relevant negative stereotypes. Efforts could also focus on improving hiring practices, perhaps by increasing the information available to employers that reduces the attribution of stereotypes to older workers. More generally, the Code of Federal Regulations covering the ADEA currently states, “Help wanted notices or advertisements may not contain terms and phrases that limit or deter the employment of older individuals. Notices or advertisements that contain terms such as age 25 to 35, young, college student, recent college graduate, boy, girl, or others of a similar nature violate the Act unless one of the statutory exceptions applies” (§1625.4).<sup>9</sup> Thus, our work can provide information to agencies that enforce age discrimination laws on job-ad language that may predict employer discrimination in hiring.

---

<sup>7</sup>The only other studies that we are aware of that used textual analysis, with machine learning, on job ads are Jaeger et al. (2020) and Tian and Zhang (2021). Jaeger et al. apply machine learning to the text from ads to classify internships into occupation categories. Tian and Zhang apply machine learning to text from ads for library science positions in China to detect explicit discriminatory preferences, including age limits for applicants.

Some correspondence studies use textual data on a more limited basis. Hanson, Hawley, and Taylor (2011) study subtle discrimination through “keywords” used by landlords responding to prospective tenants. Hanson et al. (2016) had research assistants code the helpfulness and other characteristics of mortgage loan originator responses to prospective borrowers. Tilsik (2011) identifies four words in job ads related to masculine stereotypes (decisive, aggressive, assertive, and ambitious) and links them to hiring outcomes in a study of discrimination against gay men. Nunley et al. (2015) uses textual analysis of job titles in ads to identify jobs requiring extensive customer interaction.

Correspondence studies in other markets could also make use of text data. For example, Kugelmass (2019) studies discrimination in access to appointments with mental health professionals, and Ameri et al. (2020) study discrimination in access to AirBnB rentals. Both studies use platforms in which there is text data that could potentially be analyzed.

<sup>8</sup>There are several notable examples of researchers using textual data in job ads outside of a resume correspondence study. Kuhn and Shen (2013) and Kuhn, Shen, and Zhang (2020) explore how gender preferences feature explicitly or implicitly in job ads in China, and Hellester, Kuhn, and Shen (2020) explore age and gender preferences in job ads in China and Mexico. Modestino, Shoag, and Balance (2016) use text data from job ads to document that during the recovery from the Great Recession, “downskilling” occurred, with firms reducing skill requirements in job ads. Deming and Kahn (2018) use text data in job ads to measure how ten different skills relate to wages. Marinescu and Wolthoff (2020) match text data from job ads to job application data to study the matching process between jobs and applicants. Banfi and Villena-Roldán (2019) use unique features of a job board to study how posted wages affect job applicants. Ash et al. (2020) measure gender stereotyped language in judges’ decisions on gender-related issues (e.g., workplace outcomes, sexual harassment). In an early small study, Wax (1948) found that summer resorts in Ontario, Canada, were more likely to discriminate against Jewish customers (based on names) requesting accommodations if they used phrases like “restrictive clientele” in their advertising.

<sup>9</sup>An extreme version of such language is stating maximum experience levels in job ads – as occurred recently in *Kleber v. Carefusion Corp.* – which will clearly act to exclude many older applicants. See *Kleber v. Carefusion Corp.* ([http://www.aarp.org/content/dam/aarp/aarp\\_foundation/litigation/pdf-beg-02-01-2016/kleber-amended-complaint.pdf](http://www.aarp.org/content/dam/aarp/aarp_foundation/litigation/pdf-beg-02-01-2016/kleber-amended-complaint.pdf), viewed November 8, 2017). Button (2020) discusses the ruling in this case.

## A Model of Stereotyped Language in Job Ads and Employer Discrimination

We develop a stylized model to help clarify how to interpret evidence that stereotyped language in job ads is or is not related to age discrimination. This is useful because it is conceivable that older workers' responses to stereotyped job ads could end up obscuring the relationship between stereotyped language and measured discrimination. We are not studying applicants' responses to stereotyped job-ad language because we do not observe these responses.<sup>10</sup> But we consider the implications of this potential response for our empirical analysis.

Let  $h^T$  ( $0 < h^T < 1$ ) index age discrimination in hiring by discriminatory employers.  $h^T$  is the relative hiring rate from the pool of potential old applicants compared to potential young applicants. The  $T$  superscript indicates that  $h^T$  is the "target" ratio of old vs. young employees. Employers achieve their target employment ratio by applying the relative hiring rate  $h$  to the actual applicant pool.<sup>11</sup> Non-discriminating employers have  $h^T = 1$ , so they would hire in proportion to the ratio of old versus young among potential applicants. We assume that non-discriminating employers write age-neutral ads. In contrast, discriminatory employers may write stereotyped ads that might discourage older workers from applying.

If a discriminatory employer runs an age-neutral ad, we assume the old/young ratio among actual applicants equals the ratio among potential applicants. In this case, the employer has to apply a lower relative hiring rate ( $h = h^T$ ) to the older applicants. Alternatively, they can run stereotyped ads. This lowers the relative share of the old among actual vs. potential applicants, allowing employers to hit their target ratio with hiring rate  $h$  closer to 1 than is  $h^T$ . If the response to the stereotyped ads is strong enough, the share of the old among actual applicants could be driven so low that employers do not need to use  $h < 1$  to hit their target ratio. They could even, conceivably, favor older applicants ( $h > 1$ ) if the response to stereotypes is large enough.<sup>12</sup> Thus, given the potential response to stereotypes, there may be a bias against finding discrimination in hiring rates for employers running stereotyped ads; that is, our estimates provide a lower bound. Conversely, however, we cannot get spurious evidence in favor of this hypothesis.

We would expect more potential applicant response to stereotyped ads the higher is the cost of applying. In the (lower) limit, if applying is costless, applicants ignore the stereotyped ad language. Thus, there is less bias the lower the cost of applying. (And we think the cost of applying on our job board is quite low.) We would also expect more potential applicant response when the probability of being hired declines more in response to age-stereotyped language. This has implications for the potential bias in our estimates.

<sup>10</sup>There is research suggesting that, in other contexts, job seekers respond to job-ad language, including: Belot, Kircher, and Muller (2018) and Banfi, Stefano, and Benjamín Villena-Roldán (2019) on posted wages; Kuhn et al. (2020) on gender requests; and Ibanez and Riener (2018), Leibbrandt and List (2018), and Flory et al. (2021) on affirmative action or diversity statements in recruitment materials or job ads.<sup>10</sup>

<sup>11</sup>For purposes of discussion, we equate the callback rate with the hire rate. These may differ in a fixed proportion, so that  $h = \theta \cdot cb$  (where  $h = \text{old/young hire rate}$ ,  $cb = \text{old/young callback rate}$ , and  $0 < \theta < 1$ ). The substance of this is that there is no manipulation of callbacks to young vs. old applicants to achieve hiring goals.

<sup>12</sup>A reviewer raises the possibility of a more complex response that changes the skill composition of older vs. younger applicants. For example, job ads may contain language about communications skills, which older workers perceive as a stereotype, causing those with worse communications skills to apply at a lower rate. It is beyond the ability of our data to test for variation in effects based on skills of applicants that are potentially related to age stereotypes, but this is an interesting question for further research.

To develop these ideas more explicitly, we cast them in the context of our regression model. With non-discriminatory hiring and no shaping of the applicant pool, the ratio of old hires ( $H_O$ ) to young hires ( $H_Y$ ) equals the ratio of old potential applicants ( $PA_O$ ) to young potential applicants ( $PA_Y$ ), because the non-discriminatory hiring rate ( $r = r_Y = r_O$ ) is applied to both older and younger potential applicants, implying:

$$H_O/H_Y = (r \cdot PA_O)/(r \cdot PA_Y) = PA_O/PA_Y \quad (1)$$

Discriminatory firms are defined by the parameter  $h^T$ , such that  $H_O/H_Y = h^T \cdot PA_O/PA_Y$ . There are two tools to achieve this goal: discriminatory hiring and using stereotyped language in the job ad. If a discriminatory firm engages in discriminatory hiring, but job-ad language does not deter older applicants, then  $r_O/r_Y = h^T$ ,  $H_O = h^T \cdot r_O \cdot PA_O$ ,  $H_Y = r_Y \cdot PA_O$ , and  $H_O/H_Y = h^T \cdot PA_O/PA_Y$ .

Stereotyped language in a job ad reduces the actual older applicants ( $AA_O$ ) below  $PA_O$ . For example, suppose the effect of the stereotyped language is to drive  $AA_O$  to equal  $h^T \cdot PA_O$ . In that case, the firm can apply the non-discriminatory hiring rate  $r$  to both groups, and again  $H_O/H_Y = (r \cdot AA_O)/(r \cdot PA_Y) = h^T \cdot PA_O/PA_Y$ .

We estimate models that, at their simplest, are of the form:

$$d_{exp} = \alpha + \beta S + \varepsilon. \quad (2)$$

$S$  is a measure of the “age stereotype” strength of the ad.<sup>13</sup> In our implementation,  $S$  is a continuous variable that runs from  $-1$  to  $1$ , but for what follows, treat it as rescaled so that  $S \geq 0$  (without loss of generality).  $d_{exp}$  is the experimental indicator of discrimination – for most of our analysis, whether the young applicant received a callback and the old one did not. This can be viewed as a job-ad-level measure of  $H_O/H_Y$ .

The question is whether a positive estimate of  $\beta$  indicates that firms that use ageist stereotypes in their job ads are more likely to discriminate. The potential problem is that if stereotyped job ads reduce  $AA_O$  relative to  $PA_O$ , age discrimination in hiring may not be accurately reflected in the estimate of  $\beta$ . When high  $S$  job ads reduce applications from the old, there is an omitted variable  $AA_O/PA_O$ . This omitted variable  $AA_O/PA_O$  is negatively correlated with  $S$ , through the response of older applicants. And it has a negative effect on  $d_{exp}$  because discriminating firms that receive fewer applications from older workers engage in less discrimination against older workers who apply. Thus, the estimate of  $\beta$  is biased downward – *against* finding evidence that firms that use age-stereotyped ads discriminate against older workers.

This bias is stronger the larger is the response of applicants ( $AA_O/PA_O$ ) to  $S$ . We would think this application response is higher the stronger is the relationship between  $S$  and hiring discrimination against older workers. This is helpful because it makes it less likely that

<sup>13</sup>It can be a vector corresponding to many stereotypes, but here we treat it as a single variable.



the bias from the response of older applicants obscures evidence that employers who use age-stereotyped ads discriminate against older workers.

To see this, suppose the utility of job  $j$  to person  $i$  is  $U_{ij} = \varepsilon_{ij}$ , where  $\varepsilon_{ij} \sim N(0,1)$ .<sup>14</sup> The cost of applying for a job is  $c$ . The probability of getting a job offer (callback) is  $p_y = b$  ( $0 < b < 1$ ) if young, and  $p_o = b(S)$  ( $0 < b(S) < b$ ,  $b'(S) < 0$ ) if old. That is, the probability of a job offer is lower the more age-stereotyped the job ad. An example of a function with  $0 < b(S) < b$ , is:

$$b(S) = b \cdot e^{-\eta \cdot S} (\eta > 0). \quad (3)$$

A young person applies if  $b \cdot \varepsilon > c$  (dropping the  $i$  and  $j$  subscripts) or  $\varepsilon < -c/b$ . Given the distributional assumption, the probability of applying is  $A_y = \Phi(-c/b)$ . An old person applies if  $b(S) \cdot \varepsilon > c$  (dropping the  $i$  and  $j$  subscripts) or  $\varepsilon < -c/b(S)$ , so  $A_o = \Phi(-c/b(S))$ .<sup>15</sup>

For old applicants,

$$\partial A_j / \partial S = \frac{-(-c)b'(S)}{b(S)^2} \cdot \phi\left(\frac{-c}{b(S)}\right), \quad (4)$$

where  $\phi$  denotes the standard normal density. Since  $b'(S) < 0$ ,  $\partial A_o / \partial S < 0$ , establishing that there will be a negative response of older applicants to more stereotyped job ads. We would expect  $\partial A_o / \partial S$  to be larger (a larger reduction in  $A_o$ ) the more the probability of an offer declines with  $S$ . To see this, assume the specific functional form  $b(S) = b \cdot e^{-\eta S}$ , so that  $b(S) = b$  if  $S = 0$ , and declines towards zero as  $S$  increases.

In this case, since  $b'(S) = -\eta \cdot b \cdot e^{-\eta S}$ ,

$$\partial A_j / \partial S = \frac{(-c)\eta}{b} \cdot e^{\eta S} \cdot \phi\left(\frac{-c}{b} \cdot e^{\eta S}\right) \quad (5)$$

and

$$\partial^2 A_o / \partial S \partial \eta = \left\{ \left( \frac{-c}{b} \right) \cdot e^{\eta S} - \frac{c}{b} \cdot \eta \cdot S \cdot e^{\eta S} \right\} \phi\left(\frac{-c}{b} \cdot e^{\eta S}\right) - \frac{c}{b} \cdot \eta \cdot e^{\eta S} \cdot \phi'\left(\frac{-c}{b} \cdot e^{\eta S}\right) \quad (6)$$

Since for the standard normal,  $\phi'(x) = -x \cdot \phi(x)$ :

$$\partial^2 A_o / \partial S \partial \eta = \phi\left(\frac{-c}{b} \cdot e^{\eta S}\right) \cdot \left\{ -\frac{c}{b} \cdot e^{\eta S} - \frac{c}{b} \cdot \eta \cdot S \cdot e^{\eta S} - \left(\frac{c}{b}\right)^2 \cdot \eta \cdot (e^{\eta S})^2 \right\}, \quad (7)$$

which is negative because  $b, c, \eta, S > 0$ .

<sup>14</sup>The variance can be fixed without loss of generality.

<sup>15</sup>We assume that  $\partial A_y / \partial S = 0$ , or young people do not respond to the stereotyped language. We could have the probability of an offer for a young applicant *increasing* in  $S$  – i.e., the opposite direction from old people – and the qualitative conclusions are the same.



Thinking back to our regression model (equation (2)),  $\beta$  and  $\eta$  are positively related. When  $\eta$  is larger, the true relationship between age-stereotyped ads and discrimination against older applicants is stronger. The implication of  $\partial \beta / \partial \eta < 0$ , then, is that there will be stronger negative bias in the estimate of  $\beta$  when the true value of  $\beta$  (and hence  $\eta$ ) is larger.

This is significant because it militates against bias from the response of older applicants obscuring evidence that employers who use age-stereotyped ads discriminate. This occurs because when the relationship captured by  $\beta$  (relating measured age discrimination to stereotyped ads) is weaker, the bias from the applicant response is also weaker. This makes sense because the bias comes from potential applicants responding to the likelihood that the stereotyped ad implies it is harder for an older applicant to get hired.

The implication is that a failure to find evidence linking job-ad language to employer discrimination likely reflects something else: (i) methods that are uninformative at detecting ageist stereotypes in job ads; (ii) employers not believing that job seekers respond to age stereotypes in job ads; or (iii) discriminatory employers simply do not use stereotyped language in job ads to discourage older workers from applying. In none of these cases is there an actual link between ageist stereotypes in job ads and employer age discrimination in hiring. Even more important, because there is a bias against finding evidence of such a link, evidence against the null would imply that ageist language in job ads does help predict hiring discrimination against older workers.

## Background and Data from the Resume-Correspondence Study

Our experimental measure of discrimination comes from NBB's large and comprehensive resume-correspondence study of age discrimination. The study used triplets of realistic but fictitious resumes for young (aged 29–31), middle-aged (aged 49–51), and older (aged 64–66) job applicants, sending 40,223 applications (resumes) to 13,371 job positions in 12 U.S. cities (in 11 states). This is by far the most extensive resume correspondence study of hiring discrimination to date, and the large number of job ads included in the study is critical to the methods we use in the present paper.

NBB sent applications for positions in occupations that, based on Current Population Survey data, older as well as younger individuals often take as new jobs (hence likely bridge jobs for older workers): administrative assistant and retail sales jobs for women, and retail sales, security, and janitor jobs for men.<sup>16</sup>

Figure 1 presents the main descriptive evidence from NBB. Across all occupations and genders, older applicants (age 64–66) got fewer callbacks than younger applicants. (These

<sup>16</sup>NBB sent three applications per position: always one younger applicant and two older applicants of different ages (49–51 or 64–66) or with different work experience histories. While some of the resumes sent were on average identical to isolate the effect of age, as in the usual resume-correspondence design, NBB also sent some older worker resumes with more realistic, longer work histories; arguably these applicants are more comparable to the younger applicants because their experience is commensurate with their age – like for young applicants. This was done to avoid the possibility of bias towards finding evidence of age discrimination, as older workers would not normally have the same listed work experience as younger workers. NBB also used different resume types to explore whether older workers who exhibit “bridging” behavior – moving from demanding jobs or jobs with more responsibility to jobs that are more flexible or with less responsibility – experienced more discrimination. Measured discrimination was generally insensitive to these variations.

differences were statistically significant in all cases, except for men applying for security jobs.) As Figure 1 shows, the magnitude of the discrimination against older women was larger. NBB present several more sophisticated analyses, but the basic conclusion remains the same.

## What Does Our Evidence Tell Us About Discrimination?

Why might employers use stereotyped language in job ads, and what might this predict for our analysis? One hypothesis is that employers who discriminate based on age use stereotyped language to shape the applicant pool, reducing the likelihood that age discrimination is detected. In particular, language that conveys positive stereotypes about young workers might discourage older workers from applying (as might language conveying negative stereotypes about older workers – although negative stereotypes are less common in our data). This would lead to the underrepresentation of older applicants in the applicant pool.

Why is this valuable to a discriminating employer? Presumably, the probability of an age discrimination claim (and an adverse outcome for the employer) depends on how much lower the ratio of job offers to applicants is for older applicants than for younger applicants ( $h$ , from the model). Then for the same *number* of older and younger hires, an employer who uses stereotypes that discourage older job applicants would have a lower probability of facing a discrimination claim.<sup>17</sup> Thus, we can test the hypothesis that discriminatory employers use ageist language in job ads by relating the measure of age discrimination in the resume-correspondence study (differences in callbacks for older versus young applicants) to age stereotypes in job-ad language.<sup>18</sup> This hypothesis does not necessarily distinguish between taste and statistical discrimination, but rather just tests whether employers who do not want to hire older workers use stereotypes in job ads to facilitate their discrimination.<sup>19</sup>

A second hypothesis is more closely related to statistical discrimination. Different jobs may have different requirements, which could be stated in job ads. However, employers may hold stereotypes about older job applicants' abilities to meet these job requirements – for example, assuming that older workers are less likely to be able to do heavy lifting. In this case, employers posting such ads and offering fewer callbacks to older workers would be engaging in pure statistical discrimination.

While economists are interested in the nature of discriminatory behavior, both statistical and taste discrimination are illegal under U.S. law. EEOC regulations state: “An employer may not base hiring decisions on stereotypes and assumptions about a person’s race, color,

<sup>17</sup>In legal cases, the most compelling data on hiring discrimination comes from comparing hiring rates of the group in question (older workers, in our case) relative to the applicant pool. In the absence of data on applicants, the analysis of a firm’s workforce relative to the age structure of the relevant workforce in the population is sometimes used, but such analyses pose a greater challenge to establishing evidence consistent with age discrimination.

<sup>18</sup>One might be concerned that different people in an organization write the ads and make the hiring decisions. However, the job board we use is focused on smaller employers, and our experimental protocol sharpens this focus because we do not use ads from larger companies that direct job seekers to their online job application sites (see NBB).

<sup>19</sup>Taste discrimination is discrimination that occurs because employers, employees, or customers having animus against, or a distaste for, the group in question. Statistical discrimination is defined as using actual or perceived group-level differences – such as stereotypes – to make inferences about an individual from the group and hence treat that individual differently. See Neumark (2018) for additional discussion.

religion, sex (including pregnancy), national origin, age (40 or older), disability or genetic information.”<sup>20</sup> The regulations do not refer to whether the stereotypes are correct (i.e., right on average), although from an efficiency perspective economists would likely be more concerned about incorrect stereotypes.

While a justification for statistical discrimination may simply be that it is unfair, labor economics research shows that statistical discrimination can generate inefficiencies when workers’ human capital investments respond to this discrimination (see the seminal paper by Lundberg and Startz, 1983). For example, in the context of aging, an assumption like “older workers do not understand new technology” can deter investments by older workers in acquiring new technological skills. Of course, age stereotypes can also be inefficient if they are wrong, and there is ample evidence that simple assumptions about declines in workers’ skills as they get older are sometimes incorrect (for recent evidence and a review, see Börsch-Supan, Hunkler, and Weiss, 2021).

A somewhat different and more complicated question is whether job requirements reflected in the stereotyped language in job ads, to the extent they result in less hiring of older workers, are legal. Legality generally requires an employer to show that these requirements are based on a reasonable factor other than age (RFOA), even if that factor is correlated with age.<sup>21</sup> In other words, a job requirement that is associated with less hiring of older workers is not necessarily illegal. Our evidence does not speak to the potential legality of job requirements that reflect age stereotypes. However, evidence that such job requirements are associated with hiring discrimination against older workers would prompt important questions about the validity of these job requirements, and more so if we think the first hypothesis – that employers put these requirements in job ads to discourage older workers from applying – has some validity.

We do not necessarily know – nor do we need to take a stand – on why employers discriminate based on age. They may want to avoid older workers because of taste-based discrimination or because of statistical discrimination. The potential implications for the observed relationship between stereotyped language and hiring are the same.

## Methods

An essential task in this paper is to classify job ads by the age stereotypes that appear in their text. To do this, we scrape the text and use language processing software and algorithms to identify language related to age stereotypes and to quantify the strength of the relationships.<sup>22</sup> We then use this information to test whether employers who use language in their job ads that is related to stereotypes of older workers are less likely to hire older workers – as captured in the experimental results.

<sup>20</sup>See <http://www1.eeoc.gov/laws/practices/index.cfm?renderforprint=1> (viewed September 15, 2019).

<sup>21</sup>An RFOA is defined as “a non-age factor that is objectively reasonable when viewed from the position of a prudent employer mindful of its responsibilities under the ADEA under like circumstances.” See <https://www.federalregister.gov/documents/2012/03/30/2012-5896/disparate-impact-and-reasonable-factors-other-than-age-under-the-age-discrimination-in-employment> (viewed September 15, 2019).

<sup>22</sup>Online Technical Appendix C provides more details, references to computer code, etc.

Our strategy was to specify the relationships between job-ad language and age stereotypes *ex ante*, prior to doing any analysis of which job-ad language predicts measured discrimination, and also to make the identification of which phrases from job ads predict discrimination mechanical. This dual strategy was intended to avoid (i) cherry-picking phrases from job ads that predict age discrimination, (ii) specification search to emphasize results suggesting that stereotyped phrases are associated with discrimination, and (iii) *ex post* rationalization of the results (finding which phrases in the job ads predict discrimination and then searching for age stereotypes related to these phrases).<sup>23</sup>

Our first step was to identify common age stereotypes from the research literature in industrial psychology and related fields. Second, we use computer science methods for measuring semantic similarity in text data to identify and code words and phrases in the job ads that are related to specific age stereotypes (Mikolov et al., 2013a and 2013b). Third, for each job ad, we use all the phrases in the job ad to calculate the job-ad-specific distribution of semantic similarity scores for each stereotype. We quantify the usage of stereotyped language across ads for each stereotype, based on the 95<sup>th</sup> percentile of the distribution for each stereotype in each ad.<sup>24</sup> Finally, we regress a dummy variable for observing age discrimination in our experiment on the 95<sup>th</sup> percentiles of each ad's similarity score distributions – for all of the stereotypes simultaneously. If we find a positive effect of the 95<sup>th</sup> percentile for a particular stereotype, the implication is that job-ad language related to that stereotype predicts hiring discrimination against older workers.<sup>25</sup> We explain these steps in the following subsections.

### Identifying Stereotypes of Older Workers

We conducted a detailed review of the industrial psychology, communications, and related literature to identify age stereotypes that this research identifies as applying to workers in their 50s and 60s. We relied on studies that were more likely to cover the cohorts covered by the data in NBB, since age stereotypes may change over time (Gordon and Arvey, 2004); we avoided studies published before the 1980s and studies of non-Western countries. We reviewed an extensive set of literature reviews and meta-analyses to identify the relevant studies, but we draw our stereotypes from papers that tested for stereotypes rather than papers that simply reported or aggregated the evidence on stereotypes from other studies.

We compiled lists of the stereotypes that these studies identified as applying to older workers. Since studies often have similar stereotypes but phrase them differently, we grouped very similar stereotypes into aggregate categories in a similar manner to the literature review and meta-analysis papers (e.g., Posthuma and Campion, 2007).<sup>26</sup> To focus

<sup>23</sup>This strategy is reflected in our analysis using the complete Wikipedia corpus, described below. The additional analyses we present were suggested by reviewers or the editor, which of course happened after our initial work.

<sup>24</sup>For example, consider two job ads. The first has many phrases about physical ability (a negative stereotype applied to older workers), and the second does not. Then the 95<sup>th</sup> percentile of the distribution of similarity scores to the physical ability stereotype will be much higher for the first ad.

<sup>25</sup>An alternative procedure we explored was to use machine learning methods (Elastic Net) to identify the words and phrases from the job ads that predict age discrimination in hiring, and to analyze statistically whether the selected words and phrases that predict age discrimination are related to age stereotypes, based on the semantic similarity scores. The results were qualitatively similar, but much more complicated to present, explain, and interpret. (An earlier version of the paper with these results is available upon request.) We are grateful to an anonymous reader of this paper for suggesting the simpler approach.

the analysis on stereotypes on which research agrees, we included a stereotype in our analysis only if at least two studies confirmed the stereotype.

This process led to 17 stereotypes of older workers, listed in Tables 1–3, corresponding to stereotypes related to health, personality, and skills. Of these 17 stereotypes, 11 (including all the health-related stereotypes) are negative – lower ability to learn, less adaptable, less attractive, worse communication skills, less physically able, less productive, worse with technology, less creative, worse memory, hard of hearing, and negative personality – and six are positive (more productive, dependable, careful, more experienced, better communication skills, and warm personality). Note that three pairs are contradictory: worse/better communication skills, warm/negative personality, and less/more productive. Our empirical analysis provides evidence on the effects of these age-related stereotypes in either direction, which is informative about the net effect of these related stereotypes – in favor of or against older workers.

### Matching Stereotypes to Words and Phrases in the Job Ads

The most complex part of our research is the machine-learning methods to identify words and phrases in the job ads that are related to the 17 stereotypes. The complication is that we do not expect age stereotypes to be expressed in the job ads precisely as they are in the research literature. Rather, there are many words and phrases that are potentially related to these 17 stereotypes, and the strength of their associations with age stereotypes can vary.

Our computational linguistics methods consist of two steps. First, we use machine learning to calibrate a model to identify the semantic similarity between words and phrases. In particular, we use machine learning to train a model using textual data from English-language Wikipedia.<sup>27</sup> The model has a structure that relates semantic similarities among the 885,424 words used in the job ads based on their usage in Wikipedia articles.<sup>28</sup> Second, we use this Wikipedia model to calculate the similarity between the 17 stereotypes and phrases in the job ads. We now turn to a more detailed explanation of our methods.

In the first step, we train the model using the entirety of English-language Wikipedia. The method uses neural networks, which are trained to reconstruct linguistic contexts of words. These neural networks take what would otherwise appear to be a jumble of words and sort them such that words used in similar contexts, as measured by Wikipedia, are placed closer together in a vector space. We use an algorithm called *word2vec* (Mikolov et al., 2013a and 2013b) to identify the similarity of two words using the context in which the words appear.<sup>29</sup> The *word2vec* algorithm employs a continuous “bag of words” algorithm

<sup>26</sup>For example, within the aggregate category of “Less Adaptable,” we include: “resistant to change” (McGregor and Gray, 2002; Weiss and Maurer, 2004); “adapt less well to change” (Warr and Pennington, 1993); and “[less] flexibility” (Levin, 1988).

<sup>27</sup>We use the English Wikipedia corpus as of November 3, 2017. This included 5.4 million articles. See <https://dumps.wikimedia.org/enwiki/> (viewed November 3, 2017). As is standard in the neural networks literature, we divide each Wikipedia article into paragraphs (Adafre and De Rijke, 2006). We further split the paragraphs into single sentences. Each sentence and paragraph is used as a separate document in the machine learning algorithm. The intuition is that sentences can provide information on closer relationships between words, like “ice” and “cold,” while paragraphs are needed for more general relationships, like “ice” and “Antarctica,” which are related but might be less likely to appear in the same sentence.

<sup>28</sup>Note that in the English language there are fewer than 885,424 words. For example, the *Oxford English Dictionary*, second edition, includes 171,476 words in current use (<https://www.lexico.com/en/explore/how-many-words-are-there-in-the-english-language>, viewed September 15, 2019). But the job ads include names, places, misspellings, verb conjugations, etc.

to use the context of a word's usage to predict other related words. The model produces a vector space where each unique word from Wikipedia is given a corresponding vector in a vector space. Words that are used more similarly to each other are located closer together in the vector space. This vector space is the mathematical representation of the relationships between these words, and can be used to construct a numerical measure of the distance between any two words.

The structure of the *word2vec* neural network begins with the inputs (we use as our main corpus the entirety of English-language Wikipedia). It then uses a series of linear projection functions (also known as hidden layers because the researcher does not observe them) to transform the textual data into a vector space. These hidden layers sort the inputted text-based data to identify relationships between words in the inputted text, capturing more complex relationships between words in the texts with each layer. To identify relationships in the data, the model aggregates the data and shrinks the dimensions of the textual data. Each layer takes in a series of inputs from the previous layer and projects the data onto the next layer, reducing the dimensionality of the vector without losing valuable information. These projections are linear functions that weight all the inputs to produce an output. Each linear projection function has a series of weights that transform the data, and a constant (known as the bias), which shifts the projection to improve prediction.<sup>30</sup>

As the data works its way through the model, the words entered in the input phase are shifted and sorted such that words that are semantically similar to each other are situated closer to each other in the output vector space. Each vector in the vector space acts as an address for a word, allowing us to determine the similarity between any two words based on how far they are from each other, using a numerical representation of the distance between them – the “cosine similarity score” – defined below.<sup>31</sup>

To analyze semantic similarity with massive databases like Wikipedia, the recommended vector size is between 100 and 200 nodes. The more nodes, the more precise the model will be. We picked 200 nodes to increase precision in the measurement of semantic similarity.<sup>32</sup> Hence, the actual neural network we construct takes as its input an  $885,424 \times 1$  vector containing all the words and projects it into an output matrix that is  $885,424 \times 200$ .<sup>33</sup> We

<sup>29</sup>Our application of the *word2vec* algorithm is taken from <https://radimrehurek.com/gensim/models/word2vec.html> (viewed September 15, 2019). Readers interested in learning more about this method are directed to <http://kavita-ganesan.com/gensim-word2vec-tutorial-starter-code/#.XOthPlhKiU1> (viewed September 15, 2019) for an overview of the implementation of the *word2vec* algorithm and alternative applications.

<sup>30</sup>For simplicity, imagine a neural network with two dimensions (rather than the actual 200 dimension vector space we use).  $y$  is the output of the hidden layer, which is a cardinal number such that two words closer together in meaning based on their usage in Wikipedia will have numbers closer together.  $y$  is a linear function of dummy variables for every word in the layer ( $x$ ), with weights and a bias correction that allows the projection function to shift up or down to improve the predictive power of the model:  $y = w_1x_1 + w_2x_2 + w_3x_3 + b$ . The bias correction is ( $b$ ), and the weights ( $w$ ) are the coefficients of the model.

<sup>31</sup>Appendix Figure A1 provides an illustration of how the *word2vec* algorithm operates. In this case, there are five inputs that are closely related, hence (hypothetically) belonging to a single layer. The *word2vec* algorithm takes the vector of input words and projects them to an output vector. The output vector is ordered such that words that are more closely related to each other are placed closer to each other (e.g., “muscle” is closer to “athlete” than to “carry,” based on usage in Wikipedia). This example features a  $5 \times 1$  vector projected onto a  $5 \times 1$  output vector. There is a total of five words and only one node (the second dimension of the output vector) to define the context.

<sup>32</sup>Pennington et al. (2014) show that there is a considerable gain in the accuracy from 100 to 200 nodes, but after 200, the gains are very marginal (see Figure 2 of Pennington et al., 2014).

<sup>33</sup>In our *word2vec* algorithm, the creation of this neural network begins by working from the input layer to the output to determine the optimal weights and bias in each layer of the network (“forward propagation”). This step consists of estimating the probability



use this matrix – the neural network created by our *word2vec* algorithm – to calculate the semantic similarity of two words based on the cosine similarity score (defined below).<sup>34</sup>

Once we estimate the vector space, we use these cosine similarity scores to identify the words in the job ads with usage (in Wikipedia) that is highly related to the usage (again, in Wikipedia) of the age stereotypes.<sup>35</sup> However, to this point, our explanation (and the example in Appendix Figure A1) have been based on single words. Because a single word may often fail to contain enough information about the association with a stereotype (which are typically expressed in multiple words), we instead use three-word phrases from the job ads in our analysis, or “trigrams.” We create these trigrams by removing words such as “the,” “and,” or “a” – so-called “stopping words” in language processing – and then creating all trigrams from the remaining words. The trigrams are all sets of three consecutive words, excluding stopping words. We retain the stereotypes as the words in which they are expressed in the first column of Tables 1–3, after we remove the words indicating the direction of the stereotype, such as “more” or “less.”

Then, we calculate the cosine similarity (CS) score between each stereotype and every trigram used in the entire set of job ads. Because the *word2vec* model is created using single words, we have estimated weights only for single words. To calculate the CS score between stereotypes and trigrams, we recover the weights applied to the hidden layers in the network that corresponds to the word in question, apply these weights to generate new weights for the trigrams and stereotypes, and then use the vectors of these new weights to calculate the CS score.

We first estimate the vector corresponding to the three words in the trigram (or the words in a stereotype), adding the weights element-by-element for each word.<sup>36</sup> For example, if the model uses two hidden layers, producing two weights for each word of three words in the trigram “able lift lbs,” then the total vector of weights of the trigram is computed as:

$$\text{able lift lbs} = \begin{bmatrix} 0.3 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 0.4 \\ 0.1 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 1.2 \\ 0.5 \end{bmatrix}. \quad (8)$$

---

that a word is between a set of other words. We select optimal weights and bias to minimize the errors of these predictions. But when using only forward propagation, the estimated output can have a high error rate. To improve the estimation, we update the biases and weights based on the error rate in the model’s prediction using a process known as “backward propagation.” This process of using both forward and backward propagation iterations is counted as a training iteration. For our purposes, we use five training iterations of the *word2vec* algorithm (the default setting in the *word2vec* package). After the five training iterations, we have fully calibrated the neural network and populated the vector space. Our final vector space contains one row for each of the 885,424 words used on the job ads, and 200 columns containing the estimated weights from the linear projection functions.

<sup>34</sup>For more details about cosine similarity and semantic similarity and these kinds of models, see Clark (2015) and Jurafsky and Martin (2017).

<sup>35</sup>Note that there are three pairs of stereotypes that are mirror images of each other: worse/better communication skills, warm/negative personality, and less/more productive. For these pairs, we just combine the stereotypes: communication skills, personality, and productive. Thus, we end up looking at cosine similarity scores with these 14 stereotypes. When we discuss the results, below, we explicitly consider the evidence on these ambiguous stereotypes.

<sup>36</sup>This procedure is derived from Mikolov et al. (2013c). They demonstrate that the relationships between words captured by the methods we use also capture relationships between small numbers of words (their focus is on pairs), based on addition or subtraction of the vectors corresponding to these words. As a prime example, the representation of the word queen can be roughly recovered from the representations of “king,” “man,” and “woman” – i.e., queen  $\approx$  king – man + woman.



We then estimate the CS score between these vectors for every trigram from the job ads and every stereotype. The CS score is defined as:

$$CS(\text{trigram}, \text{stereotype}) = \frac{\text{dot product}(\text{trigram}, \text{stereotype})}{\|\text{trigram}\| \|\text{stereotype}\|} \quad (9)$$

where “trigram” and “stereotype” in the equation refer to the vectors of weights.<sup>37</sup>

The CS score varies between  $-1$  and  $1$ . A score of  $-1$  means the words never appear in the same sentences or paragraphs in Wikipedia. As the CS score increases, the usage of the words becomes more similar; that is, they are used more often in the same sentences or paragraphs, suggesting that they are often used to discuss the same topic. This is what the literature defines as greater semantic similarity. If the words coincide perfectly, the CS score equals  $1$ . As an example, Appendix Figure A2 shows the distribution of CS scores of all trigrams with a particular stereotype (communication skills); the distribution is centered above zero, which makes sense since we are looking at text from job ads. To provide some examples, trigrams at the lower end of the distribution are highly unrelated. These include “christmas season near” and “hotel near seattle” (both with scores of  $-0.3$ ). Trigrams with scores close to  $0.0$  include “every Sunday pm” and “work year round.” Trigrams at the top of the distribution with scores of  $1.0$  include “excellent communication skills” and “prioritizing skills communication.”

For each job ad, we use these CS scores to calculate the distribution of semantic similarity for all three-word phrases and all stereotypes. To illustrate, consider the job ad in Figure 2. The ad contains phrases that, on the surface, could be related to age stereotypes, including, for example, “computer savvy,” “experience preferred,” “energetic,” and “customer friendly.” Figure 3 displays the distributions of the CS scores for each phrase (trigram) in the ad, with the communication skills, physical ability, and technology stereotypes. There is a wide range of CS scores in this job ad, though they are more related than unrelated (lying almost entirely above  $0$ ). The distributions are skewed with a long upper tail, indicating that, though rare, the job ad does contain highly related trigrams.

Our analysis requires a summary measure of the distribution of CS scores in each job ad for each stereotype, to quantify the usage of stereotyped language (for each stereotype) across job ads. We plotted the distributions of CS scores at the median, the 75<sup>th</sup> percentile, the 95<sup>th</sup> percentile, and the maximum. Figure 4 shows these for the same three stereotypes used in Figure 3. The histograms in Figure 4 provide a sound rationale for using the 95<sup>th</sup> percentile. The mass of the distributions is much lower using the median (or the 75<sup>th</sup> percentile). This is not surprising. If a job ad only contains a few stereotyped phrases, then the phrase with the median CS score in the ad, for a given stereotype, is likely quite unrelated to that stereotype. If we used lower percentiles, such as the median, as our measure of ageist sentiment in a job ad, we would be using variation in the language across job ads that is largely unrelated to the stereotypes we are studying. In addition, Figure 4 shows that if we used the median (or the 75<sup>th</sup> percentile), we would not pick up much range in the language across ads. This is also

<sup>37</sup>The  $\|$  notation indicates the Euclidean norm – e.g.,  $\| [x, y]^T \| = (x^2 + y^2)^{1/2}$ .

not surprising for the same reason; the phrases with, e.g., median CS scores are more likely to be generic phrases that do not vary much across job ads.

Since higher CS scores indicate a stronger relation to the stereotype, selecting a higher percentile of the CS score distribution ensures that the phrases being analyzed are more related to the stereotypes. On the other hand, if we use the maximum, we get what appears to be a good deal more noise (especially for communication). We would expect this because we are looking at extremes of the distribution, and language in the extreme upper tail but with different CS scores may not imply any real differences in behavior. Therefore, we chose to use the 95<sup>th</sup> percentile of the distributions of the CS scores to measure how stereotyped the language in a job ad is to quantify differences in the usage of stereotyped language across job ads.<sup>38</sup>

To provide more details and examples, in Figure 5, we display the distributions – for each stereotype related to health – of the 95<sup>th</sup> percentiles of the CS score computed from each job ad.<sup>39</sup> The figure displays this information for all occupations combined, and then each occupation separately. The figure shows that the trigrams in job ads are fairly weakly related to hearing and memory, but that ads with language strongly related to physical ability are relatively common. For this latter stereotype, the distributions in all occupations feature a large mass of ads with trigrams in the range of 0.6 or higher.<sup>40</sup>

### Testing which Stereotypes Predict Callback Differences by Age

Our next step is to use our job-ad-level measures of the semantic similarity between language in the ads and age stereotypes to estimate the relationships between age-stereotyped language in job ads and the likelihood that older or younger applicants received callbacks.

Our data set includes all responses to the triplet of job applications sent in response to each job ad that we could match to an employer and their job advertisement.<sup>41</sup> Our outcome

<sup>38</sup>We explore robustness to this choice in Online Appendix B, discussed below.

<sup>39</sup>As noted earlier, we combine the stereotypes that the literature indicates could be both positively or negatively associated with older workers (e.g., “warm personality” and “negative personality” becoming “personality”). The CS scores do not differentiate between the sentiment of the words. For example, phrases such as “good personality positive” and “bad personality negative” will both have a positive CS score with personality. We let the evidence on the association between stereotyped language and the experimental measure of age discrimination tell us whether, on net, language associated with these ambiguous stereotypes is associated with higher or lower hiring of older applicants.

<sup>40</sup>Appendix Figures A3 and A4 show the same information for stereotypes related to personality and related to skills, respectively. For personality, the distributions of the 95<sup>th</sup> percentiles of the CS scores are more normally (or at least symmetrically) distributed than those in Figure 5. The job ads do not include phrases highly related to many of the stereotypes, with the median of the 95<sup>th</sup> percentile in the 0.2 to 0.4 range, and the distributions do not have the large upper tails we saw for physical ability. Still, there is some variation apparent both by stereotype and occupation. For example, the distributions for careful in the security guard and the janitor ads are shifted notably to the right (and the distributions for this stereotype are furthest to the right for administrative assistant ads as well, although not as markedly). For skills, we find that job ads contain a higher frequency of trigrams related to some skill-related stereotypes. Job-ad language strongly related to ability to learn and communication skills is much more common than for the other three stereotypes in this figure, and for the stereotype shown in the other two figures. However, the upper tails are not as extreme as for physical ability. Phrases in job ads strongly related to the technology stereotype are also more common (although not as pronounced); for all occupations except janitors, the medians are closer to 0.4. Phrases strongly related to experience and productivity are less common, with medians of the distributions of the 95<sup>th</sup> percentile scores in the 0.2 to 0.3 ranges and sometimes below 0.2.

<sup>41</sup>We were able to match 34,260 job applications to 11,420 job advertisements, corresponding to 22,840 observations for older and middle-aged applicants. There are 4,266 applications that cannot be matched to a saved job ad. The most common reason was that an ad was not saved. In some cases, the ad was saved in the incorrect format and cannot be scraped. (Research assistants were instructed to save all job advertisements as an HTML file, but there were instances of advertisements being saved as a PDF or a PNG file.) In total, 87% of applications are matched to a job ad.

$(d_{exp}^{ij})$  is a dummy variable equal to one if the older applicant  $i$  did not receive a callback from employer (or, equivalently, job ad)  $j$  but the younger applicant did, and zero otherwise. If both applicants are called back, neither applicant is called back, or only the older applicant is called back (which is less common than the reverse case), we do not consider the outcome to reflect age discrimination, and code  $d_{exp}^{ij}$  as zero.<sup>42</sup> In 76% of cases, neither applicant was called back, while in 6% of cases, both applicants were called back. In 11% of cases, the older applicant was not called back and the younger applicant was, whereas the reverse occurred in 7% of cases.<sup>43</sup>

We estimate probit models for our experimental measure of discrimination. The key independent variables are the 95<sup>th</sup> percentiles of the distributions of the CS scores of each phrase in the job ad with each stereotype. Denote these percentiles, for job ad  $j$  and stereotype  $s$ , by  $P_{js}^{95}$ . We also control for the observable resume differences, using the same control variables  $X_{ij}$  as in NBB.<sup>44</sup> Thus, our model is:

$$Pr\left[d_{exp}^{ij} = 1\right] = \alpha + \sum_s \beta_s P_{js}^{95} + X_{ij}\delta + \varepsilon_{ij}. \quad (10)$$

We standardize  $P_{js}^{95}$  (for each stereotype  $s$ ) to have a mean of zero and a standard deviation of one across all the ads in the study.  $\beta_s$  then represents the effect of a one standard deviation increase in the 95<sup>th</sup> percentile of the CS score in a job ad for stereotype  $s$ .<sup>45</sup> The estimate of  $\beta_s$  then reflects both the effect of the stereotype on discrimination and how related the job-ad language is to the stereotype. A small  $\beta_s$  could be attributable to stereotype  $s$  not mattering much for discrimination, or to a one standard deviation increase in similarity to stereotype  $s$  being a small increase. From the point of view of asking which stereotypes in job-ad language predict age discrimination, this combined effect is of interest. In contrast, a comparison between the estimated effects of the same absolute change in the 95<sup>th</sup> percentile for different stereotypes is of less interest, given that, in reality, job-ad language is more closely related to some stereotypes than to others.

<sup>42</sup>In theory, it is possible to impose an even stronger definition of discrimination on the data, defining discrimination as occurring if the younger applicant is called back but neither older applicant is. The challenge in using this definition is in the construction of the triplets. All triplets had one younger applicant and two older applicants, but the older applicants could either be middle-aged or older. So in some triplets the older workers will be a mixed pair, one old and one middle-aged. In these cases, the stronger definition of discrimination would require discrimination to occur against the applicant aged 49 to 51 and the applicant aged 64 to 66. However, in NBB we generally observed stronger evidence of discrimination against older applicants than middle-aged applicants, and sometimes no discrimination against middle-aged applicants. The way we define discrimination here is more informative, as it results in separate estimates for middle-aged vs. younger applicants and older vs. younger applicants. (This issue could be avoided in future studies by simply sending pairs of applicants in response to each job ad.)

<sup>43</sup>In each triplet sent to a job opening, there was one young worker and two older workers (randomly selected to be either middle-aged or old). Our unit of observation is each middle-aged or older applicant, so that each triplet produces two observations. Thus, discrimination against an older applicant is measured independently of whether the other older worker was called back.

<sup>44</sup>Resume features include: city, order sent, skill level, unemployment status, template, and email domain.

<sup>45</sup>In Appendix Table A1, we provide the text of some trigrams to indicate how they differ one standard deviation higher in the distribution of 95<sup>th</sup> percentile CS scores. We report the text at the mean 95<sup>th</sup> percentile and the five trigrams closest to one standard deviation higher than this mean. This corresponds to the interpretation of  $\beta_s$ , although of course identification of  $\beta_s$  comes from variation across the entire distribution.

We estimate our models at the gender-age-occupation level (e.g., women aged 49 to 51 applying to sales jobs) to allow the effects of stereotypes to vary across different kinds of employers and applicants. Because each job ad has two pairs of applicants – one younger applicant and one either middle-aged or older applicant in each pair – we cluster the standard errors at the job-ad level.

## Results

We now turn to the evidence on the relationships between age stereotypes in job ads and age discrimination in hiring. Figure 6 presents a convenient summary of the results from estimates of equation (10) by age, gender, and occupation, and the full underlying estimates are reported in Table 4; both show the estimated marginal effects. The top row of Table 4 shows the percentage of job ads where the younger applicant received a callback but the older applicant did not – our experimental measure of discrimination. This occurred in 11% of cases, but the rate varies by occupation. For the regression estimates below the first row, the coefficients measure the marginal effect on the probability of age discrimination of a one standard deviation increase in the CS score at the 95<sup>th</sup> percentile of the job ads' distributions for that stereotype. For example, middle-aged men who applied for a job as janitors (in column (5)) experienced age discrimination 9.6% of the time. If the job ad included language where the 95<sup>th</sup> percentile was one standard deviation more highly related to physical ability than the average ad, measured discrimination was 3.8 percentage points (or 40%) higher.<sup>46</sup>

Figure 6 provides a compact way to visualize the estimates, displaying the estimated marginal effects and their 95% confidence intervals.<sup>47</sup> Figure 6 shows that stereotyped language in job ads is associated with hiring discrimination, especially for men. We observe significant heterogeneity across occupations in which stereotypes are associated with hiring discrimination. In some age-occupation cells, we observe only one or two significant effects, and in some cases we observe more. When we take a more aggregated view, we find that middle-aged men have the highest number of stereotypes correlated with hiring discrimination (eleven), followed by older men (3) and older women (2), while middle-aged women have the fewest (0). For older women, stereotyped language related to personality appears to be most strongly correlated with hiring discrimination. Among men, we observe health (but only for middle-aged men), personality, and skill-related stereotypes all being associated with hiring discrimination.

<sup>46</sup>Table 4 includes results from correcting for multiple hypothesis testing (p-values in square brackets). We use the Simes procedure to control the false discovery rate. The false discovery rate (e.g., at the 5% significance level) means that we are 95% confident that at least some of the rejected hypotheses are false. The more conservative approach controls the family-wise error rate, which means that we would be 95% confident that all rejected hypotheses are false. Controlling the family-wise error rate is more appropriate when there is the potential for “harm” from falsely rejecting any of the tested hypotheses. In contrast, controlling the false discovery rate is more appropriate when harm is less likely to be caused by a single true hypothesis being falsely rejected, as long as some are correctly rejected (Pike, 2011). We think the latter is more appropriate to our case, as we are testing, in general, whether age stereotypes in job ads predict hiring discrimination (as opposed, say, to testing the effects of a number of interventions on a disease). We would also suggest that the need for these corrections is less compelling in our case than in some others, as we prespecified the stereotypes we study and use machine learning to measure semantic similarity. We thus do not have the issue of, e.g., reporting a subset of results from a large set of possible analyses (as discussed, e.g., in Casey, Glennerster, and Miguel, 2012). The adjusted p-values are higher in all instances. None of the results for women remain significant, but the strongest results for men remain significant or marginally significant in many cases.

<sup>47</sup>Note that the horizontal scale is not the same in every figure – including the other versions of these figures that come later.

We now discuss these results in more detail. In general, our evidence regarding the relationships between stereotyped job-ad language and discrimination against older job applicants is consistent with what we would expect from the evidence on stereotypes in the industrial psychology literature.

### Results for Stereotypes Related to Health

In Figure 6 (and Table 4), we find that stereotypes related to the health of older workers predict higher levels of discrimination for middle-aged men but not for women (of either age group) or older men. The differences are primarily attributable to larger point estimates of the effects for middle-aged men rather than more precise estimates. For middle-aged men, the health-stereotyped language is always associated with more discrimination.

The stereotypes that are associated with increased discrimination depend on the occupation. For retail sales and janitor positions, we find that employers with job ads that used language more highly related to physical ability more often discriminate against middle-aged men (significant at the 5% level). The estimates for middle-aged men in security jobs and older men in sales jobs are in the same direction, with slightly weaker statistical evidence (significant only at the 10% level). We expected to find this based on the industrial psychology literature indicating that employers view older workers as having lower physical abilities than younger workers (see Table 1). For janitor positions, job ads with phrases for which the 95<sup>th</sup> percentile of the distribution of CS scores with physical ability were one standard deviation higher are associated with a 3.8 percentage point increase in discrimination against middle-aged men. Similarly, for sales positions, a one standard deviation increase in this 95<sup>th</sup> percentile is associated with a 3.2 percentage point increase in discrimination against middle-aged men.<sup>48</sup>

For retail sales positions, we find that employers with job ads that used language more highly related to memory more often discriminate against middle-aged men. This is expected based on the industrial psychology literature indicating that employers view older workers as having worse memories than younger workers (see Table 1). Job ads with phrases for which the 95<sup>th</sup> percentile of the distribution of CS scores with memory were one standard deviation higher are associated with a 3.0 percentage point increase in discrimination against middle-aged men.<sup>49</sup> For security guard positions, we find that employers with job ads that used language more highly related to hearing more often discriminate against middle-aged men. This evidence is consistent with the industrial psychology literature indicating that employers view older workers as having worse hearing

<sup>48</sup>In Appendix Table A1, we show that this one standard deviation increase in the CS score for physical ability is akin to changing the 95<sup>th</sup> percentile from “assistant position available” to a phrase similar in relatedness to “work preferred necessary,” “fast paced fun,” or “required flexibility required.”

<sup>49</sup>This result highlights a potential challenge in adapting methods from machine learning to our context of analyzing text data. The *word2vec* algorithm could sometimes identify trigrams from the job ads that are not meaningfully related to our age stereotypes – which we might think of as false positives – and if these trigrams happen to predict lower relative callback rates for older job applicants, these false positives could generate bias towards concluding that job-ad language related to specific age stereotypes predicts age discrimination. The stereotype “memory” may be a prime example. We find a significant correlation between using words highly related to memory and discrimination against middle-aged men. But this result could be driven in part by an ad mentioning “computers,” which was identified by the algorithm as being highly related to “memory” because of the number of times “computer memory” is mentioned in Wikipedia. Therefore, we may identify a significant correlation for memory, when in reality we are picking up a technology-related correlation.

than younger workers (see Table 1). Job ads with phrases for which the 95<sup>th</sup> percentile of the distribution of CS scores with hearing were one standard deviation higher are associated with a 2.0 percentage point increase in discrimination against middle-aged men.

### Results for Stereotypes Related to Personality

Stereotypes related to personality appear to explain discrimination for older women, but not for middle-aged women; the evidence indicates that stereotyped language related to personality is associated with less discrimination against older women. For older women in administrative jobs, employers who use language in their job ads more related to dependability discriminate less against older workers. This result is in line with the predictions of the industrial psychology literature, which indicates that employers view older workers as more dependable than younger workers (see Table 2). For administrative assistant job ads, job ads with phrases for which the 95<sup>th</sup> percentile of the distribution of CS scores with dependability were one standard deviation higher decrease measured discrimination against older women by 1.6 percentage points.

For older women in retail jobs, employers who use language in their job ads more related to the personality stereotype discriminate less against older workers. In the industrial psychology literature, employers' views of the personalities of older workers are mixed (see Table 2); some papers suggest that employers view older workers as having worse personalities, while others provide evidence that employers view them as having warm personalities. Our results suggest that in retail sales the more positive stereotype dominates for older women. For sales associate positions, job ads with phrases for which the 95<sup>th</sup> percentile of the distribution of CS scores with personality were one standard deviation higher decrease measured discrimination against older women by 3.9 percentage points.

For middle-aged men applying to janitor positions, and for both older and middle-aged men applying to retail sales positions, we find that employers with job ads with phrases more highly related to careful are less likely to discriminate against older workers. This result matches the industrial psychology literature, which indicates that employers view older workers as more careful than younger workers (see Table 2). For janitor job ads, when phrases for which the 95<sup>th</sup> percentile of the distribution of CS scores with careful were one standard deviation higher, measured discrimination was 4.2 percentage points lower. For sales associate job ads, a one standard deviation increase in this 95<sup>th</sup> percentile decreases observed discrimination by 3.5 percentage points for middle-aged applicants and 3.1 percentage points for older applicants. For middle-aged men applying to security guard positions, we observe results that are at odds with the industrial psychology literature, which finds that older workers were viewed by employers as less creative (see Table 2). In contrast, we find evidence that employers who use language related to creativity are less likely to discriminate against older workers. Job ads with phrases for which the 95<sup>th</sup> percentile of the distribution of CS scores with creative were one standard deviation higher are associated with measured discrimination being 2.6 percentage points lower. Finally, for older men in retail sales, we find that language related to adaptability is associated with increased discrimination. This finding is consistent with the industrial psychology literature, which shows that older workers are viewed as less adaptable (see Table 2). Job ads with phrases



for which the 95<sup>th</sup> percentile of the distribution of CS scores with adaptability were one standard deviation higher increase observed discrimination by 3.1 percentage points.

### Results for Stereotypes Related to Skills

For women, we do not observe any significant relationship between phrases highly related to skill-related stereotypes and observed discrimination. This is true in both administrative assistants and retail sales positions, as well as across age groups. The very small point estimates in the administrative assistant models suggest that the language is very similar on ads where the employer discriminates and ads where the employer does not. We observe larger differences in retail sales ads, but the smaller sample size results in larger standard errors. Still, the point estimates are smaller in absolute value than the significant results we have discussed above.

Among men, we observe skill-related language being associated with discrimination against older workers in several age-occupation cells. A wider range of stereotyped language predicts discrimination for middle-aged men, and we observe significant heterogeneity across occupations. In contrast, the evidence for older men appears only in janitor positions.

In the industrial psychology literature, there is disagreement about whether employers stereotype older workers as more productive or less productive (see Table 3). Thus, like for personality, this is a case where our method enables us to gauge the relative importance of the positive or negative association empirically by measuring employer behavior. Our results suggest that among the employers who are hiring janitors, the positive association between age and productivity dominates. We find that employers who use language more highly related to productivity are less likely to discriminate against middle-aged men. Job ads with phrases for which the 95<sup>th</sup> percentile of the distribution of CS scores with productivity were one standard deviation higher decrease measured discrimination by 4.9 percentage points.

For middle-aged men in security guard positions, we observe our only instance of technology-related language being correlated with hiring discrimination. Despite a significant emphasis on this in the industrial psychology literature, it does not appear that the usage of language related to technology often differs between employers who discriminate against older workers and those who do not. Our estimated correlation is in the same direction as the negative stereotypes about older workers and technology suggested by the literature (see Table 3), as we observe a higher rate of discrimination associated with job ads that use language highly related to technology. Job ads with phrases for which the 95<sup>th</sup> percentile of the distribution of CS scores with technology were one standard deviation higher increase observed discrimination against middle-aged men by 2.1 percentage points.

For middle-aged and older men applying to janitor positions, we observe significant associations between measured discrimination and job-ad language related to experience. In the industrial psychology literature, employers view older workers as more experienced, perhaps tautologically. In contrast, we find that employers with job-ad language more highly related to experience are more likely to discriminate against middle-aged and older men. Job ads with phrases for which the 95<sup>th</sup> percentile of the distribution of CS scores with



experience were one standard deviation higher increase observed discrimination against middle-aged men by 3.3 percentage points and against older men by 4.2 percentage points.

## Summary

Overall, our results suggest that ageist stereotypes may affect the hiring of older workers. Using the ageist stereotypes found in the industrial psychology literature, we show that job-ad language that is highly related to ageist stereotypes is associated with hiring discrimination. The direction of these empirical correlations is generally in the same direction as predicted by the literature. Positive stereotypes are correlated with less hiring discrimination, and negative stereotypes are correlated with more discrimination. We only find three instances where our results contradict the predictions of the industrial psychology literature, with negative stereotypes associated with less hiring discrimination. In contrast, in 13 cases the evidence is consistent with the industrial psychology literature.

## Supplemental Analyses

We now turn to alternative ways to conduct our analysis, examining how robust our results are to different choices about how to define discrimination, the construction of phrases, and alternative corpora or methods to flag stereotyped job ad language.<sup>50</sup>

### Definition of Discrimination

To this point, we have studied discrimination against older applicants, defined as a callback to younger applicants but not older applicants. It is possible that studying discrimination in favor of older applicants (against younger applicants) would detect more evidence of positive stereotypes reducing discrimination against older workers. We therefore also did analyses for this reverse outcome – when only the older applicant received a callback. For this analysis, we created separate pairs of each older applicant in the pair combined with the corresponding younger applicant (even though this means younger applicants get used in two pairs). We did this because otherwise we would have to use a more stringent definition of favoring older applicants, with both older applicants but not the younger applicant getting callbacks. In this analysis, we aggregate the comparisons of older workers to younger workers and ignore the variation in whether employers prefer middle-aged and older workers to younger workers.<sup>51</sup>

Figure 7 presents the results.<sup>52</sup> In general, while we find some evidence linking job-ad language to measured discrimination against younger applicants (or in favor of older ones), the results are less clearly consistent with the predictions from the industrial psychology literature. For women, we find that all of the significant associations are for personality stereotypes (similar to in Figure 6). The evidence points to more discrimination against younger workers when job ads use stereotyped language associated with adaptability, and less discrimination against younger workers when the job-ad language reflects the careful

---

<sup>50</sup>We discuss the most significant supplemental analyses here. Others are discussed in Online Appendix B.

<sup>51</sup>Results looking at younger versus middle-aged and younger versus older were qualitatively similar, and results are available upon request.

<sup>52</sup>Appendix Table A2 reports the estimated marginal effects shown in Figure 7.

and dependable stereotypes. These associations point in the opposite direction to what we would have predicted based on the industrial psychology literature.

For men, we only find significant associations for janitor positions. These results are less at odds with what is predicted by the industrial psychology literature. The one exception is for job-ad language related to stereotypes about hearing, for which we find evidence of more discrimination against younger workers.<sup>53</sup> For stereotyped language related to skills, we find significant effects that align with the predictions of the industrial psychology literature. Ads that feature language more related to the ability to learn are associated with less discrimination against younger workers, and ads with language related to communication skills are associated with more discrimination against younger workers. Both of these are consistent with the positive stereotype of older workers' communication skills dominating the negative stereotype (see the conflicting stereotypes in Table 3).

The results using this alternative definition of discrimination do not replicate (with the opposite sign) the results for discrimination against older workers. The simplest explanation may be that there is little discrimination in favor of older workers and little use of age-stereotyped language by employers more likely to hire older workers.

### Alternative Corpora and Methods

Finally, we compare results using alternative methods for calculating semantic similarity. Our first method uses a naïve string matching algorithm that searches for ads containing the stereotyped words. We create 14 dummy variables corresponding to each stereotype ( $D_{js}$ ). We classify each ad as containing a stereotyped if the stereotyped words from Tables 1–3 appear. If the stereotyped word or phrase appears on the ad, the dummy variable will take on a value of one and otherwise be zero. For example, our dummy variable takes on a value of one if the words “physically able” appear anywhere in the ad. If the words “physically able” do not appear in the ad, then the dummy variable takes on a value of zero.<sup>54</sup> Using these string-matching derived dummy variables, we estimate the following equation:

$$Pr \left[ d_{exp}^{ij} = 1 \right] = \alpha + \sum_s \beta_s D_{js} + X_{ij} \delta + \varepsilon_{ij}. \quad (11)$$

$\beta_s$  in this case is the change in the probability of discrimination when the job ad features the stereotyped word or phrase. The controls remain the same as in equation (10).

The second method is more closely tied to our core methods. However, we use an alternative to the CS score derived from the complete Wikipedia corpus by instead taking a more narrow set of Wikipedia articles to construct the vector space. A potential concern with using the complete Wikipedia corpus is that we may sometimes measure semantic similarity from contexts far removed from the age stereotypes, labor markets, and discrimination. We

<sup>53</sup>Consistent with the caveat raised earlier, one possibility is that the language that the machine learning identifies as associated with “hearing” is also strongly associated with “listening,” which could be related to more positive stereotypes about older workers (such as careful or dependable).

<sup>54</sup>We use most of the unique words (excluding adverbs) from the “Phrasing” column in Tables 1–3. Details are available in Online Technical Appendix C.

thus instead use an algorithm that starts from a subset of Wikipedia articles. We set this “scrapy spider” algorithm (Kouzis-Loudas, 2016) to begin with Wikipedia pages for the following: ageism; ageing; discrimination; workforce; job; employment; stereotype; sexism; recruitment; skill (labor); and human capital. The scrapy spider algorithm then add to our corpus the text of every page linked in these initial pages. It then repeats this process, adding the text of every page linked in those pages, which collects a total of 65,532 pages. In essence, the scrapy spider algorithm reweights the cosine similarity scores to ignore all connections on pages more than two links away from the starting pages.<sup>55</sup> For example, whereas using the complete Wikipedia model, “hearing” may be highly related to court cases, under the scrapy spider algorithm, these links are down-weighted because they rarely occur within close proximity to the Wikipedia pages related to age and employment.<sup>56</sup>

Figures 8A and 8B highlight how our estimates change when we use the string matching (estimating equation (11)) and the scrapy spider algorithm (estimating equation (10) with the alternative CS scores), compared to our baseline estimates with the complete Wikipedia model. The confidence intervals are much larger for the string matching. Surely this is driven in part by many of the words or phrases infrequently appearing in the job ads.<sup>57</sup>

When we compare the results from the scrapy spider corpus to the complete Wikipedia corpus, we find minimal differences. The confidence intervals for our estimates always overlap, and the point estimates are often very close. Where the estimates differ, it may be for stereotypes where the semantic similarity was being driven by correlations not related to humans or workers – for example, computer memory vs. human memory, and court hearing vs. human hearing. Consistent with this, in the different panels in Figures 8A and 8B, the confidence intervals overlap less closely for cases including memory, technology, communication, and hearing. Moreover, there are often cases for which estimates shift towards more discrimination against older workers when ads use language highly related to the stereotype; examples include technology for women in sales (middle-aged and older), adaptable for older men in security, and technology for older men in sales. However, the estimates do not uniformly shift towards more evidence of age discrimination when we classify job-ad language using the narrower corpus from the scrapy-spider algorithm.<sup>58</sup>

## Discussion and Conclusions

We develop new machine-learning techniques for analyzing complex textual data, which we apply to job ads collected in a large-scale resume-correspondence study of age discrimination. We combine the machine-learning analysis of the text of job ads with

<sup>55</sup>We chose to set the algorithm to use two links because the number of links grows exponentially.

<sup>56</sup>Appendix Table A3 compares the words nearest the 95<sup>th</sup> percentile in an ad using this scrapy spider algorithm and using the complete Wikipedia corpus, for selected stereotypes. The full table is available upon request. We see that the words at the 95<sup>th</sup> percentile seem more related to the employment context using the scrapy spider algorithm (including for the “hearing” example just noted), but that the differences can be subtle. We also show the words used in the string matching. Comparing these two algorithms to the string matching highlights the value of measuring relatedness and not simply the presence of specific words, since all of the words shown from the Wikipedia-derived models that do not match exactly would be coded as zero in equation (11).

<sup>57</sup>For example, the words “careful” and “memory” appear in only 35 and 31 job ads, respectively. More generally, for many gender-age-occupation cells there are cases where no ad featured the words on which we string match. This is most common for the words “memory,” “careful,” and “creative.”

<sup>58</sup>Although one advantage of using complete Wikipedia is that it avoids data mining, this could be avoided using scrapy spider by pre-specifying the pages to be used, if an experiment were planned that would download job ads anew.

experimental measures of age discrimination from the correspondence study to examine whether phrases in the job ads that are strongly related to ageist stereotypes predict age discrimination in hiring.

The evidence suggests that ageist stereotypes in job ads are related to employers' decisions not to call back older applicants. For both men and women, and across different occupations, we find evidence that employers who do not call back older applicants but do call back younger applicants, or vice versa, use phrases in their job ads that are related to ageist stereotypes. For men, we find evidence that age stereotypes about all three categories we consider – health, personality, and skill – predict age discrimination, and for women, age stereotypes about personality predict age discrimination. In general, the evidence is much stronger for men, and our results for men are quite consistent with the industrial psychology literature on age stereotypes, with many negative age stereotypes reflected in job-ad language predicting more hiring discrimination against older workers and some positive age stereotypes predicting the opposite.

The stronger and more robust results for men than for women suggest that stereotypes in job ads may play a more prominent role in generating observed hiring discrimination against older men than against older women, even though our correspondence study found stronger evidence of hiring discrimination against older women. A similar puzzle is that the evidence points to larger effects of stereotypes in job ads for middle-aged than for older men, despite the experimental evidence providing stronger evidence of age discrimination in hiring against older men. Why might stereotyped language matter more to older workers when measured discrimination is lower?

One potential explanation is that the motivation to use age stereotypes to discourage workers from applying may interact with the effectiveness of age discrimination laws. If age discrimination laws are less effective for a particular subgroup of older workers, then employers who do not want to hire from this subgroup do not have as great an incentive to use ageist stereotypes in job ads to shape the applicant pool to try to avoid detection of discriminatory behavior. Age discrimination laws may be less effective for older women than for older men because of the difficulty of bringing to court intersectional discrimination claims based on being both female and older (McLaughlin, 2019). Age discrimination laws may also be less effective at deterring discrimination against older men than middle-aged men because the much smaller number of older job applicants relative to middle-aged job applicants implies that the same hiring rate difference relative to young workers is less likely to be statistically significant for the older applicants.<sup>59</sup>

In addition, the list of stereotypes we compiled from the industrial psychology literature may drive the differences by gender, for two reasons. First, the industrial psychology literature

---

<sup>59</sup>For example, suppose a company expects to get 100 young applicants, 100 middle-aged applicants, and only 10 older applicants. (This is consistent with evidence reported in NBB, and also with the much lower employment rate at ages near 65 than ages near 50.) If the company were to hire 20 young, 10 middle-aged, and one older applicant, the hiring rate differential would be 10 percentage points for both middle-aged and older applicants relative to young applicants, but the middle-aged vs. young difference would be more strongly significant because of the greater number of observations. The intersectionality issue for older women vs. older men can also be interpreted in terms of statistical evidence, because the inability to present evidence based on both age and sex can preclude evidence of a negative interaction between being older and female.

on age stereotypes primarily identifies stereotypes associated with men, in which case the stereotypes may be less salient for women (or different stereotypes than those we study may matter more). Second, the stereotypes relevant in more traditionally female jobs (like administrative assistants) may be more challenging to express in job ads, leading to weaker relationships between the underlying stereotypes and phrases from the job ads.

Our evidence that measured age discrimination is related to ageist stereotypes in job ads has potential policy implications. If older workers know the relationship between ageist job-ad language and hiring discrimination, and apply less to stereotyped job ads, then comparing hiring rates and application rates by age can understate hiring-related age discrimination. On the other hand, our evidence has potentially constructive implications for enforcing age discrimination laws. If discriminatory employers use ageist language in their job ads, then barring such language may increase applications from older workers, making it harder for hiring discrimination to go undetected. And testing for age stereotypes in job ads could be used to detect firms that may discriminate based on age in hiring decisions. Of course, the methods we develop also could be applied to discrimination against other groups.

One limitation of our work is that we only study age stereotypes that appear in job ads. Employers could have other stereotypes about older workers that affect hiring, but on which our evidence is silent. On the other hand, thinking back to our two primary hypotheses – that employers who discriminate based on age use stereotyped language to shape the applicant pool, and that employers statistically discriminate based on stereotypes about older workers' ability to meet job requirements – we may be most interested in the stereotypes expressed in job ads. For example, if age-related stereotypes in job ads signal the dimensions along which employers statistically discriminate in hiring, then these are the stereotypes that need to be assessed against the RFOA criterion. Moreover, if some stereotypes are identified in the lab, but not expressed in real-world job ads, they may simply not be very relevant to real-world labor market decisions.

A second limitation is that we do not study the behavior of job applicants in response to ageist stereotypes in job ads. As we have argued, these responses can be important, reducing the need for a discriminatory employer to hire older applicants at a lower rate. As we have shown, this can create a bias against finding evidence that stereotyped job-ad language predicts age discrimination in hiring, which could, in principle at least, explain why in many cases we do not find such evidence.

A key contribution of our techniques is that they can be adapted to other contexts. The most direct application, perhaps, is to future audit or correspondence studies of labor market discrimination. There has been an explosion in these kinds of studies in recent years,<sup>60</sup> but one could reasonably argue (see Neumark, 2018) that it is time for these studies to move beyond documenting differences in outcomes and learning more about the underlying behavior. In nearly all of these studies, some type of job-ad postings are available as objects of study. It may also be possible to apply our methods to studies of

---

<sup>60</sup>See the recent book by Gaddis (2018), and the register of studies since 2005 maintained by Baert at [https://users.ugent.be/~sbaert/research\\_register.htm](https://users.ugent.be/~sbaert/research_register.htm).

discrimination in other markets – such as housing or health care – depending on what kind of information is included in the ads or postings used in the market. With relatively few changes to our methods, researchers could test for relationships between the usage of stereotyped language and the discrimination these studies measure. Moreover, these language processing techniques could be useful in studying discrimination in different parts of the process of hiring or other employment decisions, such as recommendation letters or employee evaluations.<sup>61</sup> Researchers have also applied methods related to ours to detect bias in language in other settings, such as judges' decisions (Ash, Chen, and Ornaghi, 2020).

Moreover, related methods could be used in many other areas of labor economics. For example, a recent study presents a creative analysis of text from job advertisements and course syllabi to study how education responds to skill demands (Börner et al., 2018). Anastopoulos et al. (2019) use machine learning techniques to study the effects of immigration on demands for different kinds of jobs. And research we cited earlier uses the text of job postings from sources like [CareerBuilder.com](https://www.careerbuilder.com) and *Burning Glass Technologies* to test different hypotheses about the labor market (e.g., Marinescu and Wolthoff (2020) and Banfi and Villena-Roldán (2019) on job search). In our view, the number of potential applications in labor economics, especially combining job postings with other types of text, is enormous.

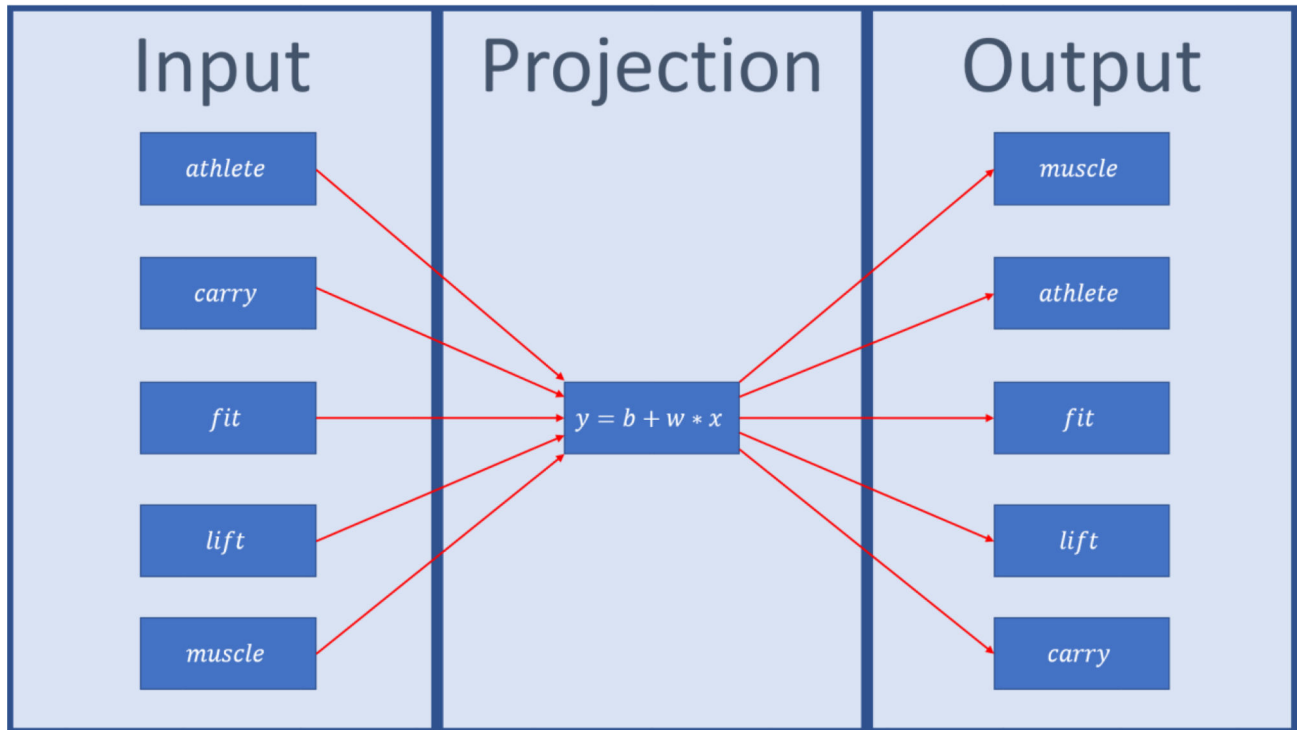
## Acknowledgments

We are grateful for helpful comments from anonymous referees, the editor, and seminar participants at the University of Bristol, University of Illinois, University of Liverpool, University of Maastricht, University of Tokyo, IZA, LISER, Southern Methodist University, and UC San Diego. We thank Hayley Alexander and Emma Tran for excellent research assistance, and are especially grateful for the help of Nanneh Chehras. Patrick Button is thankful for generous grant support from the National Institutes of Health via a postdoctoral training grant to the RAND Corporation (5T32AG000244–23), which partly funded his work on this project in 2018–2019, and to the Newcomb Institute at Tulane University for research assistant funding.

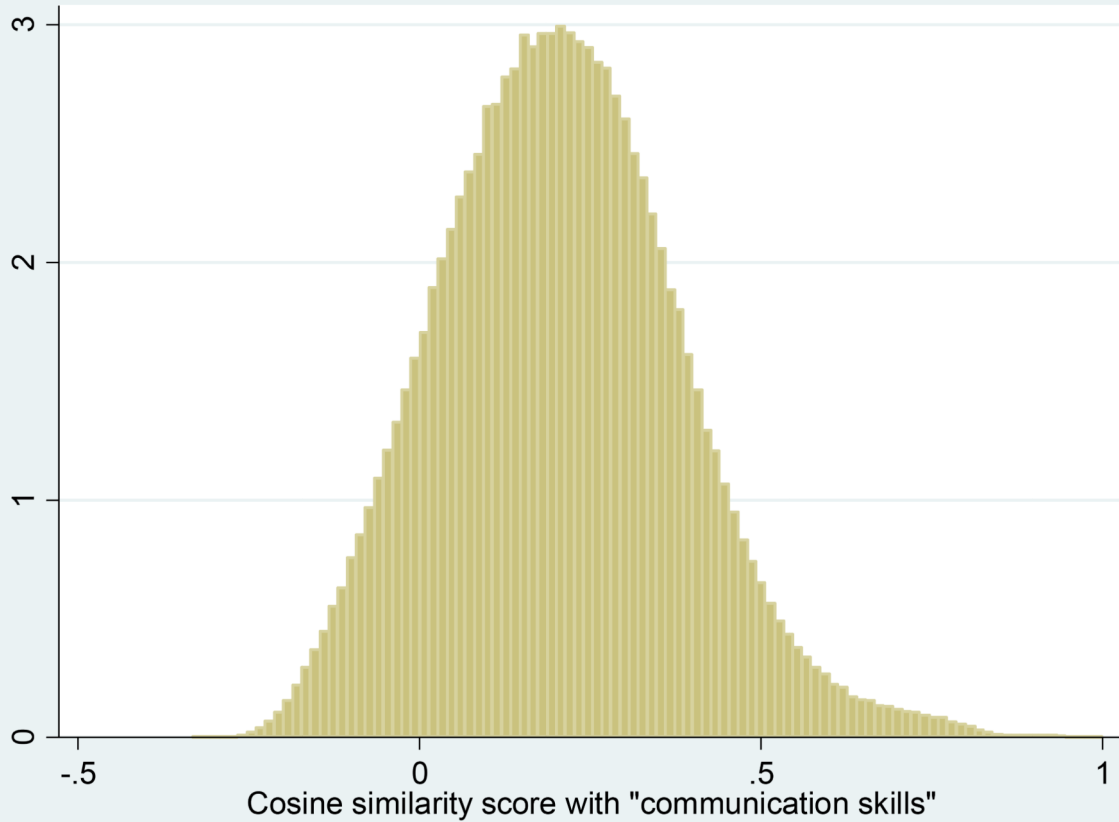
---

<sup>61</sup>For a discussion of research on letters of recommendation, see Madera et al. (2009).

## Online Appendix A: Additional Figures and Tables

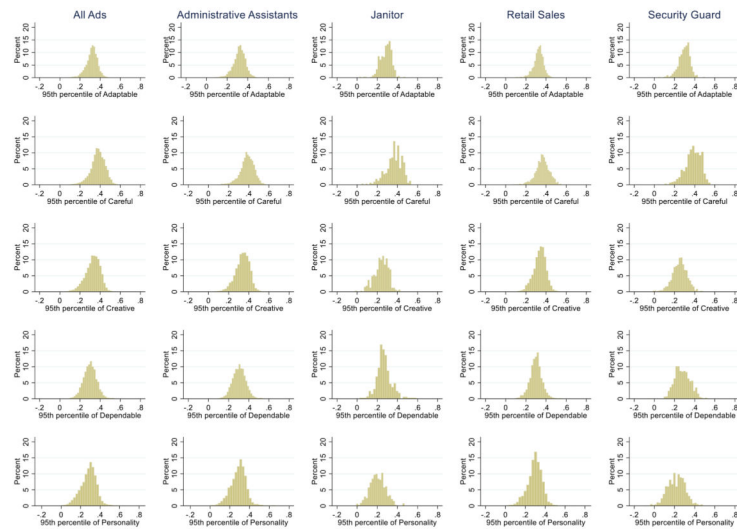
Appendix Figure A1: Visual Representation of a Hypothetical *Word2Vec* Neural Network





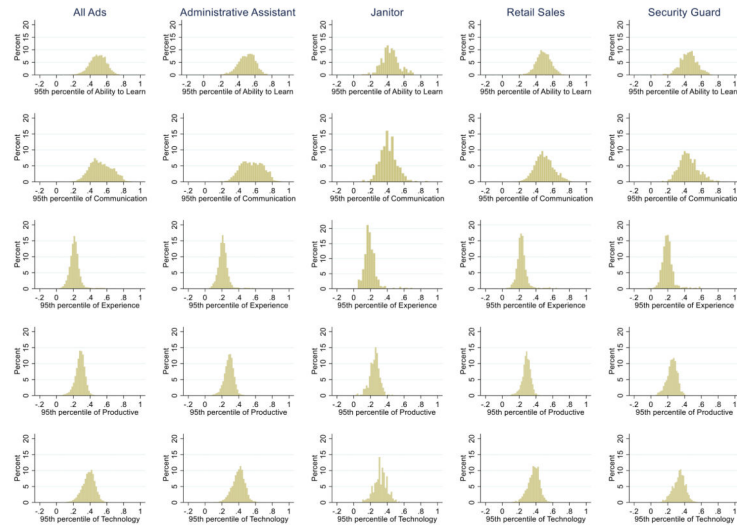
**Appendix Figure A2: Example of the Distribution of Cosine Similarity (CS) Scores**

Note: Figure reports the distribution of CS scores for all trigrams from the job ads with the communication skills stereotype. The higher the CS score, the more related the trigram is to “communication skills.”



**Appendix Figure A3: Distributions of 95<sup>th</sup> Percentiles of CS Scores for Stereotypes Related to Personality**

Note: Data come from the job ads collected in NBB. Each panel plots the distribution of CS scores at the 95<sup>th</sup> percentile for the job ads with each stereotype related to personality. The first column contains the distribution of all the ads in our sample. The remaining columns disaggregate the job ads by occupation.



**Appendix Figure A4: Distributions of 95<sup>th</sup> Percentiles of CS Scores for Stereotypes Related to Skills**

Note: Data come from the job ads collected in NBB. Each panel plots the distribution of CS score at the 95<sup>th</sup> percentile for the job ads with each stereotype related to skills. The first column contains the distribution of all the ads in our sample. The remaining columns disaggregate the job ads by occupation.

**Appendix Table A1:**

Text of Trigrams at the 95<sup>th</sup> Percentile of CS Scores within Job Ads

	Trigrams closest to 1 standard deviation above mean					
	Mean 95 <sup>th</sup> percentile	2 trigrams below	1 trigram below	Exactly 1 standard deviation above the mean	1 trigram above	2 trigrams above
	(1)	(2)	(3)	(4)	(5)	(6)
Attractive	friendly articulate personality	position seek energetic	interpersonal skills valued	passion fashion luxury	ships museum looking	cheerful ability multitask
Hearing	benefits package consideration	bridges eye care	paper work filing	hour noncommissioned officers	regard billing procedures	community seeking evening
Memory	applicant must good	background retail knowledge	oriented ability follow	time frame attitude	directs person matter	excel spreadsheets different

	Trigrams closest to 1 standard deviation above mean					
	Mean 95 <sup>th</sup> percentile	2 trigrams below	1 trigram below	Exactly 1 standard deviation above the mean	1 trigram above	2 trigrams above
	(1)	(2)	(3)	(4)	(5)	(6)
Physically able	assistant position available	fast paced fun	experience preferred necessary	work preferred necessary	required flexibility required	anything required help
Adaptable	making learning agility	must reliable apply	duties needed excellent	dedicated providing well	individual fast learner	moment respectful willing
Careful	computer skills required	basis important able	utmost professionalism integrity	processing payments must	requirements need really	assistant good verbal
Creative	position energetic detail	support vp marketing	office experience ability	involves spearheading independent	professional personal presentation	experience skills essential
Dependable	media reputation management	reliable transportation needed	service skills critical	hours retirees welcome	professional attitude well	experience excellent telephone
Personality	women feel con	correct visual image	attention detail outstanding	outgoing highly motivated	associate tribeca passion	luxury experience previous
Ability to learn	word excel must	sales skills managing	communication skills competency	career using cashiering	excellent people communication	abilities ability complete
Communication skills	politeness absolutely fundamental	backend customers required	skills abilities ability	client relation skills	filing typing computer	communication organizational skills
Experienced	evenings weekends requires	computer knowledge exciting	duties qualifications experience	level gain exposure	quickly become one	full time receptionist
Productive	fun eventful one	banker seeking ambitious	computer literate interested	tasks experience necessary	organization efficiency must	player good communication
Technology	paced integrative medical	ability learn new	willingness learn new	social media google	communication skills proficiency	computer skills proficiency

Note: For each stereotype, we construct the distribution of trigrams at the 95<sup>th</sup> percentile of all trigrams in a job ad. We identify the trigram that is closest to the mean 95<sup>th</sup> percentile trigram and the trigram that is closest to one standard deviation above the median. We also show the two trigrams immediately below 1 standard deviation above the mean 95<sup>th</sup> percentile (columns 2 and 3) and the two trigrams immediately above 1 standard deviation above the mean 95<sup>th</sup> percentile (columns 5 and 6). The CS scores in columns (2), (3), (5), and (6) were very close to those in column (4), for each row. For example, in the first row (Attractive), the CS scores in columns (2)–(6) ranged from 0.3772 to 0.3776. The CS score in column (1) was 0.3082.

**Appendix Table A2:**

**Analysis of Discrimination Against Younger Applicants**

	Female		Male		
	Admin (1)	Sales (2)	Janitor (3)	Sales (4)	Security (5)
<i>Baseline discrimination</i>	0.048	0.074	0.115	0.077	0.090

		Female		Male		
		Admin	Sales	Janitor	Sales	Security
		(1)	(2)	(3)	(4)	(5)
<i>Health stereotypes</i>		<i>Predicted sign</i>				
Attractive	Negative	0.004 (0.003)	-0.000 (0.010)	0.007 (0.022)	-0.010 (0.009)	-0.000 (0.012)
Hearing	Negative	-0.002 (0.002)	-0.009 (0.007)	<b>0.036</b> <sup>*</sup> <b>(0.015)</b>	-0.005 (0.006)	-0.005 (0.008)
Memory	Negative	-0.001 (0.003)	0.011 (0.008)	0.024 (0.020)	0.001 (0.008)	0.009 (0.009)
Physical Ability	Negative	-0.002 (0.003)	0.003 (0.009)	0.025 (0.022)	-0.006 (0.009)	0.001 (0.009)
<i>Personality stereotypes</i>		<i>Predicted sign</i>				
Adaptable	Negative	-0.003 (0.003)	<b>0.028</b> <sup>**</sup> <b>(0.010)</b>	0.021 (0.026)	0.008 (0.009)	-0.002 (0.011)
Careful	Positive	0.005 (0.003)	<b>-0.022</b> <sup>*</sup> <b>(0.009)</b>	-0.026 (0.022)	-0.004 (0.009)	0.002 (0.010)
Creative	Negative	-0.003 (0.003)	0.016 (0.008)	-0.020 (0.023)	-0.015 (0.008)	0.010 (0.012)
Dependable	Positive	<b>-0.006</b> <sup>*</sup> <b>(0.003)</b>	-0.001 (0.008)	0.003 (0.018)	-0.013 (0.008)	0.010 (0.009)
Personality	Positive/Negative	-0.002 (0.003)	-0.010 (0.008)	0.018 (0.023)	-0.001 (0.008)	0.004 (0.010)
<i>Skills stereotypes</i>		<i>Predicted sign</i>				
Ability to Learn	Negative	0.000 (0.003)	-0.007 (0.010)	<b>-0.056</b> <sup>*</sup> <b>(0.027)</b>	-0.004 (0.010)	-0.008 (0.012)
Communication Skills	Positive/Negative	0.005 (0.003)	-0.007 (0.011)	<b>0.085</b> <sup>*</sup> <b>(0.037)</b>	0.014 (0.011)	-0.023 (0.014)
Experienced	Positive	0.002 (0.002)	-0.003 (0.006)	-0.018 (0.015)	-0.008 (0.006)	0.005 (0.006)
Productive	Positive/Negative	-0.004 (0.003)	-0.003 (0.008)	-0.011 (0.025)	0.015 (0.009)	-0.018 (0.011)
Technology	Positive	0.002 (0.003)	-0.005 (0.008)	-0.035 (0.018)	-0.003 (0.007)	-0.001 (0.009)
N		14136	2834	646	3274	1804
Adjusted R <sup>2</sup>		0.035	0.040	0.082	0.042	0.093

Note: See notes to Table 4. The difference is that the outcome is now defined as discrimination against younger applicants. Note that the sample sizes do not add up because we do not observe reverse discrimination in all city-occupation cells, so those observations are excluded from the probit.

\* indicates that the estimate is statistically significant at the 5% level.

\*\* indicates that the estimate is statistically significant at the 1% level.

**Appendix Table A3:**

Comparing Words from Alternative Corpora

Stereotype	Complete Wikipedia		Spider-crawl Wikipedia		Strict string match	
	Word	CS Score	Word	CS Score	Word	
Attractive	unattractive	0.7458	desirable	0.6811	attractive	
	appealing	0.6966	advantageous	0.6603		
	desirable	0.6878	unattractive	0.6250		
	unappealing	0.6699	appealing	0.6150		
	alluring	0.6612	hospitable	0.5845		
	enticing	0.6471	palatable	0.5710		
	ideal	0.6290	favorable	0.5578		
	advantageous	0.6268	apt	0.5564		
	agreeable	0.6207	adaptable	0.5564		
	interesting	0.6175	comfortable	0.5479		
	Hearing	hearings	0.5888	judgment	0.4814	hearing
		testifying	0.5629	deaf	0.4631	
testimony		0.5403	hearings	0.4614		
arraignment		0.5272	tinnitus	0.4585		
pleading		0.5148	earplugs	0.4549		
sentencing		0.5078	testifying	0.4473		
testified		0.5051	cochlear	0.4447		
committal		0.4971	proceedings	0.4426		
questioning		0.4929	auditory	0.4397		
complaint		0.4853	trial	0.4374		
Memory	memories	0.6628	memories	0.7152	memory	
	brain	0.5274	amnesia	0.5725		
	recollection	0.5185	hippocampus	0.5631		
	cpu	0.5138	cognition	0.5625		
	remembering	0.5051	cognitive	0.5617		
	eidetic	0.4975	retrieval	0.5589		
	scratchpad	0.4933	recollection	0.5514		
	cache	0.4873	episodic	0.5478		
	rom	0.4862	perceptual	0.5418		
	consciousness	0.4770	reconsolidation	0.5093		
Physically able	unable	0.6887	unable	0.6910	physically able	
	enough	0.6050	willing	0.5993		
	willing	0.5912	enough	0.5800		
	trying	0.5904	psychologically	0.5680		
	ability	0.5865	trying	0.5513		
	needed	0.5696	unwilling	0.5365		

Stereotype	Complete Wikipedia		Spider-crawl Wikipedia		Strict string match
	Word	CS Score	Word	CS Score	Word
	attempting	0.5556	expected	0.5363	
	allowed	0.5549	anxious	0.5294	
	psychologically	0.5535	attempting	0.5281	
	unwilling	0.5505	eager	0.5271	
Number of articles	appx. 5,500,000		65,532		
Number of words	885,424		260,073		

Note: The table presents examples of result from training models on different corpora. Complete Wikipedia includes all the articles; it has 885,424 words and 200-dimensional vectors. Spider-crawl (scrapy spider) Wikipedia starts with articles referring to stereotyped3, ageism, and labor markets, as explained in the text. Both use 200-dimensional vectors. The “Word” column lists the words with the highest similarity scores; the scores are reported in the “CS Score” column.

## Online Appendix B: Supplemental Analyses

### Construction of Phrases

One robustness analysis we performed was to alter how many words we use in job-ad phrases. As our baseline, we used trigrams. We made this choice prior to analyzing the relationships between the phrases, stereotypes, and measured discrimination, to avoid the risk of cherry-picking phrase length to get stronger results. Nonetheless, we can use our same methods to construct similarity scores for phrases of any length.

Thus, subsequent to our baseline analysis, we redid our analysis using two-word and four-word phrases. The results, shown in Appendix Figures B1A and B1B, provide evidence that the results we obtained using three-word phrases are quite robust to using two- or four-word phrases.<sup>62</sup> The signs of the estimates are similar using different phrase lengths. For women, varying the length of phrases does not result in more significant estimates, indicating that the lack of strong associations between stereotyped language and discrimination against older women is not an artifact of phrase length. There is some heterogeneity in the statistical significance of the estimates, even though the estimates are similar in sign and magnitude. For men, the results from Figure 6 are very robust to changing the phrase length. Many of the stereotypes we found to predict discrimination using trigrams have similar estimated coefficients for predicting discrimination using either bigrams or quadgrams, although the statistical significance varies (less so for men than for women). Appendix Figure B1B suggests that the results are perhaps more robust when comparing trigrams to quadgrams and less robust when comparing bigrams to trigrams.

This may be because two-word phrases less reliably reflect the underlying stereotype – as evidence by the mean 95<sup>th</sup> percentile CS scores being lower for bigrams.<sup>63</sup>

<sup>62</sup>The regression results are reported in Appendix Tables B1A and B1B.

<sup>63</sup>In addition to this supplemental analysis, in Appendix Tables B2A and B2B and Appendix Figure B2, we replicate our baseline results using the median and the maximum CS scores rather than the 95<sup>th</sup> percentile. The results are consistent with the concerns we discussed earlier about using the median or maximum. In a number of age-occupation-cells, we find different stereotypes predicting discrimination at different percentiles. This variation is driven by two factors. First, the estimates (regardless of significance) are often very different at the median than they are in the upper tail, flipping signs when we move from the median to the 95<sup>th</sup> percentile, but

## Coding of Experience

We also explore modifying the method of characterizing age stereotypes in ads when an ad seeks more or less of a stereotypical characteristic. The semantic similarity model struggles when the text of a job ad features adverbs modifying the stereotyped characteristic. This seems most salient for the stereotype that older workers have more experience. Job ads that state they want a worker with ten years of experience will use language that is just as related to experience as job ads that state that they only require one to two years of experience or that no experience is required. Our results for experience are then simply the average direction of the relationship. However, because language associated with experience may positively affect the hiring of older workers and language associated with low levels of experience or no experience may negatively affect the hiring of older workers, our results are potentially biased towards zero. We therefore tried to modify the coding of experience to determine the impact of this potential problem.

In particular, we identified every sentence in the job ads with the word “experience.” We then searched for preceding or following words (10 words before or after) that indicated a context other than work experience (such as customer experience or shopping experience). We also capture instances of “commensurate” and “depending,” which typically appear in the text to indicate that pay is based on experience. We coded these instances as not requiring experience. We then identified all instances where the words “no” or “not” were used immediately before or after experience, which allowed us to capture phrases such as “no experience was necessary” or “experience not needed,” and coded these instances as not requiring experience. To code low experience requirements, we search ten words before or after the use of experience for any instance of “year/years.” If “year” or “years” is preceded by the number 1 or 2 and there are no numbers higher than 2, we code it as low experience required. No or low experience could, of course, indicate a preference for a younger worker.

These kinds of refinements of the semantic coding will always suffer from false positives and false negatives because the meanings of words are subtle, and combinations of words can give rise to different meanings, which algorithmic coding cannot definitively capture. After many rounds of refinements to arrive at the recoding described above, we found that our algorithm had a false positive rate of 24% for coding “no experience” and a false positive rate of 33% for coding “low experience.”<sup>64</sup> These false positive rates reflect the difficulty of designing this kind of algorithm to capture semantic meaning based on specific words or phrases. This difficulty, even in this reasonably simple context in which one might want to “override” the machine learning, demonstrated the value of the machine learning – even putting aside the clear advantages of machine learning for protecting from subjectivity and increasing replicability and transparency.<sup>65</sup>

---

much more rarely comparing results using the 95th percentile vs. the maximum. The implication is that more of the results at the median conflict with expectations based on the industrial psychology literature. This is not surprising for the median, where many of the phrases are unrelated to the stereotype. There are also some differences between the 95th percentile and the maximum, although rarely in terms of sign for the larger coefficient estimates. Looking back at Figure 4, this is not surprising, since the shapes of the distributions are sometimes very different for the maximum than for the 95th percentile.

<sup>64</sup>Our subjective coding of a random sample of ads, and comparisons to the algorithmic coding, are available upon request.

<sup>65</sup>Ningrum et al. (2020) explore a similar approach to classifying job ads by whether they include references to specific groups (e.g., gender or age), fine tuning how the language is read to better capture how these group identifiers are referenced. We strongly



To test how controlling for no or low experience requirements in a job ad impacts the estimates, we interact dummy variables for the job ad having no experience requirement ( $N_j$ ) or a low experience requirement ( $L_j$ ) with the cosine similarity score for experience ( $P_{jExp}^{95}$ ). We continue to control for the other thirteen stereotypes, so our model becomes:

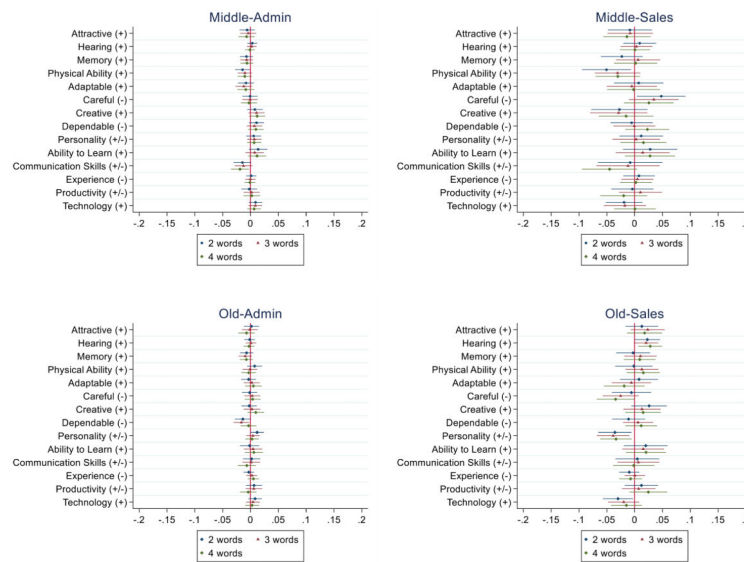
$$Pr\left[d_{exp}^{ij} = 1\right] = \alpha + \sum_s \beta_s P_{js}^{95} + \gamma_1 N_j + \theta_1 \left(P_{jExp}^{95} \times N_j\right) + \gamma_2 L_j + \theta_2 \left(P_{jExp}^{95} \times L_j\right) + X_{ij}\delta + \varepsilon_{ij} . \quad (B1)$$

Appendix Table B3A reports the results for women. In the baseline estimates, we found no significant relationship between stereotyped language related to experience and discrimination. We find no evidence that job ad language related to no or low experience is associated with discrimination, nor do we find evidence of interactions of indicators for low or no experience with the CS score with experience. Appendix Table B3B reports the results for men. In this case, our refinement produces more nuanced results in a handful of cases, but for many age-occupation clusters the results are the same. In the baseline estimates, we find no significant correlation between stereotyped language related to experience and discrimination, except for janitors (both middle-aged and old). When we introduce the controls and interactions for job ads using language indicating no or low experience required, we find significant increases in discrimination when the ad uses phrases that indicate low experience in retail sales for middle-aged men. Ads with this language increase the rate of discrimination by 19.5 percentage points. We find no evidence that language indicating no experience required is correlated with discrimination in any age-occupation cluster (for men or women).

When we allow the language related to experience to have different effects on discrimination when the job ad features language related to no or low experience, we find significant differences for middle-aged janitors and older men in retail sales. In both cases, the rates of discrimination are lower on ads that feature language highly related to experience and use phrases indicating they require only one to two years of experience, compared to the omitted group of language indicating that more experience is required. These somewhat non-intuitive results lead us to conclude that this exercise may not be reliable at providing better evidence on the effects of job ad language related to experience.<sup>66</sup>

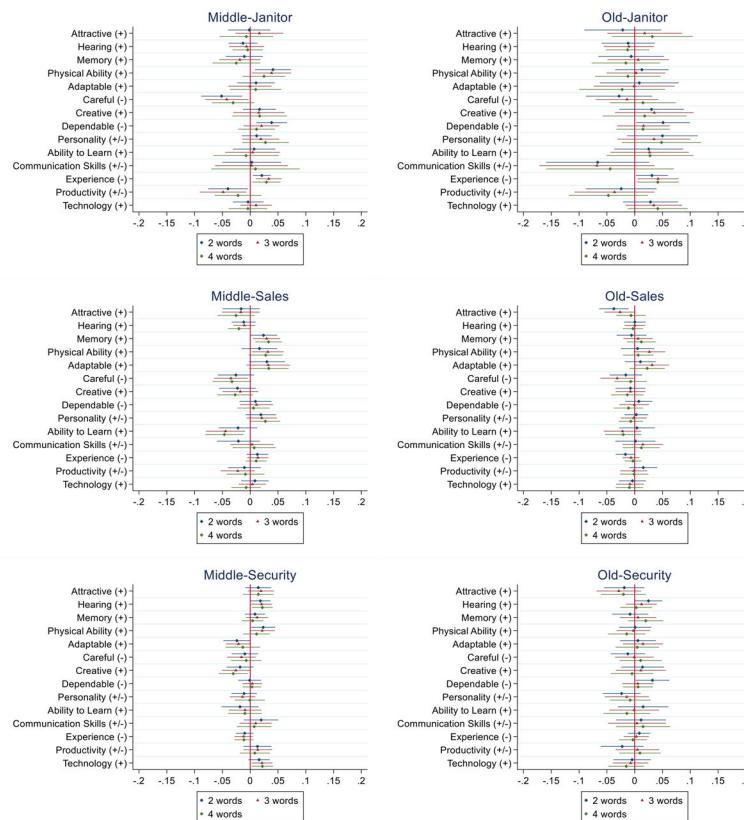
---

suspect that this approach would be far less suitable to characterizing the relationship between job-ad language and stereotypes, since stereotypes can be expressed in many ways, making it more important to measure linguistic similarity even when words do not match.<sup>66</sup> It is also the case, though, that at the end of the day this exercise may not matter too much, because our coding (based on a random sample) indicated that nearly three-quarters of job ads with language related to experience had the meaning of more experience required or preferred.



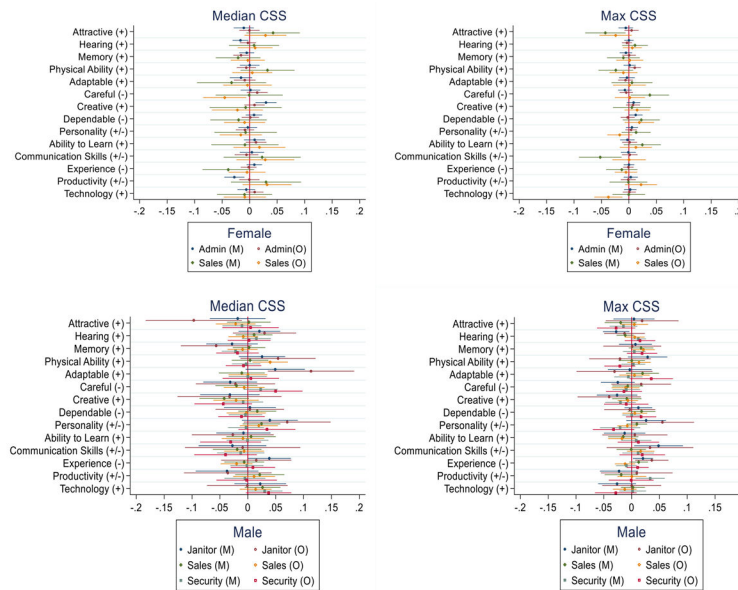
**Appendix Figure B1A: Varying the Number of Words in Phrases, Women**

Note: See notes to Figure 6. The difference is that here we vary the number of words in a phrase used to calculate the cosine similarity scores.



**Appendix Figure B1B: Varying the Number of Words in Phrases, Men**

Note: See notes to Figure 6. The difference is that here we vary the number of words in a phrase used to calculate the cosine similarity scores.



**Appendix Figure B2: Varying the Percentile**

Note: Figure reports estimated marginal effects from equation (10) and 95% confidence intervals, using the median CS score and the maximum CS score. Positive predictions indicate we expect to see more discrimination against older applicants, negative predictions indicate we expect to see less discrimination against older applicants. Standard errors are clustered at the job-ad level. The age of the sample is indicated in parentheses, M for middle-aged and O for old.

**Appendix Table B1A:**

Varying the Number of Words in Phrases, Women, Effects on Discrimination Against Older Applicants

# words in phrase	Middle-Admin			Middle-Sales			Old-Admin			Old-Sales		
	2	3	4	2	3	4	2	3	4	2	3	4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>Health stereotypes</i>												
Attractive (+)	-0.006 (0.007)	-0.004 (0.007)	-0.007 (0.007)	-0.009 (0.020)	-0.008 (0.021)	-0.014 (0.022)	0.002 (0.007)	-0.001 (0.007)	-0.007 (0.008)	0.013 (0.015)	0.023 (0.015)	0.018 (0.016)
Hearing (+)	0.003 (0.005)	0.002 (0.004)	-0.001 (0.005)	0.009 (0.015)	0.003 (0.015)	0.001 (0.014)	-0.002 (0.005)	0.001 (0.005)	-0.003 (0.005)	0.023 (0.012)	0.021 (0.011)	<b>0.028</b> <b>(0.011)</b>
Memory (+)	-0.007	-0.007	-0.006	-0.023	0.006	0.002	-0.007	-0.009	-0.007	-0.003	0.010	0.009

# words in phrase	Middle-Admin			Middle-Sales			Old-Admin			Old-Sales		
	2	3	4	2	3	4	2	3	4	2	3	4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Physical Ability (+)	(0.006) <b>-0.014*</b> <b>(0.007)</b>	(0.006) -0.010 (0.007)	(0.006) -0.010 (0.007)	(0.019) <b>-0.051*</b> <b>(0.022)</b>	(0.020) -0.031 (0.021)	(0.020) -0.030 (0.020)	(0.006) 0.008 (0.007)	(0.006) -0.001 (0.007)	(0.006) -0.003 (0.007)	(0.016) -0.001 (0.017)	(0.015) 0.013 (0.015)	(0.015) 0.016 (0.015)
<i>Personality stereotypes</i>												
Adaptable (+)	-0.008 (0.007)	-0.012 (0.008)	-0.008 (0.008)	0.007 (0.023)	-0.005 (0.023)	-0.002 (0.024)	-0.003 (0.007)	0.002 (0.007)	0.005 (0.008)	0.008 (0.017)	-0.006 (0.018)	-0.019 (0.019)
Careful (-)	-0.001 (0.007)	-0.001 (0.007)	-0.003 (0.007)	0.048* (0.022)	0.034 (0.023)	0.026 (0.023)	-0.001 (0.007)	0.003 (0.007)	0.004 (0.007)	-0.006 (0.018)	-0.025 (0.017)	<b>-0.034*</b> <b>(0.017)</b>
Creative (+)	0.008 (0.007)	0.011 (0.007)	0.012 (0.007)	-0.027 (0.026)	-0.029 (0.026)	-0.015 (0.025)	-0.002 (0.007)	0.003 (0.008)	0.010 (0.008)	0.026 (0.016)	0.013 (0.017)	0.015 (0.016)
Dependable (-)	0.011 (0.007)	0.007 (0.007)	0.010 (0.007)	-0.005 (0.019)	-0.001 (0.019)	0.023 (0.020)	-0.014 (0.007)	<b>-0.016*</b> <b>(0.007)</b>	-0.003 (0.007)	-0.011 (0.015)	0.006 (0.014)	0.012 (0.014)
Personality (+/-)	0.006 (0.007)	0.007 (0.006)	0.007 (0.006)	0.012 (0.020)	0.003 (0.022)	0.016 (0.021)	0.012 (0.006)	0.004 (0.006)	0.003 (0.006)	<b>-0.035*</b> <b>(0.015)</b>	<b>-0.039*</b> <b>(0.015)</b>	<b>-0.033*</b> <b>(0.014)</b>
<i>Skills stereotypes</i>												
Ability to Learn (+)	0.014 (0.008)	0.008 (0.008)	0.012 (0.008)	0.028 (0.025)	0.014 (0.025)	0.028 (0.023)	-0.002 (0.009)	0.005 (0.008)	0.006 (0.008)	0.020 (0.020)	0.015 (0.019)	0.020 (0.018)
Comm Skills (+/-)	-0.015 (0.008)	-0.013 (0.008)	<b>-0.019*</b> <b>(0.008)</b>	-0.008 (0.030)	-0.012 (0.029)	-0.045 (0.025)	0.002 (0.008)	0.001 (0.008)	-0.007 (0.008)	0.005 (0.020)	0.007 (0.019)	-0.002 (0.019)
Experienced (-)	0.001 (0.005)	-0.000 (0.005)	-0.001 (0.005)	0.008 (0.014)	0.005 (0.015)	0.002 (0.015)	-0.003 (0.005)	0.002 (0.005)	0.005 (0.005)	-0.010 (0.009)	0.000 (0.009)	-0.007 (0.010)
Productive (+/-)	-0.002 (0.007)	0.002 (0.007)	0.002 (0.008)	-0.004 (0.019)	0.010 (0.020)	-0.020 (0.022)	0.006 (0.007)	0.006 (0.008)	-0.004 (0.008)	0.012 (0.015)	0.007 (0.015)	0.024 (0.017)
Technology (+)	0.009 (0.006)	0.009 (0.006)	0.007 (0.006)	-0.019 (0.017)	-0.018 (0.019)	0.001 (0.019)	0.008 (0.006)	0.005 (0.006)	0.003 (0.006)	<b>-0.030*</b> <b>(0.014)</b>	-0.020 (0.014)	-0.015 (0.014)
N	6,822	6,822	6,821	986	986	986	7,321	7,321	7,320	1,856	1,856	1,856

Note: See notes to Table 4.

\* indicates that the estimate is statistically significant at the 5% level.

\*\* indicates that the estimate is statistically significant at the 1% level.

**Appendix Table B1B:**  
 Varying the Number of Words in Phrases, Men, Effects on Discrimination Against Older Applicants

# words in phrase	Middle-Junior			Middle-Sales			Middle-Security			Old-Junior			Old-Sales			Old-Security		
	2	3	4	2	3	4	2	3	4	2	3	4	2	3	4	2	3	4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
<i>Health stereotypes</i>																		
Attractive (+)	-0.002 (0.019)	0.017 (0.022)	-0.007 (0.024)	-0.016 (0.017)	-0.017 (0.017)	-0.025 (0.017)	0.015 (0.012)	0.019 (0.012)	0.015 (0.014)	-0.021 (0.035)	0.018 (0.034)	0.032 (0.037)	<b>-0.037</b> <b>**</b>	-0.026 (0.014)	-0.007 (0.013)	-0.019 (0.018)	-0.029 (0.020)	-0.020 (0.021)
Hearing (+)	-0.013 (0.013)	-0.007 (0.016)	-0.004 (0.014)	-0.012 (0.011)	-0.010 (0.010)	<b>-0.020</b> <b>*</b>	<b>0.018</b> <b>*</b>	<b>0.020</b> <b>*</b>	<b>0.022</b> <b>*</b>	-0.012 (0.024)	-0.010 (0.023)	-0.013 (0.020)	0.000 (0.010)	0.000 (0.009)	-0.003 (0.009)	<b>0.025</b> <b>*</b>	0.012 (0.014)	0.003 (0.015)
Memory (+)	-0.011 (0.017)	-0.019 (0.019)	-0.025 (0.022)	0.024 (0.012)	<b>0.030</b> <b>**</b>	<b>0.033</b> <b>**</b>	0.009 (0.009)	0.012 (0.010)	0.004 (0.010)	-0.006 (0.030)	0.006 (0.028)	-0.016 (0.031)	-0.006 (0.014)	0.006 (0.013)	0.012 (0.013)	-0.008 (0.017)	0.006 (0.017)	0.020 (0.016)
Physical Ability (+)	<b>0.041</b> <b>*</b>	<b>0.038</b> <b>*</b>	0.025 (0.020)	0.017 (0.016)	<b>0.032</b> <b>*</b>	<b>0.028</b> <b>*</b>	<b>0.023</b> <b>*</b>	0.022 (0.011)	0.011 (0.012)	0.013 (0.012)	0.002 (0.024)	-0.012 (0.027)	0.005 (0.030)	0.026 (0.015)	0.006 (0.015)	0.001 (0.014)	-0.002 (0.015)	-0.014 (0.015)
<i>Personality stereotypes</i>																		
Adaptable (+)	0.011 (0.017)	-0.000 (0.020)	0.010 (0.023)	0.030 (0.016)	0.032 (0.020)	0.033 (0.018)	-0.024 (0.012)	-0.021 (0.011)	-0.013 (0.014)	0.008 (0.036)	-0.001 (0.037)	-0.023 (0.039)	0.010 (0.014)	<b>0.031</b> <b>*</b>	0.023 (0.016)	0.006 (0.016)	0.015 (0.018)	0.005 (0.020)
Careful (-)	<b>-0.051</b> <b>**</b>	<b>-0.042</b> <b>*</b>	-0.031 (0.019)	-0.026 (0.017)	<b>-0.035</b> <b>*</b>	-0.033 (0.018)	-0.010 (0.012)	-0.016 (0.013)	-0.007 (0.014)	-0.028 (0.030)	-0.014 (0.029)	0.015 (0.030)	-0.016 (0.015)	<b>-0.031</b> <b>*</b>	-0.007 (0.015)	-0.012 (0.016)	-0.001 (0.018)	0.011 (0.019)
Creative (+)	0.017 (0.015)	0.015 (0.023)	0.017 (0.025)	-0.023 (0.017)	-0.018 (0.016)	-0.027 (0.017)	-0.018 (0.012)	<b>-0.026</b> <b>*</b>	<b>-0.030</b> <b>*</b>	0.030 (0.030)	0.035 (0.036)	0.018 (0.039)	-0.008 (0.014)	-0.008 (0.014)	-0.013 (0.015)	0.014 (0.020)	0.011 (0.023)	-0.005 (0.019)
Dependable (-)	<b>0.038</b> <b>**</b>	0.020 (0.016)	0.011 (0.017)	0.010 (0.015)	0.011 (0.015)	0.006 (0.015)	-0.001 (0.010)	0.004 (0.009)	0.003 (0.008)	<b>0.051</b> <b>*</b>	0.016 (0.024)	0.015 (0.024)	0.007 (0.012)	-0.001 (0.014)	-0.011 (0.013)	<b>0.032</b> <b>*</b>	0.006 (0.014)	0.006 (0.014)
Personality (+/-)	0.012 (0.014)	0.019 (0.017)	0.027 (0.021)	0.019 (0.014)	0.021 (0.014)	0.027 (0.014)	-0.011 (0.012)	-0.014 (0.012)	-0.001 (0.014)	0.050 (0.032)	0.035 (0.033)	0.048 (0.036)	0.003 (0.011)	-0.002 (0.012)	-0.007 (0.011)	-0.024 (0.017)	-0.014 (0.020)	-0.008 (0.019)
<i>Skills stereotypes</i>																		
Ability to Learn (+)	0.007 (0.019)	0.004 (0.025)	-0.007 (0.030)	-0.022 (0.018)	<b>-0.045</b> <b>*</b>	<b>-0.047</b> <b>**</b>	-0.018 (0.017)	-0.009 (0.015)	-0.010 (0.015)	0.025 (0.032)	0.027 (0.036)	0.028 (0.040)	0.004 (0.016)	-0.022 (0.017)	-0.021 (0.017)	0.015 (0.023)	-0.001 (0.023)	-0.014 (0.021)
Common Skills (+/-)	0.003 (0.027)	0.002 (0.033)	0.010 (0.040)	-0.021 (0.020)	0.003 (0.020)	0.007 (0.020)	0.020 (0.016)	0.010 (0.015)	0.007 (0.016)	-0.067 (0.047)	-0.068 (0.053)	-0.044 (0.059)	0.001 (0.018)	0.015 (0.019)	0.012 (0.017)	0.011 (0.023)	0.004 (0.027)	0.015 (0.025)
Experienced (-)	<b>0.021</b> <b>**</b>	<b>0.033</b> <b>**</b>	<b>0.029</b> <b>*</b>	0.013 (0.010)	0.014 (0.010)	0.011 (0.010)	-0.010 (0.008)	-0.012 (0.008)	-0.011 (0.009)	<b>0.031</b> <b>*</b>	<b>0.042</b> <b>*</b>	<b>0.042</b> <b>*</b>	-0.017 (0.009)	-0.007 (0.008)	-0.003 (0.008)	0.008 (0.010)	0.003 (0.011)	-0.003 (0.013)
Productive (+/-)	<b>-0.040</b> <b>*</b>	<b>-0.049</b> <b>*</b>	-0.022 (0.021)	-0.010 (0.015)	-0.022 (0.016)	-0.009 (0.017)	0.013 (0.013)	0.014 (0.013)	0.008 (0.014)	-0.024 (0.032)	-0.036 (0.037)	-0.047 (0.036)	0.015 (0.013)	-0.002 (0.012)	-0.001 (0.013)	-0.023 (0.020)	0.005 (0.020)	0.009 (0.019)
Technology (+)	-0.004 (0.014)	0.011 (0.014)	-0.004 (0.018)	0.009 (0.012)	0.004 (0.012)	-0.007 (0.013)	0.016 (0.010)	<b>0.021</b> <b>*</b>	<b>0.022</b> <b>*</b>	0.028 (0.009)	0.034 (0.010)	0.042 (0.025)	-0.004 (0.026)	-0.009 (0.027)	-0.009 (0.013)	-0.005 (0.012)	-0.007 (0.017)	-0.015 (0.016)
N	311	311	311	1,612	1,612	1,610	954	954	953	318	318	318	1,680	1,680	1,680	932	932	931

Note: See notes to Table 4.

\* indicates that the estimate is statistically significant at the 5% level.

\*\* indicates that the estimate is statistically significant at the 1% level.

Appendix Table B2A:

Varying the Percentile, Women, Effects on Discrimination Against Older Applicants

Percentile	Middle-Admin			Middle-Sales			Old-Admin			Old-Sales		
	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>Health stereotypes</i>												
Attractive (+)	-0.011 (0.009)	-0.004 (0.007)	-0.006 (0.007)	0.043 (0.025)	-0.008 (0.021)	<b>-0.043</b> <sub>*</sub> <b>(0.019)</b>	-0.000 (0.009)	-0.001 (0.007)	0.005 (0.007)	0.029 (0.019)	0.023 (0.015)	-0.024 (0.015)
Hearing (+)	<b>-0.017</b> <b>(0.007)</b>	0.002 (0.004)	0.000 (0.004)	0.008 (0.023)	0.003 (0.015)	0.011 (0.012)	-0.003 (0.007)	0.001 (0.005)	-0.004 (0.005)	0.010 (0.016)	0.021 (0.011)	0.006 (0.009)
Memory (+)	-0.005 (0.007)	-0.007 (0.006)	-0.005 (0.006)	-0.021 (0.021)	0.006 (0.020)	-0.010 (0.015)	<b>-0.016</b> <sub>*</sub> <b>(0.007)</b>	-0.009 (0.006)	0.000 (0.006)	-0.003 (0.016)	0.010 (0.015)	0.002 (0.013)
Physical Ability (+)	0.000 (0.009)	-0.010 (0.007)	0.001 (0.006)	0.032 (0.025)	-0.031 (0.021)	-0.024 (0.016)	-0.006 (0.009)	-0.001 (0.007)	0.011 (0.006)	0.005 (0.019)	0.013 (0.015)	-0.010 (0.013)
<i>Personality stereotypes</i>												
Adaptable (+)	-0.016 (0.010)	-0.012 (0.008)	-0.004 (0.007)	-0.033 (0.032)	-0.005 (0.023)	0.006 (0.019)	-0.009 (0.010)	0.002 (0.007)	-0.006 (0.007)	-0.004 (0.023)	-0.006 (0.018)	0.001 (0.015)
Careful (-)	0.002 (0.009)	-0.001 (0.007)	-0.007 (0.006)	-0.001 (0.031)	0.034 (0.023)	<b>0.038</b> <sub>*</sub> <b>(0.018)</b>	0.014 (0.010)	0.003 (0.007)	-0.005 (0.006)	<b>-0.045</b> <sub>*</sub> <b>(0.020)</b>	-0.025 (0.017)	0.002 (0.014)
Creative (+)	<b>0.030</b> <sub>**</sub> <b>(0.010)</b>	0.011 (0.007)	0.009 (0.006)	-0.007 (0.033)	-0.029 (0.026)	0.005 (0.018)	0.009 (0.009)	0.003 (0.008)	0.008 (0.007)	-0.022 (0.024)	0.013 (0.017)	0.015 (0.012)
Dependable (-)	0.008 (0.008)	0.007 (0.007)	0.012 (0.006)	-0.020 (0.026)	-0.001 (0.019)	0.023 (0.017)	0.001 (0.008)	<b>-0.016</b> <sub>*</sub> <b>(0.007)</b>	-0.002 (0.007)	-0.009 (0.019)	0.006 (0.014)	0.019 (0.014)
Personality (+/-)	-0.003 (0.009)	0.007 (0.006)	0.006 (0.006)	-0.007 (0.029)	0.003 (0.022)	0.013 (0.014)	-0.008 (0.008)	0.004 (0.006)	0.004 (0.005)	-0.016 (0.020)	<b>-0.039</b> <sub>**</sub> <b>(0.015)</b>	-0.017 (0.012)
<i>Skills stereotypes</i>												
Ability to Learn (+)	0.009 (0.010)	0.008 (0.008)	-0.003 (0.007)	-0.009 (0.031)	0.014 (0.025)	0.024 (0.017)	0.012 (0.010)	0.005 (0.008)	0.001 (0.007)	0.018 (0.024)	0.015 (0.019)	0.013 (0.015)
Comm Skills (+/-)	0.004 (0.011)	-0.013 (0.008)	-0.001 (0.007)	0.023 (0.036)	-0.012 (0.029)	<b>-0.052</b> <sub>**</sub> <b>(0.020)</b>	-0.006 (0.011)	0.001 (0.008)	0.002 (0.007)	0.028 (0.027)	0.007 (0.019)	0.000 (0.016)
Experienced (-)	0.009 (0.007)	-0.000 (0.005)	0.000 (0.005)	-0.039 (0.024)	0.005 (0.015)	-0.013 (0.015)	-0.002 (0.007)	0.002 (0.005)	-0.001 (0.005)	-0.005 (0.017)	0.000 (0.009)	-0.005 (0.010)

Percentile	Middle-Admin			Middle-Sales			Old-Admin			Old-Sales		
	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Productive (+/-)	<b>-0.028</b> **	0.002	0.003	0.030	0.010	-0.001	-0.001	0.006	-0.001	0.032	0.007	0.022
	(0.009)	(0.007)	(0.007)	(0.033)	(0.020)	(0.018)	(0.009)	(0.008)	(0.007)	(0.023)	(0.015)	(0.015)
Technology (+)	-0.006	0.009	0.003	-0.009	-0.018	-0.000	0.009	0.005	0.001	-0.009	-0.020	<b>-0.037</b> **
	(0.008)	(0.006)	(0.006)	(0.025)	(0.019)	(0.015)	(0.008)	(0.006)	(0.006)	(0.020)	(0.014)	(0.013)
N	6,822	6,822	6,827	986	986	987	7,321	7,321	7,330	1,856	1,856	1,861

Note: See notes to Table 4. There are sometimes more observations for the maximum because of ads with small numbers of trigrams for which percentiles could not be accurately calculated.

\* indicates that the estimate is statistically significant at the 5% level.

\*\* indicates that the estimate is statistically significant at the 1% level.

### Appendix Table B2B:

#### Varying the Percentile, Men, Effects on Discrimination Against Older Applicants

Percentile	Middle-Janitor			Middle-Sales			Middle-Security			Old-Janitor			Old-Sales			Old-Security		
	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
<i>Health stereotypes</i>																		
Attractive (+)	-0.018 (0.026)	0.017 (0.022)	0.004 (0.019)	0.002 (0.020)	-0.017 (0.017)	-0.020 (0.015)	-0.010 (0.018)	0.019 (0.012)	-0.015 (0.013)	<b>-0.097</b> **	0.018 (0.034)	0.019 (0.033)	-0.022 (0.018)	-0.026 (0.014)	0.005 (0.013)	0.005 (0.026)	-0.029 (0.020)	-0.028 (0.017)
Hearing (+)	0.021 (0.019)	-0.007 (0.016)	<b>0.012</b> (0.017)	0.011 (0.017)	-0.010 (0.010)	-0.011 (0.008)	0.015 (0.013)	<b>0.029</b> (0.010)	0.012 (0.007)	0.030 (0.029)	-0.010 (0.023)	-0.013 (0.019)	-0.008 (0.014)	0.000 (0.009)	0.006 (0.009)	0.002 (0.020)	0.012 (0.014)	0.015 (0.014)
Memory (+)	-0.028 (0.024)	-0.019 (0.019)	0.007 (0.015)	0.002 (0.015)	<b>0.030</b> (0.012)	0.017 (0.011)	-0.022 (0.013)	0.012 (0.010)	0.005 (0.010)	-0.057 (0.032)	0.006 (0.028)	0.001 (0.027)	-0.009 (0.014)	0.006 (0.013)	<b>0.021</b> (0.010)	-0.019 (0.019)	0.006 (0.017)	0.019 (0.014)
Physical Ability (+)	0.026 (0.021)	<b>0.038</b> (0.018)	0.029 (0.018)	0.004 (0.017)	<b>0.032</b> (0.015)	0.000 (0.012)	-0.003 (0.013)	0.022 (0.012)	0.003 (0.011)	0.054 (0.034)	0.002 (0.027)	-0.021 (0.028)	<b>0.040</b> (0.016)	0.026 (0.015)	0.013 (0.011)	-0.008 (0.016)	-0.002 (0.015)	-0.021 (0.014)
<i>Personality stereotypes</i>																		
Adaptable (+)	0.049 (0.027)	-0.000 (0.020)	-0.003 (0.020)	-0.011 (0.021)	0.022 (0.020)	0.020 (0.015)	0.000 (0.017)	-0.021 (0.011)	-0.006 (0.013)	<b>0.113</b> **	-0.001 (0.037)	-0.031 (0.035)	0.000 (0.018)	<b>0.031</b> (0.016)	0.005 (0.014)	0.005 (0.026)	0.015 (0.018)	0.025 (0.020)
Careful (-)	-0.032 (0.025)	<b>0.012</b> (0.020)	-0.025 (0.016)	-0.020 (0.019)	<b>0.035</b> (0.015)	-0.007 (0.014)	0.023 (0.016)	-0.016 (0.013)	-0.009 (0.033)	-0.022 (0.036)	-0.014 (0.029)	0.017 (0.027)	-0.007 (0.019)	-0.031 (0.015)	-0.011 (0.013)	<b>0.050</b> (0.025)	-0.001 (0.018)	-0.014 (0.016)
Creative (+)	-0.032 (0.027)	0.015 (0.023)	-0.026 (0.019)	-0.042 (0.023)	-0.018 (0.016)	-0.009 (0.013)	-0.008 (0.018)	<b>0.026</b> (0.013)	-0.020 (0.014)	-0.033 (0.048)	0.035 (0.036)	-0.040 (0.029)	-0.021 (0.021)	-0.008 (0.014)	-0.007 (0.009)	-0.044 (0.026)	0.011 (0.023)	-0.010 (0.020)
Dependable (-)	0.004 (0.024)	0.020 (0.016)	0.012 (0.013)	0.017 (0.017)	0.011 (0.015)	0.018 (0.013)	-0.004 (0.015)	0.004 (0.009)	0.000 (0.009)	0.004 (0.031)	0.016 (0.024)	-0.003 (0.024)	-0.003 (0.018)	-0.001 (0.014)	0.006 (0.011)	-0.011 (0.021)	0.006 (0.014)	0.017 (0.014)
Personality (+/-)	0.039 (0.026)	0.019 (0.017)	0.026 (0.018)	0.025 (0.017)	0.021 (0.014)	0.009 (0.009)	-0.003 (0.006)	-0.014 (0.006)	-0.021 (0.006)	0.071 (0.040)	0.035 (0.033)	0.055 (0.029)	0.020 (0.017)	-0.002 (0.012)	-0.008 (0.009)	0.034 (0.026)	-0.014 (0.020)	-0.032 (0.019)
<i>Skills stereotypes</i>																		
Ability to Learn (+)	-0.008 (0.025)	0.004 (0.025)	-0.013 (0.019)	0.006 (0.023)	<b>0.045</b> (0.018)	-0.015 (0.014)	-0.000 (0.019)	-0.009 (0.015)	0.008 (0.014)	-0.027 (0.037)	0.027 (0.036)	0.006 (0.030)	-0.009 (0.019)	-0.022 (0.017)	-0.017 (0.013)	-0.031 (0.028)	-0.001 (0.023)	0.012 (0.019)



Percentile	Middle-Junior			Middle-Sales			Middle-Security			Old-Junior			Old-Sales			Old-Security		
	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max	Median	95 <sup>th</sup>	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Comm Skills (+/-)	-0.028 (0.031)	0.002 (0.033)	<b>0.048</b> **	-0.019 (0.022)	0.003 (0.020)	-0.001 (0.013)	-0.013 (0.022)	0.010 (0.015)	0.012 (0.014)	-0.009 (0.053)	-0.068 (0.053)	0.033 (0.040)	-0.007 (0.018)	0.015 (0.019)	0.016 (0.013)	-0.040 (0.029)	0.004 (0.027)	0.019 (0.021)
Experienced (-)	0.039 (0.020)	<b>0.033</b> **	<b>0.020</b> **	-0.006 (0.018)	0.014 (0.010)	0.013 (0.009)	-0.002 (0.015)	-0.012 (0.008)	-0.009 (0.010)	0.015 (0.032)	<b>0.042</b> **	<b>0.036</b> **	-0.020 (0.015)	-0.007 (0.008)	-0.012 (0.009)	0.009 (0.020)	0.003 (0.011)	0.011 (0.010)
Productive (+/-)	-0.037 (0.029)	<b>-0.049</b> **	-0.023 (0.017)	0.021 (0.022)	-0.022 (0.016)	-0.019 (0.015)	-0.005 (0.019)	0.014 (0.013)	<b>0.033</b> **	-0.036 (0.040)	-0.036 (0.037)	0.009 (0.033)	0.011 (0.019)	-0.002 (0.012)	0.000 (0.013)	-0.002 (0.028)	0.005 (0.020)	-0.001 (0.021)
Technology (+)	0.022 (0.024)	0.011 (0.014)	-0.026 (0.017)	0.026 (0.016)	0.004 (0.012)	0.002 (0.011)	0.030 (0.016)	<b>0.021</b> **	0.002 (0.009)	-0.001 (0.012)	0.034 (0.026)	0.000 (0.027)	0.014 (0.015)	-0.009 (0.013)	-0.012 (0.011)	0.037 (0.021)	-0.007 (0.016)	-0.028 (0.019)
N	311	311	311	1,612	1,612	1,612	954	954	956	318	318	318	1,680	1,680	1,680	932	932	932

Note: See notes to Table 4 and Appendix Table B2A.

\* indicates that the estimate is statistically significant at the 5% level.

\*\* indicates that the estimate is statistically significant at the 1% level.

**Appendix Table B3A:**

**Refinements to the Coding of Experience, Women, Effects on Discrimination Against Older Applicants**

	Middle-Admin			Middle-Sales			Old-Admin			Old-Sales		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Experience CS Score	-0.000 (0.005)	-0.001 (0.005)	-0.002 (0.005)	0.005 (0.015)	0.011 (0.015)	0.013 (0.016)	0.002 (0.005)	0.004 (0.005)	0.004 (0.005)	0.000 (0.009)	0.006 (0.010)	0.007 (0.010)
No Experience Needed		-0.002 (0.015)	-0.002 (0.015)		-0.041 (0.038)	-0.043 (0.038)		-0.029 (0.017)	-0.029 (0.017)		-0.032 (0.025)	-0.033 (0.025)
Experience CS Score × No Experience Needed		0.017 (0.015)	0.017 (0.015)		-0.015 (0.030)	-0.014 (0.030)		0.020 (0.015)	0.020 (0.015)		0.006 (0.027)	0.007 (0.027)
1-2 Years Experience			0.053 (0.039)			-0.051 (0.059)			0.006 (0.033)			-0.032 (0.056)
Experience CS Score × 1-2 Years Experience			-0.003 (0.024)			0.014 (0.065)			-0.012 (0.022)			-0.038 (0.039)
N	6822	6822	6822	986	986	986	7321	7321	7321	1856	1856	1856

Note: Table reports estimated marginal effects. For each age-occupation group, the first column estimates equation (10), but reports only the coefficient for the 95<sup>th</sup> percentile of Experience. In the second column, we include a dummy the ad indicating no experience was needed and an interaction between the no experience needed dummy and the Experience CS score. In the third columns, we add a dummy for only requiring one to two years of experience and the interaction between the low experience dummy and the Experience CS score (equation (10)). Standard errors are clustered at the job ad level.

\* indicates that the estimate is statistically significant at the 5% level.

\*\* indicates that the estimate is statistically significant at the 1% level.

**Appendix Table B3B:**

Refinements to the Coding of Experience, Men, Effects on Discrimination Against Older Applicants

	Middle-Junior			Middle-Sales			Middle-Security			Old-Junior			Old-Sales			Old-Security		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Experience CS Score	0.033**	0.034**	0.033**	0.014	0.012	0.011	-0.012	-0.010	-0.009	0.042*	0.041*	0.041*	-0.007	-0.007	-0.008	0.003	-0.002	-0.004
	(0.012)	(0.012)	(0.012)	(0.010)	(0.010)	(0.010)	(0.008)	(0.009)	(0.009)	(0.019)	(0.019)	(0.019)	(0.008)	(0.008)	(0.008)	(0.011)	(0.012)	(0.013)
No Experience Needed		-0.021	-0.022		0.014	0.006		-0.020	-0.018		-0.034	-0.033		0.003	-0.000		0.033	0.034
		(0.034)	(0.033)		(0.024)	(0.025)		(0.027)	(0.027)		(0.057)	(0.056)		(0.020)	(0.020)		(0.039)	(0.039)
Experience CS Score + No Experience Needed		0.070	0.064		0.010	0.014		0.014	0.013		0.097	0.098		0.020	0.024		0.041	0.040
		(0.060)	(0.058)		(0.023)	(0.024)		(0.026)	(0.026)		(0.076)	(0.077)		(0.023)	(0.023)		(0.038)	(0.038)
1-2 Years Experience			-0.054			0.195*			-0.029			-0.012			0.051			0.011
			(0.046)			(0.096)			(0.037)			(0.124)			(0.035)			(0.077)
Experience CS Score + 1-2 Years Experience			-0.038*			-0.045			-0.010			-0.011			-0.049*			0.061
			(0.019)			(0.030)			(0.027)			(0.073)			(0.020)			(0.061)
N	311	311	311	1,612	1,612	1,610	954	954	953	318	318	318	1,680	1,680	1,680	932	932	931

Note: See note to Appendix Table B3A.

\* indicates that the estimate is statistically significant at the 5% level.  
 \*\* indicates that the estimate is statistically significant at the 1% level.

**Online Appendix C: Technical Appendix**

This appendix provides explanation and code for our core methods.

**1. Extracting the Trigrams**

We begin by identifying all the trigrams in a job ad using the following python code. This proceeds in a number of steps.

First, the code below reads the file “Callback.csv” (which contains the text of all the ads saved as part of NBB) containing all the ads and does some manipulation in preparation for further processing.<sup>i</sup>

<sup>i</sup>A user of this code will of course have to change the file names and folders.

```
dfads = pd.read_csv("D:\RAWorkAgeDiscrim\ML\CallBack.csv")
dfads.head()

# Substitute blank spaces with x so the program runs later

dfads["Body"] = dfads["Body"].replace(" ", "x", regex=True)
dfads["Body"] = dfads["Body"].replace(np.nan, "x", regex=True)

# Dropping duplicates

dfads=dfads.drop_duplicates(subset="jobid", keep="first")

#Fixing index
dfads.index = np.arange(0,len(dfads))

# Keep only job id and body
match = dfads[["jobid", "Body"]]
```

Second, we create the program *review\_to\_words* that eliminates spaces, caps, punctuation, and stop words.

```
def review_to_words(add):
    # Function to convert a raw review to a string of words
    # The input is a single string (a raw movie review), and
    # the output is a single string (a preprocessed movie review)
    #
    # 1. Remove HTML (I don't need this part already cleaned html)
    #review_text = BeautifulSoup(raw_review).get_text()
    #
    # 2. Remove non-letters
    letters_only = re.sub("[^a-zA-Z]", "", add)
    #
    # 3. Convert to lower case, split into individual words
    words = letters_only.lower().split()
    #
    # 4. In Python, searching a set is much faster than searching
    # a list, so convert the stop words to a set
    stops = set(stopwords.words("english"))
    #
    # 5. Remove stop words
    meaningful_words = [w for w in words if not w in stops]
    #
    # 6. Join the words back into one string separated by space,
    # and return the result.
    return(" ").join(meaningful_words)

num_adds = match["Body"].size
```

Third, we run the program to clean each ad. This means that we are eliminating symbols, non-English characters, punctuation, and stop words. The objective is to keep only relevant words that convey meaning potentially related to stereotypes.

```
clean_add = review_to_words( match["Body"][2] )
print clean_add
print match["Body"][2]

print "Cleaning and parsing the training set movie reviews..."
clean_train_adds = []
for i in xrange( 0, num_adds ):
    # If the index is evenly divisible by 1000, print a message
    if( (i+1)%1000 == 0 ):
        print "Review %d of %d\n" % ( i+1, num_adds )
        clean_train_adds.append( review_to_words( match["Body"][i] ) )
match["Body_Clean"] = clean_train_adds
```

Fourth, we import *nlTK* to make trigrams for each ad.

```
# Creating trigrams
#Note that you need to download> nltk.download("punkt")

from nltk.util import ngrams
from nltk import word_tokenize
```

Finally, in the code below, in the first line, using `ngrams(word_tokenize())`, we separate all the trigrams in each ad. The line `match = match[["jobid", "trigram"]]` separates trigrams by job ad, creating a row with each trigram separated by commas. The following line outputs them into an *csv* file with a column for the trigrams in each ad.

```
match["trigram"] = match["Body_Clean"].apply(lambda row: list(ngrams(word_tokenize(row),
3)))

match = match[["jobid", "trigram"]]

match.to_excel("D:\\RAWorkAgeDiscrim\\SScoresMatch\\matchtrigram.xlsx")
```

Appendix Figure C1, Panels A and B, show the files created at these two steps, with the trigrams separated by commas (Panel A) and the trigrams then split into columns (Panel B).

*Panel A. Raw Output Panel*

A	B	C	D	E	F
	jobid	trigram			
0	4826529596	[(music, fest, time), (fest, time, year), (tir			
1	4833493607	[(flower, patch, currently), (patch, curren			
2	4833543304	[(seeking, career, change), (career, chang			
3	4835083297	[(interior, decorator, trainee), (decorator,			
4	4835197759	[(opportunity, become, key), (become, ke			
5	4841996184	[(furniture, warehouse, currently), (warel			
6	4843542824	[(looking, person, least), (person, least, y			
7	4843588254	[(seeking, exceptional, hotel), (exception			
8	4843935707	[(local, retail, store), (retail, store, seekin			
9	4844967381	[(yale, enforcement, services), (enforcem			
10	4845013578	[(dealership, parts, counter), (parts, coun			
11	4845385448	[(seasonal, part, time), (part, time, front),			

*B. Arranged in Columns*

A	B	C	D	E	F
	jobid	trigram			
0	4826529596	music fest fest time	time year	year right	
1	4833493607	flower pa patch curr	currently	seeking q	
2	4833543304	seeking ca career cha	change lo	looking ge	
3	4835083297	interior de decorator	trainee fu	furniture	
4	4835197759	opportuni become k	key memt	member c	
5	4841996184	furniture warehouse	currently	seeking hi	
6	4843542824	looking pe person le	least year	years den	
7	4843588254	seeking e; exception	hotel fron	front desk	
8	4843935707	local retail retail stor	store seek	seeking re	
9	4844967381	yale enfor enforcem	services s	seeking ui	
10	4845013578	dealership parts cour	counter p	position n	
11	4845385448	seasonal part time	time front	front desk	

**Appendix Figure C1:**  
Trigrams from Job Ads

## 2. Calculating the Semantic Similarity between Words using *Word2Vec*

*Word2Vec* is an algorithm that calculates vectors for a vocabulary and estimates the relationships between different words. The steps to use the *Word2Vec* algorithm are:

1. Download all Wikipedia articles<sup>ii</sup> (or the corresponding corpus used) and store the information on a computer's hard drive.
2. Read all the articles and separate them into paragraphs and sentences.

<sup>ii</sup>The database (*Wikipedia dump*) articles are available at <https://dumps.wikimedia.org/enwiki/>.

3. Clean text: drop all HTML code, punctuation, and symbols; and substitute caps for lowercase.
4. The python module *gensim* then generates all the vectors.
5. Identify the words used together in sentences and paragraphs and count the number of times that a pair of words appear together, for all the words in the corpus.
6. Estimate the probability that two words appear together.
7. Create 200-dimensional vectors for each word. The 200 dimensions are a measure of the vocabulary that describes the characteristics of each word. The method simplifies all the information using hidden layers, and instead of having all the words as a context, we only need 200 weights. In this step, similar words (because they share the same context) have similar vectors; as a result, the cosine similarity score (CSS) between them will be closer to one.
8. Store the vectors in a matrix for future use.

The code to do this is now described in more detail. All the coding is in Python 2.7. Use “—future—” to make compatible all the coding to Python 3.x.

This first section of the code imports all the modules needed to run the code correctly, and takes some other initial steps.

*gensim*: Train the model on Wikipedia. For information on how the genism program works, please see <https://radimrehurek.com/gensim/>.<sup>iii</sup>

*logging*: This has multiple purposes. Logging retains information in some parts of the code, like usernames, passwords, the directory where the files are, etc. It can print part of the process, give you warnings, and restart the kernel (python) and other functions. All these errors and warnings are stored and printed. It is not relevant to the algorithm that we are using.

*multiprocessing*: This is used to scan the computer’s number of processors (workers in python) and use the maximum number to compute the estimations of *Word2Vec* more quickly.

*os*: This is necessary to navigate through the folder and files on one’s computer.

*re*: This includes operations like detecting caps, symbols, and numbers and dropping them, adding them, replacing them, etc.

*sys*: This module provides access to some variables used or maintained by the interpreter and functions that interact strongly with the interpreter. It is always available. In our code it is used to read the pathname of the operating system.

---

<sup>iii</sup>In our view, this is the best source for understanding the *Word2Vec* algorithm. It both provides code and references to many related articles.

```

from __future__ import print_function
import gensim
import logging
import multiprocessing
import os
import re
import sys

```

“*nlk* is imported specifically to tokenize all the paragraphs and sentences in Wikipedia’s articles.

*nlk.data.load*(“tokenizers/punkt/english.pickle”): This sets English as the language to analyze the text.

*time*: This times the process.

*logging.basicConfig*: This sets up the format, and sets alerts, warnings, and errors (coded as levelname).

```

import nltk.data
tokenizer = nltk.data.load("tokenizers/punkt/english.pickle")
from time import time

logging.basicConfig(format='%(asctime)s : %(levelname)s : %(message)s',
                    level=logging.INFO)

```

We define a function that cleans all the text in Wikipedia’s articles, eliminating punctuation, some HTML code, and keeping only the relevant text.

*def*: This indicates that we are defining a new function.

*re.compile*: This creates a pattern of the symbols to search for and match in the html code.

*re.sub*: This part of the code looks for the compiled pattern created in *re.compile*, and replaces it with “blank.” If the pattern isn’t found, it returns the desired text.

```

def cleanhtml(raw_html):
    cleanr = re.compile('<.*?>')
    cleantext = re.sub(cleanr, "", raw_html)
    return cleantext

```

Next, we define the Windows path where all the Wikipedia files are saved (usually on the hard drive). The program will ask for the directory at the beginning. After that, it extracts all the articles from a “dump” file available in *Wikimedia dump* (see footnote iii).

```

class MySentences(object):
    def __init__(self, dirname):
        self.dirname = dirname

```

Next, we put together some functions to read all the articles to clean paragraphs and sentences.

*for root*: This is a loop through the directories and files on the folder where the Wikipedia files are saved.

*for filename in files*: This loops through the files.

*sline*: This separates the articles into sentences.

*cleanhtml*: This cleans the sentences of each article.

```
def __iter__(self):
    for root, dirs, files in os.walk(self.dirname):
        for filename in files:
            file_path = root + '/' + filename
            for line in open(file_path):
                sline = line.strip()
                if sline == " ":
                    continue
                rline = cleanhtml(sline)
```

Once the sentences are clean, they are tokenized (split) into words.

*Tokenizer.tokenize()*: split sentences into words.

The rest of the lines replace caps with lowercase and drop symbols.

```
tokenized_line = ''.join(tokenizer.tokenize(rline))
is_alpha_word_line = [word for word in
                      tokenized_line.lower().split()
                      if word.isalpha()]
yield is_alpha_word_line

if __name__ == '__main__':
    if len(sys.argv) != 2:
        print("Please use python train_with_gensim.py data_path")
        exit()
    data_path = sys.argv[1]
    begin = time()
```

We now explain the *Word2Vec* model, which transforms all the words and sentences into vectors. We use *gensim* (a python module) to do this. Our parameters are as follows:

*size* = 200 dimensions. Generally, to analyze semantic similarity with massive databases like Wikipedia, the recommended size is between 100 and 150 dimensions (nodes); see Pennington et al. (2014). The more nodes, the more precise the model will be. We picked 200 nodes to increase the precision of the semantic similarity; we did not exceed 200 because it is computationally cumbersome.

*window* = 10. The window defines how many words before and after the target word are used to estimate the probability that they are together in the same context. It is recommended in user forums (at <https://stackoverflow.com/>) to use 10 words (the default). If one selects fewer words, then the CSS will be more strict. Only words that are very close in a sentence are related.

*workers* = multiprocessing.cpu\_count(). This defines how many cores of the processor are used to calculate everything. "multiprocessing.cpu\_count()" identifies the number of cores in the computer and uses it; the code results in using all the cores available in the computer to calculate the vectors faster.



```

sentences = MySentences(data_path)
model = gensim.models.Word2Vec(sentences,
    size=200,
    window=10,
    min_count=10,
    workers=multiprocessing.cpu_count())
model.save("data/model/word2vec_gensim")

model.wv.save_word2vec_format("D:\RAWorkAgeDiscrim\Wikicorpus\wikiextractor\word2vec_org",
    "D:\RAWorkAgeDiscrim\Wikicorpus\wikiextractor\vocabulary", binary=False)
    
```

The last line in the code estimates how many seconds it took to calculate all the vectors.

```

end = time()
print ("Total procesing time: %d seconds" % (end - begin))
    
```

The calculations are saved as a large matrix. Each row of the matrix corresponds to a word from Wikipedia, and the columns include the weights on the 200 vectors that corresponding to words. The matrix is saved in an *numpy* file that can only be read in python.

### 3. Merging with Trigrams from Job Ads

We create an Excel sheet ("*3gram*") with the corresponding CSS of all the trigrams in the job ads with all of the stereotypes (Appendix Figure C2).<sup>iv</sup>

Unnamed: 0	Word	productive	adaptable	warm	person	communicat	creative	ability	learn	experienced	dependable	attractive	memory	careful	hearing	technology	physically	at	personality
0	catalogs coordinate shipping	0.08990942	0.05313969	0.19760552	0.20634335	0.06176708	0.07348558	0.09472074	0.09854512	-0.02646285	0.02381901	0.06710053	0.16429107	0.23960878	0.01468514	0.19740577			
1	sales southwest ranches	0.10536893	0.04885759	0.03147481	0.05422942	-0.05438927	0.11572261	0.01313831	0.01774774	0.01193705	0.17657453	0.18822904	-0.1800341	0.10438366	0.09096844	0.11477691			
2	prospective clients success	0.16838222	0.12669513	0.10045204	0.32829291	0.23495799	0.26456618	0.17890905	0.22814527	0.25742427	-0.01805643	0.19022837	0.11348414	0.15883344	0.07011159	0.16120988			
3	appointment scheduling writing	0.03171875	-0.04546606	0.14023437	0.25051039	0.34466651	0.12218836	0.02775814	0.11804051	-0.06377374	0.08669132	0.20935424	0.27556646	0.07871884	-0.06923272	0.19688022			
4	simon plaza williston	-0.19512244	-0.11748193	-0.05212216	-0.19663957	0.04461493	-0.19098368	-0.2260315	-0.11025863	-0.02714494	0.02908025	-0.13633309	-0.03011486	-0.05571298	-0.15911523	0.02598622			
5	skills transportation needed	0.34593928	0.32764688	0.04814171	0.72389758	0.24205811	0.53654891	0.17179905	0.38766405	0.21872398	0.09463752	0.3626931	0.04849161	0.43690085	0.45936465	0.02726017			
6	would include microsoft	0.05507252	0.1282331	0.07825295	0.12048694	0.203337	0.18017448	0.15983821	0.0466962	0.07154812	-0.00294341	0.15222321	0.01580893	0.17686015	0.18597004	0.05789839			
7	fit position meet	0.01135917	0.21331298	0.14873748	0.19418238	0.00165645	0.29324073	0.02988006	0.15057229	0.12077843	0.04659567	0.03246775	0.17623547	0.03006035	0.12220551	0.09303959			
8	duties within assigned	0.15287887	0.06351158	0.08917952	0.19239633	-0.01984116	0.07527398	0.0183964	0.10259384	0.10679884	-0.03612029	0.11287006	-0.06852864	-0.03495055	0.25914133	0.11236193			
9	essential interested applying	0.40528563	0.37279681	0.08665893	0.38399914	0.28855851	0.38470435	0.16601919	0.21265782	0.43854371	0.01634677	0.46559557	0.03277518	0.25115901	0.35520077	0.09226359			
10	requirements conditions listed	0.09562268	0.22609578	0.00514583	0.19581731	-0.10460529	0.10000484	0.11359353	0.06344416	0.05290517	0.01450045	0.17482537	0.00380651	0.18993343	0.19036239	0.06781746			
11	team html look	-0.01515079	0.14093253	0.16503231	0.15821129	0.16901063	0.27929473	-0.02010506	0.10175705	0.15552515	0.11618007	0.02922575	0.03741275	0.14442667	0.08497276	0.16283206			
12	group busy medical	0.15953976	0.03862157	0.01099385	0.24548081	0.19480364	0.03549661	0.05429187	0.14650725	0.07932108	-0.04571284	0.17161372	0.06139317	0.21899821	0.12465939	-0.04119808			
13	data entry south	-0.01073593	0.03443835	0.06579509	0.16544062	-0.09061332	0.08198768	-0.13353833	0.00821663	0.04550007	0.13534744	-0.01189012	-0.09288855	0.27372527	0.00562626	-0.12731037			
14	environment multiple daily	0.20707737	0.3114644	0.29219025	0.40150142	0.28879637	0.31495175	0.06950984	0.10042349	0.17320518	0.17719334	0.20981269	0.02840904	0.35128736	0.13873798	0.19426003			
15	monday saturday right	-0.15820372	-0.11020786	0.09490361	-0.04754947	-0.11052027	0.05618757	-0.15060109	-0.07671998	-0.07000691	-0.01794079	-0.02857125	0.18608308	-0.17914858	0.08227927	-0.00218246			
16	seeking friendly patient	0.13009594	0.15739988	0.27222458	0.24603693	0.04549244	0.33473244	0.16545109	0.21027303	0.25369608	0.10025571	0.34812698	0.28539941	0.08464491	0.40854532	0.1889111			
17	daily evening events	-0.1608668	-0.0626367	0.17546733	0.09843986	0.02043072	0.02739242	-0.02296785	-0.06856505	-0.02023997	0.03499395	-0.0318891	0.15581027	-0.0547977	-0.03331282	0.10391671			
18	individuals provide security	0.1268528	0.29196921	-0.00695784	0.42222661	0.09578838	0.3035934	0.08727432	0.08990557	0.13403897	0.12934859	0.13164648	0.12627788	0.3121103	0.15283234	-0.01201946			

Appendix Figure C2.

Example of the Database of CSS's between Trigrams and Stereotypes

We then want to array these by job ad. To do this we match the trigrams in each ad with the values in sheet *3gram*, using the function *vlookup* in Excel. Once we have the CSS of the trigrams for each job ad (Appendix Figure C3), we can easily estimate the mean, the variance, the median, and the 90th and 95th percentiles.

<sup>iv</sup>The transition of the vectors used to compute the CS scores from words to trigrams is explained in the paper.

jobid	1	2	3	4	5	6	7	8	9	10	11	12	13	14
4826529596	-0.10024114	-0.09708729	-0.01844186	0.01448191	-0.03113504	0.01227904	-0.028702	0.00467881	-0.04675245	-0.04064595	-0.01487927	-0.07361431	-0.05576007	0.01312138
4833493607	-0.00617863	0.04613127	0.13083598	0.13241211	0.07699008	0.00921701	0.03383891	0.00771914	0.06467864	0.16539715	0.29615444	0.27665299	0.14728437	0.04790819
4833543304	0.21578708	0.17278934	0.03393763	-0.0699803	-0.21475951	-0.18009485	-0.09426907	-0.0203431	0.02943952	0.09283339	0.23393318	0.1111477	0.10658752	0.0616859
4835083297	0.03589931	0.07311814	0.09851844	0.07501265	0.10140617	0.04791774	0.05062216	0.15721832	0.24757797	0.19007869	0.13659912	0.10788237	0.13320187	0.10031367
4835197759	0.22598884	0.16380179	0.04292243	0.04810782	-0.00053314	0.06695009	-0.04439499	0.12912951	0.13929945	0.17689294	0.08671943	0.16388877	0.18513478	0.1695627
4841996184	0.03220937	0.04563214	0.17953841	0.27428922	0.31797147	0.32621011	0.26503557	0.05160114	0.03055433	0.0673307	0.2169556	0.18767482	0.22436081	0.2372003
4843542824	0.11018354	0.0939221	0.08907165	0.11684136	0.29699579	0.34429896	0.2828511	0.29532799	0.24810651	0.24442217	0.13075891	0.16802828	0.16100369	0.16328786
4843588254	0.08232513	-0.11876513	-0.2668415	-0.07415906	0.06851102	0.2120795	0.16378568	0.17194752	0.08303604	0.14349811	0.19362704	0.18216772	0.24480003	0.28867501
4843935707	0.01724329	-0.00311806	-0.00311805	0.0984218	0.09500656	0.07983962	0.13013262	0.12435132	0.01721883	-0.0234438	0.05834471	0.19795772	0.24787803	0.1176051
4844967381	0.06123802	0.08094908	0.04454608	0.01026424	-0.05688218	0.02195812	0.02264082	0.07050962	0.06763348	0.07946321	0.13165614	0.1298659	0.10242955	0.08656864
4845013578	-0.02677057	-0.00638406	-0.05178473	-0.02342796	-0.09080653	-0.03263257	0.03076433	0.08058264	0.04988318	0.00063334	0.13504805	0.16772181	0.19037443	0.05785627
4845385448	0.16387667	-0.07936481	-0.16221321	-0.21648592	-0.01154722	0.18576795	0.22141823	0.28458473	0.10574394	0.12344759	-0.00180374	0.15091908	0.24716361	0.33783776
4846715732	0.00396486	0.03477761	0.00195647	-0.14778571	-0.16713858	-0.15529579	0.1444006	0.2309681	0.3593688	0.23717786	0.25128782	0.11241271	0.06510139	-0.08271141
4847961364	-0.01452504	0.05434846	0.24676783	0.22714891	0.20228805	0.04652498	-0.08945298	-0.16283928	-0.1491413	0.02857964	0.14077154	0.28889999	0.16714802	0.1566406
4849587887	-0.12659535	-0.13430925	-0.18079892	-0.15327126	-0.12263446	-0.04791224	0.01008615	0.09241977	0.09241977	0.09241977	-0.0060811	-0.04221516	0.02635401	0.06603011
4849822816	0.12484151	0.10832646	0.12743169	0.05535002	-0.00200086	0.05923638	0.05923638	0.06952	-0.05810596	-0.05242701	-0.05864092	-0.05228057	0.0044537	-0.08131877
4853496101	0.02508972	0.15762445	0.08941705	0.10315267	-0.06815395	-0.00076564	-0.03403388	0.0524486	-0.05848509	-0.0565788	-0.07487626	-0.02046918	0.1047399	0.17303221
4853950964	0.31863141	0.10225229	0.05329071	0.02865437	-0.04688453	-0.14647791	-0.15236194	-0.02816193	-0.04039276	-0.08843983	-0.08673836	0.0506306	0.23228753	0.21930367
4854071360	0.02147976	-0.01066627	0.14758992	0.18806379	0.13778502	0.03352794	0.15261887	0.20403506	0.22562669	0.09278277	0.2243678	0.11057214	0.25189361	0.10366898
4854184121	0.13234863	0.11073428	0.1277038	-0.02811816	-0.10372508	-0.20075628	-0.10033982	0.01887384	0.12432531	0.09552225	0.15402076	0.17795062	0.23658381	0.23422804
4854184958	0.01950624	0.02677039	0.04314319	0.06319994	-0.00332793	-0.05309774	-0.137986	-0.10930039	-0.07223731	-0.06222705	-0.09666509	-0.1154443	-0.06697967	0.01219462

**Figure C3.**  
Example of Job Ads with CSS for each Trigram

#### 4. A Narrower Corpus

We used another algorithm to train our model on a subset of Wikipedia articles – articles related to age, stereotypes, and labor markets. We used a python module called *scrapy* to facilitate this task. We use the following steps:

1. First, we defined a set of Wikipedia articles related to ageism and labor markets, using the following articles: ageism, aging, discrimination, workforce, job, employment, stereotype, sexism, recruitment, skill\_(labor), and human\_capital.
2. Second, the algorithm scrapes all the paragraphs and sentences from this set of articles.
3. Third, algorithm creates “spiders” that click all the links in these articles, assembles those articles, and scrapes them.
4. Fourth, it repeats this on the new set of articles. (We chose to use two levels.)
5. Finally, once we have this new, smaller corpus, we apply the *gensim* algorithm to train a new model. We refer to this model as the “scrapy-spider” model.

Below, we discuss the code used in these steps.

First, we import all the modules that we are going to use. The most important module is *scrapy*, which scrapes articles from Wikipedia.

```
import re
import logging
from tqdm import tqdm
import scrapy
from scrapy.crawler import CrawlerProcess
```

We name the process.

```
class discrim5(scrapy.Spider):
    name = "discrim5"
```

To construct our smaller corpus, we want to restrict the clicks only to those in the Wikipedia domain. Thus, if the spiders find a link outside Wikipedia, they ignore it.

```
allowed_domains = ["en.wikipedia.org"]
```

We define the start urls (all related to ageism and labor markets).

```
start_urls = [
    "https://en.wikipedia.org/wiki/Ageism",
    "https://en.wikipedia.org/wiki/Ageing",
    "https://en.wikipedia.org/wiki/Discrimination",
    "https://en.wikipedia.org/wiki/Workforce",
    "https://en.wikipedia.org/wiki/Job",
    "https://en.wikipedia.org/wiki/Employment",
    "https://en.wikipedia.org/wiki/Stereotype",
    "https://en.wikipedia.org/wiki/Sexism",
    "https://en.wikipedia.org/wiki/Recruitment",
    "https://en.wikipedia.org/wiki/Skill_(labor)",
    "https://en.wikipedia.org/wiki/Human_capital",
]
```

We also define some settings. The most important are:

*feed format*: This specifies that the output will be in *csv* format.

*depth limit*: This defines how many times the spiders will click links on the website (article) – or the number of levels we use.

```
custom_settings = {
    'LOG_LEVEL': logging.WARNING,
    'FEED_FORMAT': 'csv',
    'FEED_URI': 'discrim5.csv',
    'DEPTH_LIMIT': 2,
}
```

In the next section, we define the codes for parsing all the articles and keeping only essential words (no symbols or HTML code).

*parse (parsing)*: Reads all the html code and extracts the title of the article and the main text.

*next\_page*: This reads all the links in the article.

*remove\_html\_tags*: This removes all the html code (like head, body, etc.).

```

def parse(self, response):
    for quote in response.css('.mw-parser-output > p'):
        yield {
            'quote': quote.extract(),
            'title': response.css("#firstHeading").extract(),
        }

    for next_page in response.css('div.mw-parser-output > p > a'):
        yield response.follow(next_page, self.parse)

def __remove_html_tags__(self, text):
    html_tags = re.compile('<.*?>')
    return re.sub(html_tags, "", text)

```

Here we just run all the processes and programs that we define above.

```

process = CrawlerProcess()
process.crawl(discrim5)
process.start()

```

Next, we import *pandas*, a spreadsheet module to save all the scraped text to spreadsheets.

```

import pandas as pd
dfquotes = pd.read_csv("discrim5.csv")
dfquotes.head()

```

We clean the articles further (*scrapy-spider* only eliminates tags) using *BeautifulSoup* (another python module). This step drops any symbol or HTML code that was not identified by *scrapy*. The code includes:

*num\_articles*: This command is used to identify the number of articles. It is used in the next loop when we clean the articles.

```

#Cleaning
from bs4 import BeautifulSoup
num_articles = dfquotes["quote"].size

```

We loop to clean the text using *BeautifulSoup*.<sup>v</sup> The loop cleans the body text (*clean\_p*) and the titles (*clean\_title*).

*Database*: This matches the cleaned titles with the cleaned body.

```

clean_p = []
clean_title = []
for i in tqdm(range(0, num_articles)):
    clean_p.append(BeautifulSoup(dfquotes["quote"].iloc[i]).get_text())
    clean_title.append(BeautifulSoup(dfquotes["title"].iloc[i]).get_text())

dfquotes["title_c"] = clean_title
dfquotes["body_c"] = clean_p

database = dfquotes[["title_c", "body_c"]]

```

<sup>v</sup>We have to do this for *scrapy-spider* because when we scrape directly from the online articles, html codes are included. They are already removed from the Wikipedia dump for the entire corpus.

Once we have the clean articles in a spreadsheet, we use *nunique()* to count the number of articles.

```
# Number of articles
database["title_c"].nunique()
```

The following code saves all the data.

```
database.to_excel("D:\RAWorkAgeDiscrim\SScoresMatch\discrim.xlsx")
database.to_csv("D:\RAWorkAgeDiscrim\SScoresMatch\discrim.csv")

#database=database.drop_duplicates(subset="title_c", keep="first")
#database.to_excel("D:\RAWorkAgeDiscrim\SScoresMatch\titles.xlsx")
```

The next part of the code (another file) trains the model. The code is very similar to the code used for training with the whole Wikipedia corpus. However, instead of reading all the Wikipedia dump, we read from the file we created in the last step (discrim.csv).

```
dfads = pd.read_csv("D:\RAWorkAgeDiscrim\SScoresMatch\discrim.csv")
dfads.head()
dfads["body_c"] = dfads["body_c"].replace(" ", "x", regex=True)
dfads["body_c"] = dfads["body_c"].replace(np.nan, "x", regex=True)

# Load the punkt tokenizer
tokenizer = nltk.data.load('tokenizers/punkt/english.pickle')

# Define a function to split a review into parsed sentences
def review_to_sentences( review, tokenizer, remove_stopwords=False ):
    # Function to split a review into parsed sentences. Returns a
    # list of sentences, where each sentence is a list of words
    #
    # 1. Use the NLTK tokenizer to split the paragraph into sentences
    raw_sentences = tokenizer.tokenize(review.strip())
    #
    # 2. Loop over each sentence
    sentences = []
    for raw_sentence in raw_sentences:
        # If a sentence is empty, skip it
        if len(raw_sentence) > 0:
            # Otherwise, call review_to_wordlist to get a list of words
            sentences.append( review_to_wordlist( raw_sentence, \
                remove_stopwords ))
    #
    # Return the list of sentences (each sentence is a list of words,
    # so this returns a list of lists
    return sentences

sentences = [] # Initialize an empty list of sentences

print ("Parsing sentences from training set")
for review in tqdm(dfads["body_c"]):
    sentences += review_to_sentences(review, tokenizer)

# Training the model

model = gensim.models.Word2Vec(sentences, size=200, window=10,
                               min_count=5, workers=multiprocessing.cpu_count())

model.init_sims(replace=True)

model.save("D:\RAWorkAgeDiscrim\Wikicorpus\wikiextractor\Spiders\model_spiders")
```



We also tried the Google News corpus. We did not scrape or train this model because it is available in gensim. We need to import it using the following code:

```
import gensim
from gensim.models import Word2Vec
import gensim.downloader as api
# This files download the google news model (run only if you don't have it)

#import gensim.downloader as api
#This part of the code retrieve the google news model.
model = api.load('word2vec-google-news-300')
```

We decided not to use Google News corpus because it is not fully clean. It includes stop words like “the,” “not,” “there,” etc. Also, symbols like “\_” and “\$.” Therefore we discarded the model because it is not comparable with the other ones.

## 5. Comparing Different Models

In this section we discuss comparing our two models. (This could be applied easily to a model trained on another corpus.) We compare them based on examining what words they rank as most related to a stereotype. In the example below, we ask the model to give us the 10 most similar words to the stereotype “attractive,” based on the CSS. (This could be modified for any number, of course.)

First, we use the full Wikipedia model:

```
model =
Word2Vec.load("D:\RAWorkAgeDiscrim\Wikicorpus\wikiextractor\word2vec_gensim")

#Size of the model:
print (model.wv.syn0.shape)

#Test
model.most_similar("attractive")
Output:
[('unattractive', 0.7458347678184509),
 ('appealing', 0.6965897679328918),
 ('desirable', 0.6877966523170471),
 ('unappealing', 0.6699395179748535),
 ('alluring', 0.6611822843551636),
 ('enticing', 0.6471317410469055),
 ('ideal', 0.6289623975753784),
 ('advantageous', 0.6268125772476196),
 ('agreeable', 0.6207250952720642),
 ('interesting', 0.6174749732017517)]
```

Then, we use the scrapy-spider model:

```

model =
Word2Vec.load("D:\RAWorkAgeDiscrim\Wikicorpus\wikiextractor\Spiders\model_spiders")

print(model.wv.syn0.shape)
model.most_similar("attractive")
Output:
[('desirable', 0.6811007857322693),
 ('advantageous', 0.6603072881698608),
 ('unattractive', 0.6249780654907227),
 ('appealing', 0.6150048971176147),
 ('hospitable', 0.5845235586166382),
 ('palatable', 0.5709764361381531),
 ('favorable', 0.5578073263168335),
 ('apt', 0.5564395189285278),
 ('adaptable', 0.5564374923706055),
 ('comfortable', 0.5478556156158447)]

```

Both models calculate similar words that are correlated with the stereotype “attractive.” For instance, both have the words “desirable” and “unattractive.” However, the scrapy-spider model finds the words “appealing” and “adaptable,” which appear to be more related to job markets and stereotypes.

## 6. Trying to Refine Treatment of “Experience” in Job-Ad Language

As we explain in the paper, we examined in greater detail job-ad language related to experience because this language could sometimes mean more of the characteristic, sometimes less, and sometimes be unrelated. Therefore, we created an algorithm that coded the level of experience required in a job ad to determine the impact of this potential blind spot in our model on our results.<sup>vi</sup>

## References

- AARP. 2000. *American Business and Older Employees*. AARP: Washington, DC.
- Adafre Sisa F., and de Rijke Maarten. 2006. “Finding Similar Sentences Across Multiple Languages in Wikipedia.” In *Proceedings of the EACL Workshop on New Text*. Trento, Italy.
- Ameri Mason, Rogers Sean Edmund, Schur Lisa, and Kruse Douglas. 2020. “No Room at the Inn? Disability Access in the New Sharing Economy.” *Academy of Management Discoveries* 6(2): 175–205.
- Anastopoulos L. Jason, George Borjas, Gavin Cook, and Lachanski Michael. 2019. “(Machine) Learning About Immigration’s Impact in Local Labor Markets with Classified Text.” Unpublished paper.
- Armstrong-Stassen Marjorie, and Schlosser Francine. 2008. “Benefits of a Supportive Development Climate for Older Workers.” *Journal of Managerial Psychology* 23(4): 419–437.
- Ash Elliott, Chen Daniel L., and Ornaghi Arianna. 2020. “Stereotypes in High-Stakes Decisions: Evidence from U.S. Circuit Courts.” Unpublished paper.
- Baert Stijn, Norga Jennifer, Thuy Yannick, and Van Hecke Marieke. 2016. “Getting Grey Hairs in the Labour Market. An Alternative Experiment on Age Discrimination.” *Journal of Economic Psychology* 57: 86–101.
- Banfi Stefano, and Villena-Roldán Benjamín. 2019. “Do High-Wage Jobs Attract More Applicants? Directed Search Evidence from the Online Labor Market.” *Journal of Labor Economics* 37 (3): 715–46.
- Barnett Julie. 1998. “Sensitive Questions and Response Effects: An Evaluation.” *Journal of Managerial Psychology* 13(1/2): 63–76.

<sup>vi</sup>This code is available upon request.

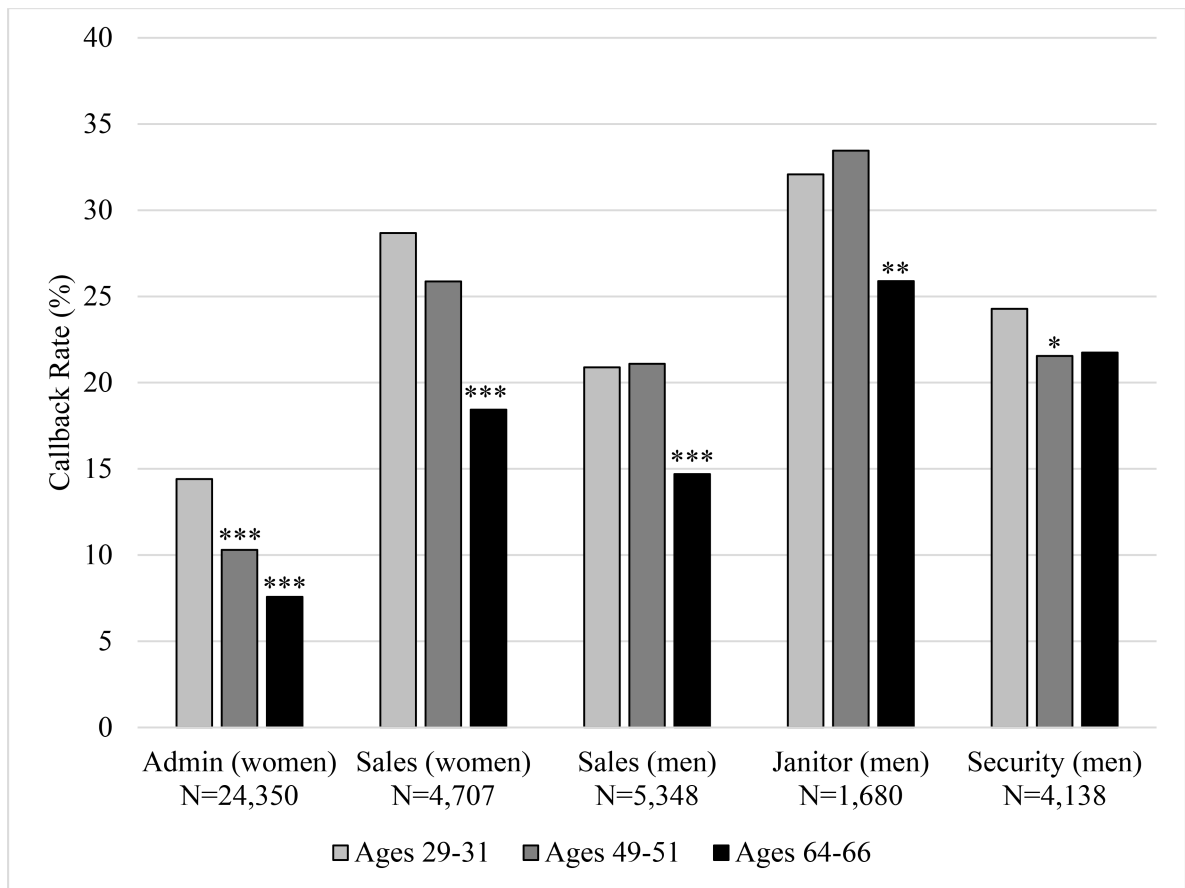


- Belot Michele, Kircher Philipp, and Paul Muller P 2018. “How Wage Announcements Affect Job Search – A Field Experiment.” IZA Discussion Paper No. 11814.
- Bendick Marc Jr., Brown Lauren E., and Wall Kennington. 1999. “No Foot in the Door: An Experimental Study of Employment Discrimination Against Older Workers.” *Journal of Aging & Social Policy* 10(4): 5–23. [PubMed: 10724770]
- Bendick Marc Jr., Jackson Charles W., and Romero J. Horacio. 1997. “Employment Discrimination Against Older Workers: An Experimental Study of Hiring Practices.” *Journal of Aging & Social Policy* 8(4): 25–46.
- Bertrand Marianne, and Duflo Esther. 2017. “Field Experiments on Discrimination.” In Banerjee AV and Duflo E (Eds.), *Handbook of Economic Field Experiments*, 309–93. North-Holland: Springer.
- Börner Katy et al. 2018. “Skill Discrepancies between Research, Education, and Jobs Reveal the Critical Need to Supply Soft Skills for the Data Economy.” *Proceedings of the National Academy of Sciences* 115(50): 12630–7.
- Börsch-Supan Axel, Hunkler Christian, and Weiss Matthias. 2021. “Big Data at Work: Age and Labor Productivity in the Service Sector.” *Journal of the Economics of Ageing* 10: 100319.
- Button Patrick. 2020. “Population Aging, Age Discrimination, and Age Discrimination Protections at the 50th Anniversary of the Age Discrimination in Employment Act.” In Czaja S, Sharit J, and James J (Eds.), *Current and Emerging Trends in Aging and Work*, 163–88. New York, NY: Springer.
- Carlsson Magnus, and Eriksson Stefan. 2019. “The Effect of Age and Gender on Labor Demand – Evidence from a Field Experiment.” *Labour Economics* 59: 173–83.
- Casey Katherine, Glennerster Rachel, and Miguel Edward. 2012. “Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan.” *Quarterly Journal of Economics* 127: 1755–812.
- Clark Stephen. 2015. “Vector Space Models of Lexical Meaning.” In Lapin S and Fox C (Eds.), *Handbook of Contemporary Semantics*, 493–522. Oxford: Blackwell.
- Crew James C. 1984. “Age Stereotypes as a Function of Race.” *Academy of Management Journal* 27(2): 431–35.
- Dedrick Esther J., and Dobbins Gregory H.. 1991. “The Influence of Subordinate Age on Managerial Actions: An Attributional Analysis.” *Journal of Organizational Behavior* 12(5): 367–77.
- Deming David, and Kahn Lisa B.. 2018. “Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals.” *Journal of Labor Economics* 36 (S1): S337–69.
- Farber Henry S, Silverman Dan, and von Wachter Till M.. 2017. “Factors Determining Callbacks to Job Applications by the Unemployed: An Audit Study.” *RSF: The Russell Sage Foundation Journal of the Social Sciences* 3(3): 168–201.
- Farber Henry S., Herbst Chris M., Silverman Dan, and von Wachter Till. 2019. “Whom Do Employers Want? The Role of Recent Employment and Unemployment Status and Age.” *Journal of Labor Economics* 37(2): 323–49.
- Finkelstein Lisa M., Higgins Kelly D., and Clancy Maggie. 2000. “Justifications for Ratings of Older and Young Job Applicants: An Exploratory Content Analysis.” *Experimental Aging Research* 26(3): 263–83. [PubMed: 10919070]
- Finkelstein Lisa M., Burke Michael J., and Raju Nanbury S.. 1995. “Age Discrimination in Simulated Employment Contexts: An Integrative Analysis.” *Journal of Applied Psychology* 80(6): 652–63.
- Finkelstein Lisa M., and Burke Michael J.. 1998. “Age Stereotyping at Work: The Role of Rater and Contextual Factors on Evaluations of Job Applicants.” *Journal of General Psychology* 125(4): 317–45. [PubMed: 9951407]
- Finkelstein Lisa M., Ryan Katherine M., and King Eden B.. 2013. “What Do the Young (Old) People Think of Me? Content and Accuracy of Age-Based Metastereotypes.” *European Journal of Work and Organizational Psychology* 22(6): 633–57.
- Fiske Susan T., Cuddy Amy J.C., Glick Peter, and Xu Jun. 2002. “A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition.” *Journal of Personality and Social Psychology* 82(6): 878–902. [PubMed: 12051578]
- Fix Michael, and Struyk Raymond J., Eds. 1993. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C.: The Urban Institute Press.

- Flory Jeffrey A., Leibbrandt Andreas, Rott Christina, and Stoddard Olga. 2021. "Increasing Workplace Diversity: Evidence from a Recruiting Experiment at a Fortune 500 Company." *Journal of Human Resources* 56(2): 73–92.
- Gaddis S. Michael. 2018. "An Introduction to Audit Studies in the Social Sciences." In Gaddis SM (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*. New York: Springer.
- Gordon Randall A., and Arvey Richard D.. 2004. "Age Bias in Laboratory and Field Settings: A Meta-Analytic Investigation." *Journal of Applied Social Psychology* 34(3): 468–92.
- Krumpal Ivar. 2013. "Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review." *Quality and Quantity* 47(4): 2025–47.
- Hanson Andrew, Hawley Zachary, and Taylor Aryn. 2011. "Subtle Discrimination in the Rental Housing Market: Evidence from e-mail Correspondence with Landlords." *Journal of Housing Economics* 20(4): 276–84.
- Hanson Andrew, Hawley Zachary, Martin Hal, and Liu Bo. 2016. "Discrimination in Mortgage Lending: Evidence from a Correspondence Experiment." *Journal of Urban Economics* 92: 48–65.
- Hellester Miguel D., Kuhn Peter, and Shen Kailing. 2020. "The Age Twist in Employers' Gender Requests." *Journal of Human Resources* 55(2): 482–69.
- Hendrick Jennifer J., Knox V. Jane, Gekoski William L., and Dyne Kate J.. 1988. "Perceived Cognitive Ability of Young and Old Targets." *Canadian Journal on Aging* 7(3): 192–203.
- Hummert Mary Lee, Garstka Teri A., Shaner Jaye L., and Strahm Sharon. 1994. "Stereotypes of the Elderly Held by Young, Middle-aged, and Elderly Adults." *Journal of Gerontology* 49(5): P240–9. [PubMed: 8056949]
- Hummert Mary Lee, Garstka Teri A., and Shaner Jaye L.. 1995. "Beliefs About Language Performance: Adults' Perceptions About Self and Elderly Targets." *Journal of Language and Social Psychology* 14(3): 235–59.
- Ibanez Marcela, and Reiner Gerhard. 2018. "Sorting through Affirmative Action: Three Field Experiments in Colombia." *Journal of Labor Economics* 36(2): 437–78.
- Jaeger David. A, Nunley John M., Seals R. Alan, and Wilbrandt Eric. 2020. "The Demand for Interns." NBER Working Paper No. 26729.
- Johnson Richard W., Kawachi Janette, and Lewis Eric K.. 2009. "Older Workers on the Move: Recareering in Later Life." Washington, DC: AARP Public Policy Institute.
- Jurafsky Daniel, and Martin James H.. 2017. "Vector Semantics." In *Speech and Language Processing, Third Edition (draft)*, <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>.
- Karpinska Kasia, Henkens Kène, and Schippers Joop. 2013. "Retention of Older Workers: Impact of Managers' Age Norms and Stereotypes." *European Sociological Review* 29(6): 1323–35.
- Kite Mary E., Deaux Kay, and Miele Margaret. 1991. "Stereotypes of Young and Old: Does Age Outweigh Gender?" *Psychology and Aging* 6(1): 19–27. [PubMed: 2029363]
- Kouzis-Loukas Dimitrios. 2016. *Learning Scrapy*. Packt Publishing Ltd.
- Krings Franciska, Sczesny Sabine, and Kluge Annette. 2011. "Stereotypical Inferences as Mediators of Age Discrimination: The Role of Competence and Warmth." *British Journal of Management* 22(2): 187–201.
- Kroon Anne C., Van Selm Martine, ter Hoeven Claartje L., and Vliegenthart Rens. 2016. "Reliable and Unproductive? Stereotypes of Older Employees in Corporate and News Media." *Ageing and Society* 38(1): 166–91.
- Kugelmass Heather. 2019 "Just the Type with Whom I Like to Work: Two Correspondence Field Experiments in an Online Mental Health Care Market." *Society and Mental Health* 9(3): 350–65.
- Kuhn Peter, and Shen Kailing. 2013. "Gender Discrimination in Job Ads: Evidence from China." *Quarterly Journal of Economics* 128(1): 287–336.
- Kuhn Peter, Shen Kailing, and Zhang Shuo. 2020. "Gender-Targeted Job Ads in the Recruitment Process: Facts from a Chinese Job Board: Evidence from China." *Journal of Development Economics* 147: 102531.
- Lahey Joanna. 2008. "Age, Women, and Hiring: An Experimental Study." *Journal of Human Resources* 43(1): 30–56.

- Lawrence Barbara S. 1988. "New Wrinkles in the Theory of Age: Demography, Norms, and Performance Rating." *Academy of Management Journal* 31(2): 309–37.
- Leibbrandt Andreas, and List John A.. 2018. "Do Equal Employment Opportunity Statements Backfire? Evidence from a Natural Field Experiment." NBER Working Paper No. 25035.
- Levin William C. 1988. "Age Stereotyping: College Student Evaluations." *Research on Aging* 10(1): 134–48.
- Lundberg Shelly J., and Startz Richard. 1983. "Private Discrimination and Social Intervention in Competitive Labor Markets." *American Economic Review* 73(3): 340–7.
- McCann Robert M., and Keaton Shaughan A.. 2013. "A Cross Cultural Investigation of Age Stereotypes and Communication Perceptions of Older and Younger Workers in the USA and Thailand." *Educational Gerontology* 39(5): 326–41.
- McLaughlin Joanne S. 2019. "Limited Legal Recourse for Older Women's Intersectional Discrimination Under the Age Discrimination in Employment Act." *Elder Law Journal* 26: 287–321.
- McGregor Judy, and Gray Lance. 2002. "Stereotypes and Older Workers: The New Zealand Experience." *Social Policy Journal of New Zealand* 18: 163–77.
- Madera Juan M., Hebl Michelle R., and Martin Randi C.. 2009. "Gender and Letters of Recommendation for Academia: Agent and Communal Differences." *Journal of Applied Psychology* 94(6): 1591–99. [PubMed: 19916666]
- Maestas Nicole. 2010. "Back to Work: Expectations and Realizations of Work after Retirement." *Journal of Human Resources* 45(3): 718–48. [PubMed: 24791018]
- Marinescu Ioana Elena, and Wolthoff Ronald. 2020. "Opening the Black Box of the Matching Function: The Power of Words." *Journal of Labor Economics* 38(2): 535–68.
- Maurer Todd J., Barbeite Frank G., Weiss Elizabeth M., and Lippstreu Micheal. 2008. "New Measures of Stereotypical Beliefs about Older Workers' Ability and Desire for Development: Exploration among Employees Age 40 and Over." *Journal of Managerial Psychology* 23(4): 395–418.
- Mikolov Tomas, Chen Kai, Corrado Greg, and Dean Jeffrey. 2013a. "Efficient Estimation of Word Representations in Vector Space." Unpublished paper, ICLR Workshop.
- Mikolov Tomas, Sutskever Ilya, Chen Kai, Corrado Gregory S., and Dean Jeffrey. 2013b. "Distributed Representations of Words and Phrases and their Compositionality." In *Advances in Neural Information Processing Systems* 26: 3111–19.
- Mikolov Tomas, Yih Wen-tau, and Zweig Geoffrey. 2013c. "Linguistic Regularities in Continuous Space Word Representations." In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*: 746–51.
- Modestino Alicia Sasser, Shoag Daniel, and Ballance Joshua. 2016. "Downskilling: Changes in Employer Skill Requirements over the Business Cycle." *Labour Economics* 41: 333–47.
- Neumark David. 2018. "Experimental Research on Labor Market Discrimination." *Journal of Economic Literature* 56(3): 799–866.
- Neumark David, Burn Ian, and Button Patrick. 2019. "Is It Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment." *Journal of Political Economy* 127(2): 922–70.
- Neumark David, Burn Ian, and Button Patrick. 2017. "Age Discrimination and Hiring of Older Workers." Federal Reserve Board of San Francisco Economic Letter #2017–06.
- Neumark David, Burn Ian, and Button Patrick. 2016. "Experimental Age Discrimination Evidence and the Heckman Critique." *American Economic Review Papers & Proceedings* 106(5): 303–08.
- Neumark David, Burn Ian, Button Patrick, and Chehras Nanneh. 2019. "Do State Laws Protecting Older Workers from Discrimination Reduce Age Discrimination in Hiring? Evidence from a Field Experiment." *Journal of Law and Economics* 62(2): 373–402. [PubMed: 32051647]
- Ningrum Panggih Kusuma, Pansombut Tatdow, and Ueranantasun Attachai. 2020. "Text Mining of Online Job Advertisements to Identify Direct Discrimination During Job Hunting Process: A Case Study in Indonesia." *Plos One* 15(6): e0233746. [PubMed: 32497044]

- Nunley John M., Pugh Adam, Romero Nicholas, and Seals R. Alan. 2015. "Racial Discrimination in the Labor Market for Recent College Graduates: Evidence from a Field Experiment." *B.E. Journal of Economic Analysis and Policy*, 15(3): 1093–125.
- Pennington Jeffrey, Socher Richard, and Manning Christopher D.. 2014. "GloVe: Global Vectors for Word Representation." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Pike Nathan. 2011. "Using False Discovery Rates for Multiple Comparisons in Ecology and Evolution." *Methods in Ecology and Evolution* 2(3): 278–82.
- Pitt-Catsoupes Marcie, Smyer Micheal A., Matz-Costa Christina, and Kane Katherine. 2007. "The National Study Report: Phase II of the National Study of Business Strategy and Workforce Development." The Center on Aging & Work at Boston College. [http://www.bc.edu/content/dam/files/research\\_sites/agingandwork/pdf/publications/RH04\\_NationalStudy.pdf](http://www.bc.edu/content/dam/files/research_sites/agingandwork/pdf/publications/RH04_NationalStudy.pdf).
- Posthuma Richard A., and Campion Michael A.. 2007. "Age Stereotypes in the Workplace: Common Stereotypes, Moderators, and Future Research Directions." *Journal of Management* 35(1): 158–88.
- Riach Peter A., and Rich Judith. 2006. "An Experimental Investigation of Age Discrimination in the French Labour Market." IZA Discussion Paper No. 2522.
- Ryan Ellen Bouchard. 1992. "Beliefs About Memory Changes Across the Adult Life Span." *Journal of Gerontology: Psychological Sciences* 47(1): 41–6.
- Ryan Ellen Bouchard, Kwong See Sheree, Meneer W. Bryan, and Trovato Diane. 1992. "Age-Based Perceptions of Language Performance Among Younger and Older Adults." *Communication Research* 19(4): 423–43.
- Ryan Ellen Bouchard, and See Sheree Kwong. 1993. "Age-Based Beliefs about Memory Changes for Self and Others across Adulthood." *Journal of Gerontology* 48(4): 199–201.
- Schmidt Daniel F., and Boland Susan M.. 1986. "Structure of Perceptions of Older Adults: Evidence for Multiple Stereotypes." *Psychology and Aging* 1(3): 255–60. [PubMed: 3267406]
- Singer MS 1986. "Age Stereotypes as a Function of Profession." *Journal of Social Psychology* 126(5): 691–92. [PubMed: 3807340]
- Stewart Mark A., and Ryan Ellen Bouchard. 1982. "Attitudes toward Younger and Older Adult Speakers: Effects of Varying Speech Rates." *Journal of Language and Social Psychology* 1(2): 91–109.
- Tian Ye and Zhang Jingbei. 2021. "Employment Discrimination Analysis of Library and Information Science Based on Entity Recognition." *Journal of Academic Librarianship* 47(2): 102325.
- Tilcsik András. 2011. "Pride and Prejudice: Employment Discrimination against Openly Gay Men in the United States." *American Journal of Sociology* 117(2): 586–626.
- Truxillo Donald M., McCune Elizabeth A., Bertolino Marilena, and Fraccaroli Franco. 2012. "Perceptions of Older Versus Younger Workers in Terms of Big Five Facets, Proactive Personality, Cognitive Ability, and Job Performance." *Journal of Applied Social Psychology* 42(11): 2607–26.
- van Borm Hannah, Burn Ian, and Baert Stijn. 2019. "What Does a Job Candidate's Age Signal to Employers?" IZA Discussion Paper No. 12849.
- van Dalen Harry, Henkens Kene, and Schippers JJ. 2009. "Dealing with Older Workers in Europe: A Comparative Survey of Employers' Attitudes and Actions." *Journal of European Social Policy* 19(1): 47–60.
- Warr Peter, and Pennington Janet. 1993. "Views about Age Discrimination and Older Workers." In Taylor P, et al. (Eds.), *Age and Employment: Policies, Attitudes, and Practice*, 75–106. London: Institute of Personnel Management.
- Wax Sidney L. 1948. "Discrimination by Summer Resorts in Ontario." *Information and Comment: Committee on Social and Economic Studies of the Canadian Jewish Congress* 7, June, 10–13.
- Weiss Elizabeth M, and Maurer Todd J.. 2004. "Age Discrimination in Personnel Decisions: A Reexamination." *Journal of Applied Social Psychology* 34(8): 1551–62.



**Figure 1: Comparisons of Job Applicant Callback Rates by Age**

Note: A callback is defined as an interview invitation or similar positive interest in an applicant. Figure is reproduced from Neumark, Burn, and Button (2017) using data from NBB. \* indicates that the estimate is statistically significant at the 10% level. \*\* indicates that the estimate is statistically significant at the 5% level. \*\*\* indicates that the estimate is statistically significant at the 1% level

**Part Time Insides Sales/Customer Service Rep**

[Company Name] is currently seeking Part -Time Customer Service Representative/Inside Sales Consultant to work in our beautiful Los Angeles showroom located on Westwood Blvd. This person will be responsible for maintaining, growing and acquiring new revenue from the existing Fashion customer base, as well as, generating new business. The ideal candidate will be operationally minded and have excellent sales abilities. Candidate must be reliable, energetic, customer-friendly, detail-oriented, and possess excellent time management skills.

Our Mission is to provide an unparalleled experience by fostering an environment that inspires teamwork and unrivaled passion for exceptional service.

[Company Name] is the only furniture rental company that solely services the California market. Based in San Diego, [Company Name] serves all of California, from San Diego, through The Bay Area. Please visit our website for more information [Company Website]

Essential duties and responsibilities include:

- Manage the day-to-day relationships with existing customer base and new prospects in partnership with field counterparts in assigned territory.
- Take orders and inquiries via incoming calls, walk in appointments, and email.
- Provide excellent customer service.
- Perform order and data entry with precision in a given time frame.
- Assist with residential and home staging furniture rentals.
- Assist outside sales reps with client issues & various projects.
- Provide support for internal and external customers.
- Assist in producing leads for new business
- Show a willingness to do whatever is needed to support the team

The position is heavily customer service oriented, and includes order processing. You will work with customers both over the phone and in person.

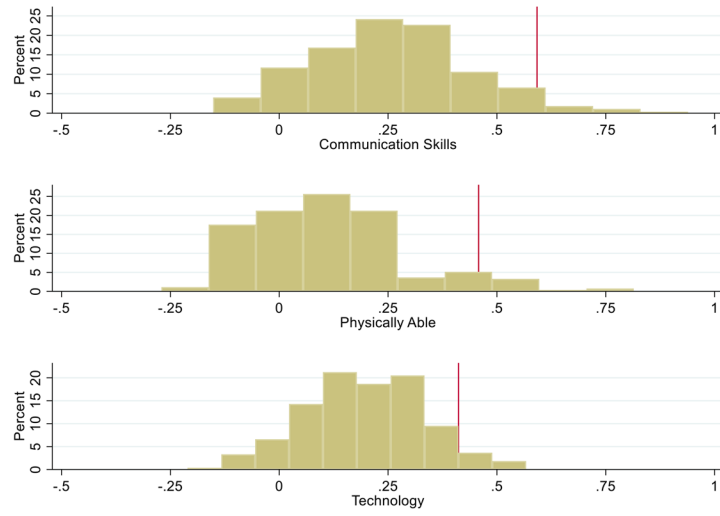
Requirements:

- Must possess excellent telephone etiquette, communication and organizational skills
- Candidate must be self motivated and be able to work independently.
- Applicants should have exceptional verbal and written skills, specifically the ability to coordinate multiple activities for numerous accounts.
- Candidate must maintain a willingness and ability to learn and understand our products and procedures.
- Effective time management
- Must be computer savvy; experience with MS Outlook and Office required
- 2-4 years of inside sales/customer service experience preferred.
- Property Management, Design/Merchandising, and Retail Sales experience a plus
- Must be available to work Saturdays

Our company is a stable, employee oriented organization, providing a fun, team oriented atmosphere.

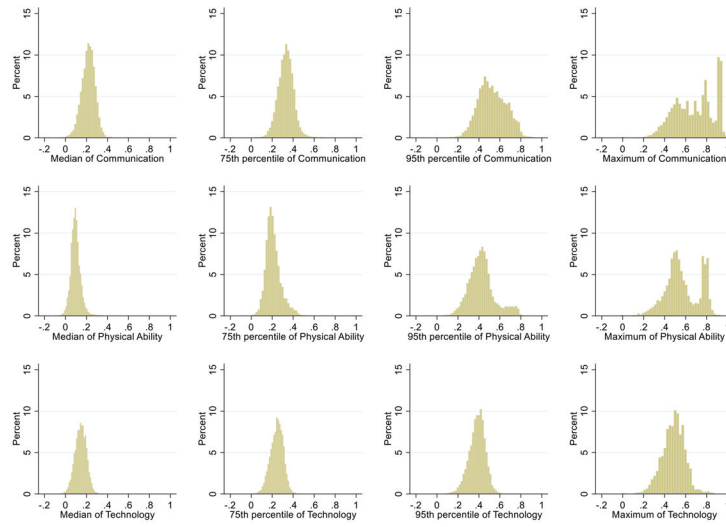
**Figure 2: Text of a Job Ad**

Note: Text drawn from actual job ad applied to in NBB, posted in June 2015. Company name and contact details have been removed.



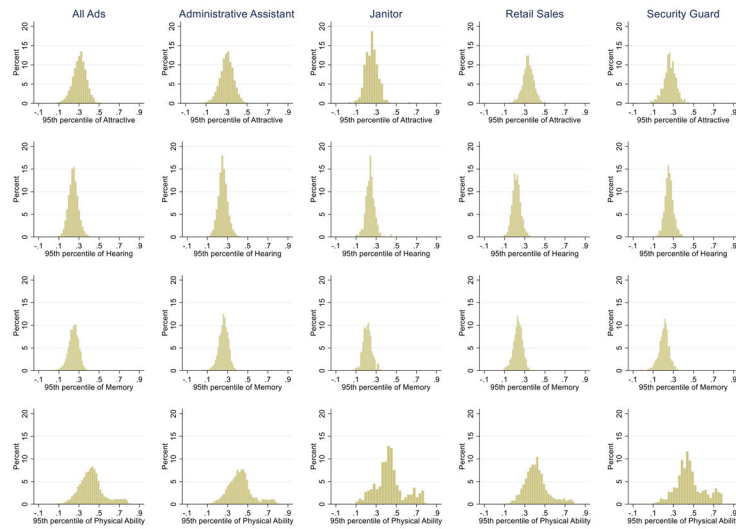
**Figure 3: Distributions of CS Scores within Example Ad (from Figure 2)**  
 Note: Based on the text from the job ad in Figure 2, these histograms plot the distributions of CS scores for all trigrams in the example job ad, for three stereotypes. The solid line in each graph indicates the 95<sup>th</sup> percentile.



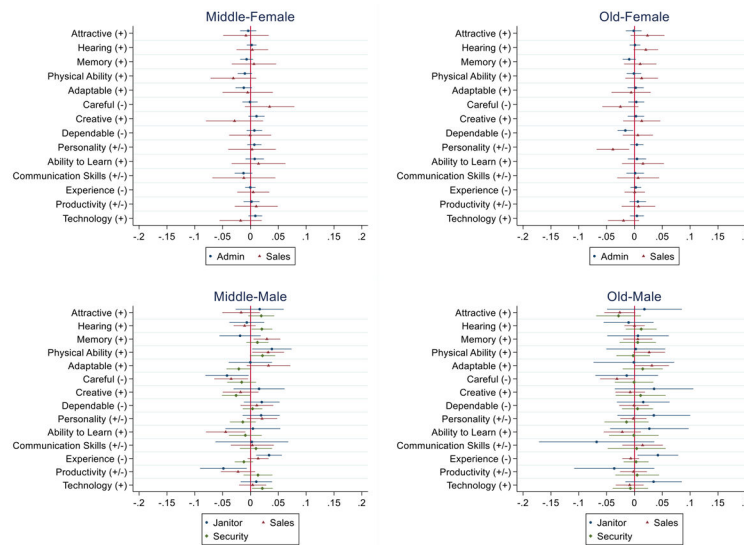


**Figure 4: Alternative Percentile Distributions of CS Scores of Job-Ad Trigrams with Select Stereotypes**

Note: Data come from the job ads collected in NBB. The distributions are for all the ads in our sample. Units on horizontal axes are the CS scores for the indicated percentiles/values of the distributions. Complete set of figures available upon request.

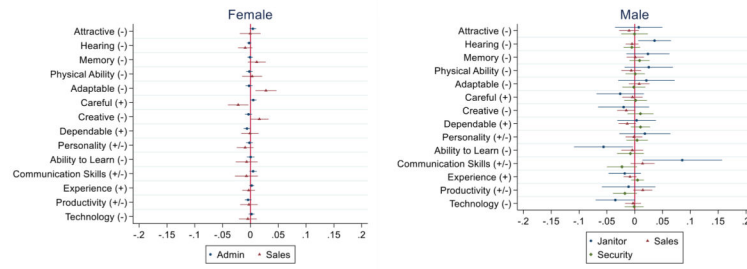


**Figure 5: Distributions of 95<sup>th</sup> Percentiles of CS Scores for Stereotypes Related to Health**  
 Note: Data come from the job ads collected in NBB. Each panel plots the distribution of CS scores at the 95<sup>th</sup> percentile for the job ads with each stereotype related to health. The first column contains the distribution of all the ads in our sample. The remaining columns disaggregate the job ads by occupation.



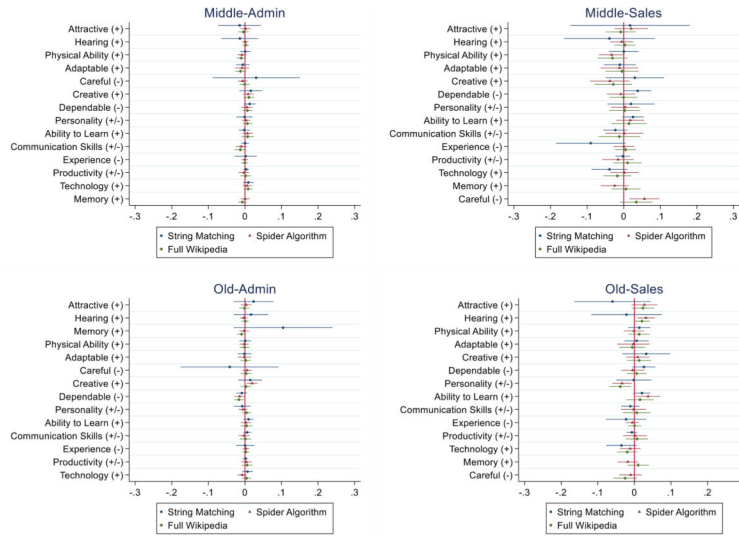
**Figure 6: Baseline Results by Gender, Age, and Occupation**

Note: Figure reports estimated marginal effects from equation (10) and 95% confidence intervals. Predicted signs of each coefficient according to the industrial psychology literature are presented in parentheses. Positive predictions indicate we expect to see more discrimination against older applicants, negative predictions indicate we expect to see less discrimination against older applicants. Standard errors are clustered at the job-ad level.



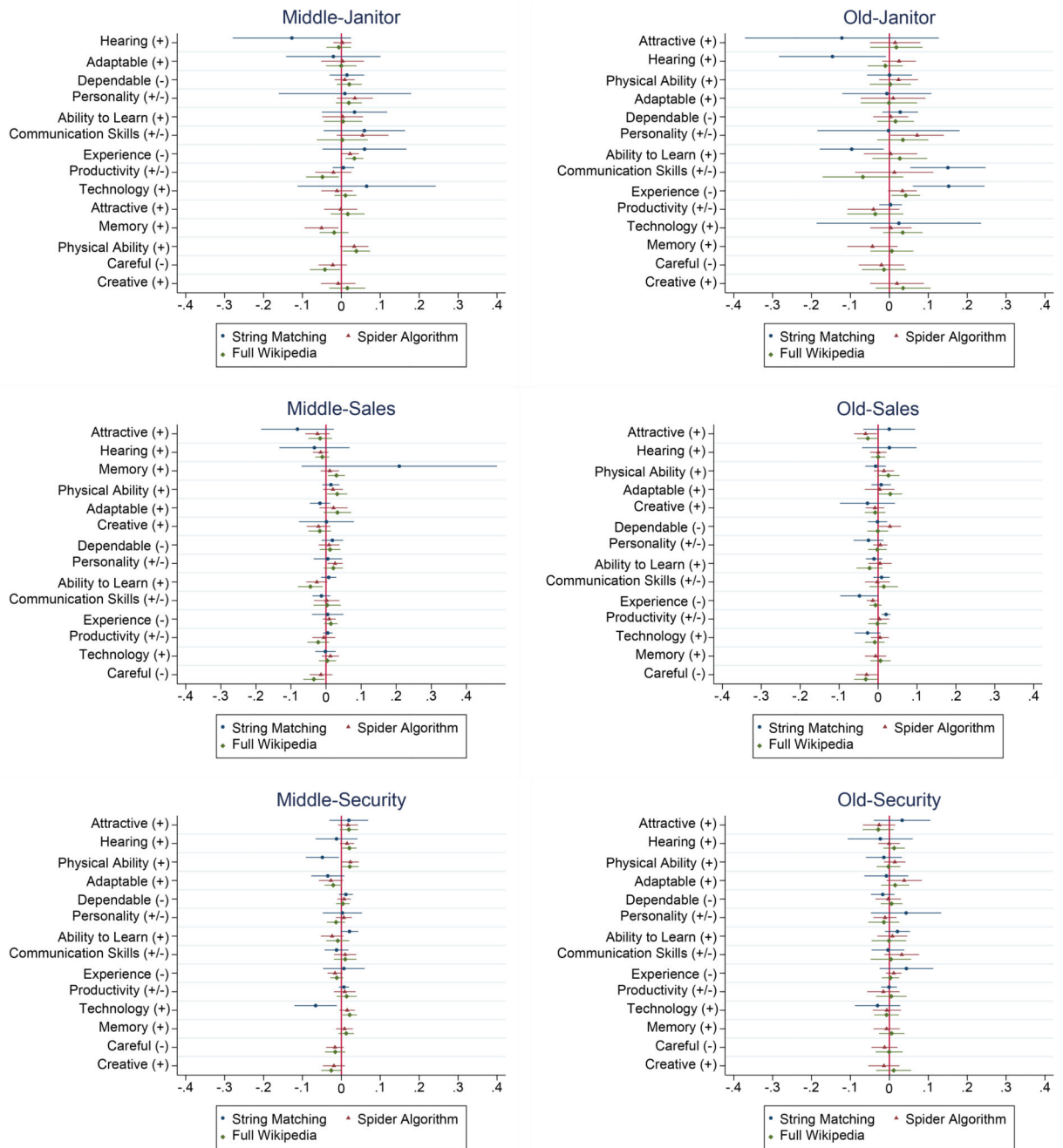
**Figure 7: Analysis of Discrimination Against Younger Applicants**

Note: See notes to Figure 6. The difference is that the outcome is now defined as discrimination against younger applicants. Predicted signs are reported in parentheses. Positive predictions indicate we expect to see more discrimination against younger applicants, negative predictions indicate we expect to see less discrimination against younger applicants.



**Figure 8A: Comparing Results with Alternative Corpora/Methods, Women**

Note: This figure reports the estimated marginal effects and 95% confidence intervals as we vary the corpora or method used to identify stereotyped language. The string matching reports the coefficient of the dummy variable indicating that the stereotype appeared in a job ad. Positive predictions indicate we expect to see more discrimination against older applicants, negative predictions indicate we expect to see less discrimination against older applicants. For age-occupation-gender cells where the word does not appear on any ad or is perfectly correlated with discrimination, no coefficient is reported. This occurs most often for memory, careful, and creative. The spider algorithm model and the complete Wikipedia models report the coefficients from Equation (10), with the complete Wikipedia model being identical to the baseline results. The string matching uses equation (11). Standard errors have been clustered at the job ad level.



**Figure 8B: Comparing Results with Alternative Corpora/Methods, Men**  
Note: See Figure 8A.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1:**

## Stereotypes about Older Workers' Health

Aggregate Stereotype	Phrasing	Source
Less Attractive	"wrinkled," "unattractive," "not neat"	Kite et al. (1991)
	"less attractive"	Levin (1988)
	"worse-looking when older"	Zepelin, Sills, and Heath (1987)
Hard of Hearing	"hard of hearing"	Kite et al (1991)
	"worse hearing," "think people speak too softly," "frustrated when not hearing," "think other people speak too fast," "often ask others to repeat"	Ryan et al. (1992)
	"worse hearing"	Hummert, Gartska, and Shaner (1995)
Worse Memory	"Worse memory"	Hendrick et al. (1988)
	"Worse memory"	Ryan (1992)
	"Worse memory"	Ryan and Kwong See (1993)
	"Worse memory"	Hummert, Gartska, and Shaner (1995)
Less Physically Able	"lower physical capacity"	Kroon et al. (2016) (p. 16)
	"[worse] physical capability and health"	van Dalen, Henkens, and Schippers (2009) (p. 21)
	"sedentary," "physically handicapped," "slow moving," "sick," "shaky hands," "fragile," "poor posture"	Schmidt and Boland (1986)
	"less qualified for a physically demanding job"	Finkelstein, Burke, and Raju (1995)
	"tired," "scared of becoming sick or incompetent"	Hummert et al. (1994)
	"[lower] activity," "[less] energy," "[worse] health," "[less] speed"	Levin (1988) (p. 142)
	"less physically active," "unhealthy," "moves slowly"	Kite et al. (1991)
	"worse psychomotor speed"	Hendrick et al. (1988)



**Table 2:**

## Stereotypes about Older Workers' Personality

Aggregate Stereotype	Phrasing	Source
Less Adaptable	"[less] flexible in doing different tasks," "[less likely to] try new approaches"	AARP (2000) (p. 6)
	"occupationally flexible"	Karpinska et al. (2013)
	"[more] flexibility"	Levin (1988) (p. 142)
	"[less likely to] adapt to change," "[less likely to] grasp new ideas"	Lyon and Pollard (1997) (p. 252)
	"older workers are less flexible than younger workers."	McCann and Keaton (2013)
	"resistant to change"	McGregor and Gray (2002)
	"find difficult to change," "old-fashioned"	Schmidt and Boland (1986)
	"adapt less well to change," "are less able to grasp new ideas"	Warr and Pennington (1993) (p. 89)
	"resistant to change"	Weiss and Maurer (2004)
Careful	"talks of past," "focuses away from future toward past"	Kite et al. (1991)
	"less flexible," "more old-fashioned"	Stewart and Ryan (1982)
	"think before they act"	Lyon and Pollard (1997) (p. 251)
	"older workers are more cautious than younger workers."	McCann and Keaton (2013)
Less Creative	"cautiousness," "self-discipline"	Truxillo et al. (2012) (p. 2623)
	"think before they act"	Warr and Pennington (1993) (p. 89)
	"better practical judgment," "better common sense"	Hendrick et al. (1988)
Less Creative	"[lower] creativity"	Levin (1988) (p. 142)
	"[lower] creativity"	van Dalen, Henkens, and Schippers (2009) (p. 21)
Dependable	"loyal"	AARP (2000) (p. 6)
	"[more] stability"	Crew (1984) (p.433)
	"more reliable," "committed to the organization"	van Dalen, Henkens, and Schippers (2009) (p. 21)
	"stable"	Finkelstein, Burke, and Raju (1995)
	"trustworthy," "reliability," "commitment"	Kroon et al. (2016) (p. 16)
	"are loyal to the organization"	Lyon and Pollard (1997) (p. 251)
	"reliability," "loyalty," "job commitment"	McGregor and Gray (2002)
	"loyal to the company," "are reliable"	Pitt-Catsoupes et al. (2007) (p. 8)
	"more loyal to the organization" "more reliable"	Warr and Pennington (1993) (p. 89)
Negative Personality	"more stable"	Singer (1986)
	"more trustworthy"	Stewart and Ryan (1982)
	"dejected," "poor," "hopeless," "unhappy," "lonely," "insecure," "complains a lot," "grouchy," "critical," "miserly"	Kite et al. (1991)
	"[less] pleasantness"	Levin (1988) (p. 143)
	"ill-tempered," "bitter," "demanding," "complaining," "annoying," "humorless," "selfish," "prejudiced," "suspicious of strangers," "easily upset," "miserly," "snobbish"	Schmidt and Boland (1986)
"[less] friendliness," "[less] cheerfulness"	Truxillo et al. (2012) (p. 2623)	

Aggregate Stereotype	Phrasing	Source
Warm Personality	“warm,” “good-natured,” “benevolent,” “amicable”	Krings, Sczesney, and Kluge (2010)
	“Warm personality”	Kroon et al. (2016) (p. 16)
	“more conscientious”	Warr and Pennington (1993) (p. 89)
	“warm”	Fiske et al. (2002)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3:**

## Stereotypes about Older Workers' Skills

Aggregate Stereotype	Phrasing	Source
Lower Ability to Learn	"will [not] participate in training programs"	AARP (2000) (p. 6)
	"learn new techniques" "personal development"	Armstrong-Stassen and Schlosser (2008)
	"[less] potential for development"	Crew (1984) (p.433)
	"lack willingness to be trained"	van Dalen, Henkens, and Schippers (2009) (p. 21)
	"training more appropriate for younger workers"	Dedrick and Dobbins (1991) (p. 373)
	"[less] ability and willingness to learn"	Kroon et al. (2016) (p. 16)
	"[less likely to] want to be trained"	Lyon and Pollard (1997) (p. 252)
	"Less interest in learning."	Maurer et al. (2008)
	"learn less quickly," "are less interested in being trained"	Warr and Pennington (1993) (p. 89)
	"less potential for development"	Finkelstein, Burke, and Raju (1995)
"lower potential for development"	Singer (1986)	
Better Communication Skills	"[better] interpersonal skills"	Crew (1984) (p.433)
	"better social skills"	van Dalen, Henkens, and Schippers (2009) (p. 21)
	"more interpersonally skilled"	Kroon et al. (2016) (p. 16)
	"sincere when talking," "tells more enjoyable stories"	Ryan et al. (1992)
Worse Communication Skills	"less interpersonally skilled"	Finkelstein and Burke (1998) (p. 331)
	"unable to communicate"	Schmidt and Boland (1986)
	"worse interpersonal skills"	Singer (1986)
	"talks slowly," "less sociable," "has few friends"	Kite, Deaux, and Meile (1991)
	"worse conversational skills," "hard to understand when noisy," "lose track of who said what," "lose track of topic," "lose track of what talked about," "hard to speak if pressed for time," "use fewer difficult words," "recognize meanings of fewer words"	Ryan et al. (1992)
	"less outgoing," "quieter voice," "more hoarse"	Stewart and Ryan (1982)
More Experienced	"solid experience"	AARP (2000) (p. 6)
	"[more] experience"	Finkelstein, Higgins, and Clancy (2000)
	"[more] experience"	Finkelstein, Ryan, and King (2013)
	"have useful experience"	Lyon and Pollard (1997) (p. 251)
	"having more experience which is useful in the job"	Warr and Pennington (1993) (p. 89)
More Productive	"strong work ethic"	Pitt-Catsoupes et al. (2007) (p. 8)
	"working harder"	Warr and Pennington (1993) (p. 89)
Less Productive	"[lower] performance capacity"	Crew (1984) (p.433)
	"attributed low performance more to the stable factor of lack of ability when the subordinate was old"	Dedrick and Dobbins (1991) (p. 368)
	"less economically beneficial"	Finkelstein and Burke (1998) (p. 331)
	"high performance rating is positively related with youth"	Lawrence (1988) (p. 328)
	"[less] competence"	Levin (1988) (p. 142)

Aggregate Stereotype	Phrasing	Source
	"younger workers are seen as having higher performance capacity"	Singer (1986) (p. 691)
Worse with Technology	"[less likely to] understand new technologies" "[less likely to] learn new technologies," "[less] comfortable with new technologies"	AARP (2000) (p. 6)
	"lack capacity to deal with new technologies"	van Dalen, Henkens, and Schippers (2009) (p. 21)
	"[less] technological competence" "[less] technological adaptability"	Kroon et al. (2016) (p. 16)
	"[less likely to] accept new technology"	Lyon and Pollard (1997) (p. 252)
	"Older workers adapt to new technology slower than younger workers." "Younger workers are less fearful of technology than older workers.	McCann and Keaton (2013)
	"problems with technology"	McGregor and Gray (2002)
	"less readily accept the introduction of new technology"	Warr and Pennington (1993) (p. 89)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4:**

Baseline Results by Gender, Age, and Occupation, Effects on Discrimination Against Older Applicants

		Female				Male					
		Middle-Admin	Middle-Sales	Old-Admin	Old-Sales	Middle-Janitor	Middle-Sales	Middle-Security	Old-Janitor	Old-Sales	Old-Security
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Baseline discrimination</i>		0.100	0.155	0.103	0.140	0.096	0.104	0.089	0.148	0.102	0.121
<i>Health stereotypes</i>		<i>Predicted sign</i>									
Attractive	Positive	-0.004 (0.007) [0.797]	-0.008 (0.021) [0.967]	-0.001 (0.007) [0.893]	0.023 (0.015) [0.455]	0.017 (0.022) [0.714]	-0.017 (0.017) [0.419]	0.019 (0.012) [0.243]	0.018 (0.034) [0.823]	-0.026 (0.014) [0.242]	-0.029 (0.020) [0.969]
Hearing	Positive	0.002 (0.004) [0.797]	0.003 (0.015) [0.967]	0.001 (0.005) [0.893]	0.021 (0.011) [0.455]	-0.007 (0.016) [0.864]	-0.010 (0.010) [0.419]	<b>0.020</b> * ( <b>0.010</b> ) [ <b>0.185</b> ]	-0.010 (0.023) [0.823]	0.000 (0.009) [0.983]	0.012 (0.014) [0.969]
Memory	Positive	-0.007 (0.006) [0.558]	0.006 (0.020) [0.967]	-0.009 (0.006) [0.847]	0.010 (0.015) [0.751]	-0.019 (0.019) [0.648]	<b>0.030</b> * ( <b>0.012</b> ) [ <b>0.102</b> ]	0.012 (0.010) [0.346]	0.006 (0.028) [0.959]	0.006 (0.013) [0.913]	0.006 (0.017) [0.969]
Physical Ability	Positive	-0.010 (0.007) [0.451]	-0.031 (0.021) [0.967]	-0.001 (0.007) [0.893]	0.013 (0.015) [0.751]	<b>0.038</b> * ( <b>0.018</b> ) [ <b>0.119</b> ]	<b>0.032</b> * ( <b>0.015</b> ) [ <b>0.102</b> ]	0.022 (0.012) [0.185]	0.002 (0.027) [0.978]	0.026 (0.015) [0.242]	-0.002 (0.015) [0.969]
<i>Personality stereotypes</i>		<i>Predicted sign</i>									
Adaptable	Positive	-0.012 (0.008) [0.451]	-0.005 (0.023) [0.967]	0.002 (0.007) [0.893]	-0.006 (0.018) [0.798]	-0.000 (0.020) [0.991]	0.032 (0.020) [0.273]	-0.021 (0.011) [0.185]	-0.001 (0.037) [0.978]	<b>0.031</b> * ( <b>0.016</b> ) [ <b>0.242</b> ]	0.015 (0.018) [0.969]
Careful	Negative	-0.001 (0.007) [0.958]	0.034 (0.023) [0.967]	0.003 (0.007) [0.893]	-0.025 (0.017) [0.455]	<b>-0.042</b> * ( <b>0.020</b> ) [ <b>0.119</b> ]	<b>-0.035</b> * ( <b>0.015</b> ) [ <b>0.102</b> ]	-0.016 (0.013) [0.346]	-0.014 (0.029) [0.823]	<b>-0.031</b> * ( <b>0.015</b> ) [0.242]	-0.001 (0.018) [0.969]
Creative	Positive	0.011 (0.007) [0.451]	-0.029 (0.026) [0.967]	0.003 (0.008) [0.893]	0.013 (0.017) [0.751]	0.015 (0.023) [0.714]	-0.018 (0.016) [0.419]	<b>-0.026</b> * ( <b>0.013</b> ) [ <b>0.185</b> ]	0.035 (0.036) [0.763]	-0.008 (0.014) [0.893]	0.011 (0.023) [0.969]
Dependable	Negative	0.007 (0.007) [0.558]	-0.001 (0.019) [0.979]	<b>-0.016</b> * ( <b>0.007</b> ) [ <b>0.322</b> ]	0.006 (0.014) [0.798]	0.020 (0.016) [0.602]	0.011 (0.015) [0.522]	0.004 (0.009) [0.677]	0.016 (0.024) [0.823]	-0.001 (0.014) [0.983]	0.006 (0.014) [0.969]
Personality	Positive/ Negative	0.007 (0.006)	0.003 (0.022)	0.004 (0.006)	<b>-0.039</b> ** ( <b>0.015</b> )	0.019 (0.017)	0.021 (0.014)	-0.014 (0.012)	0.035 (0.033)	-0.002 (0.012)	-0.014 (0.020)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

		Female				Male					
		Middle-Admin	Middle-Sales	Old-Admin	Old-Sales	Middle-Janitor	Middle-Sales	Middle-Security	Old-Janitor	Old-Sales	Old-Security
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		[0.558]	[0.968]	[0.893]	<b>[0.140]</b>	[0.604]	[0.273]	[0.346]	[0.763]	[0.983]	[0.969]
<i>Skills stereotypes</i>	<i>Predicted sign</i>										
Ability to Learn	Positive	0.008	0.014	0.005	0.015	0.004	<b>-0.045</b> *	-0.009	0.027	-0.022	-0.001
		(0.008)	(0.025)	(0.008)	(0.019)	(0.025)	<b>(0.018)</b>	(0.015)	(0.036)	(0.017)	(0.023)
		[0.576]	[0.967]	[0.893]	[0.751]	[0.991]	<b>[0.102]</b>	[0.579]	[0.823]	[0.560]	[0.969]
Communication Skills	Positive/ Negative	-0.013	-0.012	0.001	0.007	0.002	0.003	0.010	-0.068	0.015	0.004
		(0.008)	(0.029)	(0.008)	(0.019)	(0.033)	(0.020)	(0.015)	(0.053)	(0.019)	(0.027)
		[0.451]	[0.967]	[0.893]	[0.798]	[0.991]	[0.869]	[0.579]	[0.763]	[0.856]	[0.969]
Experienced	Negative	-0.000	0.005	0.002	0.000	<b>0.033</b> **	0.014	-0.012	<b>0.042</b> *	-0.007	0.003
		(0.005)	(0.015)	(0.005)	(0.009)	<b>(0.012)</b>	(0.010)	(0.008)	<b>(0.019)</b>	(0.008)	(0.011)
		[0.958]	[0.967]	[0.893]	[0.959]	<b>[0.070]</b>	[0.273]	[0.318]	<b>[0.336]</b>	[0.856]	[0.969]
Productive	Positive/ Negative	0.002	0.010	0.006	0.007	<b>-0.049</b> *	-0.022	0.014	-0.036	-0.002	0.005
		(0.007)	(0.020)	(0.008)	(0.015)	<b>(0.021)</b>	(0.016)	(0.013)	(0.037)	(0.012)	(0.020)
		[0.922]	[0.967]	[0.893]	[0.798]	<b>[0.119]</b>	[0.273]	[0.391]	[0.763]	[0.983]	[0.969]
Technology	Positive	0.009	-0.018	0.005	-0.020	0.011	0.004	<b>0.021</b> *	0.034	-0.009	-0.007
		(0.006)	(0.019)	(0.006)	(0.014)	(0.014)	(0.012)	<b>(0.009)</b>	(0.026)	(0.013)	(0.016)
		[0.451]	[0.967]	[0.893]	[0.473]	[0.714]	[0.810]	<b>[0.185]</b>	[0.763]	[0.884]	[0.969]
N		6,822	986	7,321	1,856	311	1,612	954	318	1,680	932
Adjusted R <sup>2</sup>		0.026	0.055	0.026	0.047	0.208	0.055	0.164	0.128	0.060	0.076

Note: Table reports estimated marginal effects from Equation (10). Predicted signs of each coefficient according to the industrial psychology literature (see Tables 1–3). Positive predictions indicate we expect to see more discrimination against older applicants, negative predictions indicate we expect to see less discrimination against older applicants. Standard errors are clustered at the job-ad level.

\* indicates that the estimate is statistically significant at the 5% level.

\*\* indicates that the estimate is statistically significant at the 1% level. In brackets, we report the p-values corrected for multiple hypothesis testing within a regression (i.e., a column) using the Simes procedure.