I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13506

Older Workers Need Not Apply? Ageist Language in Job Ads and Age Discrimination in Hiring

Ian Burn
Patrick Button
Luis Munguia Corella
David Neumark

JULY 2020

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13506

Older Workers Need Not Apply? Ageist Language in Job Ads and Age Discrimination in Hiring

Ian Burn
University of Liverpool

Patrick Button
Tulane University and IZA

Luis Munguia Corella
University of California–Irvine

David Neumark
University of California–Irvine and IZA

JULY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

Older Workers Need Not Apply? Ageist Language in Job Ads and Age Discrimination in Hiring*

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 methods from computational linguistics and machine learning to directly identify, in a field-experiment setting, ageist stereotypes that underlie age discrimination in hiring. The methods we develop provide a framework for applied researchers 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, our evidence points to age stereotypes about all three categories we consider – health, personality, and skill – predicting age discrimination, and for women, age stereotypes about personality. 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.

JEL Classification: J14, J7

Keywords: ageist stereotypes, age discrimination, job ads, machine learning

Corresponding author:
David Neumark
Department of Economics
University of California, Irvine
3151 Social Science Plaza
Irvine, CA 92697
USA

E-mail: dneumark@uci.edu

* We are grateful for helpful comments from seminar participants at the University of Bristol, University of Illinois, University of Liverpool, University of Maastricht, University of Tokyo, IZA, LISER, and Southern Methodist University. We thank Hayley Alexander and Emma Tran for excellent research assistance. We are especially grateful for the help of Nanneh Chehras, who assisted with the early stages of this paper. 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.

Introduction

We develop and implement methods to explore the role of stereotypes in hiring discrimination using the text of job ads, and apply this method to evidence on age discrimination. We make three 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 that appear in job ads are associated with an experimental measure of hiring discrimination against older workers – the first evidence we know of that can establish relationships between age-related stereotypes and actual employer behavior. Third, our analysis provides evidence on whether employers with less intent to hire older workers – as captured in our experimental results – use ageist language in their ads.

The most credible evidence on discrimination in hiring comes from field experiments – more specifically, resume-correspondence studies (Fix and Struyk, 1993; Gaddis, 2018; Neumark, 2018). These studies have been applied to discrimination based on race, ethnicity, gender, age, and other group membership (e.g., disability). Age discrimination, in particular, is of great policy interest in the United States and other countries because of rapidly aging populations. Low labor force participation rates of older individuals accentuate the contribution of aging populations to rising dependency ratios, straining the finances of many 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 important 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, or 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 to re-enter the workforce.

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 the 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 (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; Neumark et al., 2019).

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).[1] 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 the black box of why or how employers discriminate based on age. Specifically, we use the text data from the job ads in 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.

Little is known about which stereotypes employers act on when making actual hiring decisions. The industrial psychology literature mostly 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. Even in the unusual case in which researchers use a sample of managers with hiring experience, in their actual roles as managers they may not act on these stereotypes. 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

[1] 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.

evidence on actual age discrimination in hiring.[2] We use the text data in the thousands of job advertisements from our field experiment, and explore what job-ad language related to age stereotypes predicts age discrimination in hiring (as measured by the younger applicant being called back but not the older applicant). For example, one stereotype against 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.[3] For men, our evidence points to age stereotypes about all three categories we consider – health, personality, and skill – predicting age discrimination, and for women, age stereotypes about personality. 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.[4]

This paper makes three main contributions. First and most important, we produce evidence on which age-related stereotypes that appear in job ads are associated with hiring discrimination against older workers. Understanding which stereotypes underlie age discrimination can point to policy responses for reducing age discrimination. For example, job training, job coaching, or educational campaigns can focus on addressing the relevant negative stereotypes, or efforts could be focused on improving hiring practices, perhaps by increasing the information available to employers that reduces the attribution of stereotypes to older workers to whom they do not apply.

Second, our analysis provides evidence on whether employers with less intent to hire older workers –

[2] 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.

[3] Of course, our methods do not speak to the role of stereotypes held by employers that are not manifested in job ads.

[4] 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.

as captured in our experimental results – use ageist language in their ads. 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.[5] 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). Thus, our work can provide information to agencies that enforce age discrimination laws on job-ad language that may predict employer discrimination in hiring.

Finally, we are one of the first to develop detailed techniques, leveraging machine learning and textual analysis, to analyze the text data in job ads from field experiments on discrimination.[6] As audit and correspondence studies expand to study more markets, there are potentially more ways to leverage text data.[7] We build significantly upon earlier research leveraging text data from job ads by applying machine learning methods to the text analysis, rather than just searching for key phrases.[8] Our approach allows researchers to analyze text data when phrasing is complex, varied, and not always obvious (e.g., the numerous ways one

---

[5] See *Kleber v. Carefusion Corp.* (http://www.aarp.org/content/dam/aarp/aarp_foundation/litigation/pdf-beg-02-01-2016/kleber-amended-complaint.pdf, viewed November 8, 2017). Button (2019) discusses the ruling in this case.

[6] The only other study that we are aware of that used textual analysis, with machine learning, on job ads is Jaeger et al. (2020). They apply machine learning to the text from ads to classify internships into occupation categories. Most other correspondence studies do not analyze textual data, but there are some that do so on a limited basis. Hanson, Hawley, and Taylor (2011) is a notable example; they study subtle discrimination through "keywords" used by landlords responding to prospective tenants. Hanson et al. (2016) had research assistants subjectively (and blindly) code the helpfulness and other characteristics of mortgage loan originator responses to prospective borrowers. Tilcsik (2011) identifies four words in job ads related to masculine stereotypes (decisive, aggressive, assertive, and ambitious) and links those to hiring outcomes in a study of discrimination against gay men. Nunley (2015) uses textual analysis of job titles in ads to identify jobs that require extensive customer interaction.

[7] For example, Kugelmass (forthcoming) does a small correspondence study of discrimination in access to appointments with mental health professionals, who have on-line profiles, and Ameri et al. (2017) do a correspondence study of discrimination in access to AirBnB rentals. Both studies use platforms in which there is text data that could potentially by analyzed.

[8] 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 (2018) 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 (2019) 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.

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.

Background and Data from the Resume-Correspondence Study

To obtain estimates of age discrimination in hiring, NBB conducted a large and comprehensive resume-correspondence study of age discrimination. The study used realistic but fictitious resumes for young (aged 29-31), middle-aged (aged 49-51), and older (aged 64-66) job applicants. NBB sent three resumes to each job opening, one younger applicant and two resumes randomly selected from resumes for middle-aged or older applicants. Here, we summarize the key features of the study so that the job-advertisement data we exploit in the present paper can be understood.

The study entailed sending 40,223 applications (resumes) to 13,371 job positions in 12 cities (in 11 states). This is by far the largest 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, according to 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. 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.[9] NBB tracked callbacks – interview requests or similar

[9] 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. We also used different resume types to explore whether older workers who exhibit “bridging” behavior – the movement 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 the work experience history on the resume (NBB).

positive responses from employers – and compared them by age.

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 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 a number of more sophisticated analyses, but the basic conclusion remains the same.

Conceptual Framework

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 try to shape the applicant pool, to reduce the likelihood that age discrimination is detected. Using language that conveys positive stereotypes related to young workers might discourage older workers from applying for the job (as might language conveying negative stereotypes related to older workers – although that seems less likely and is, in fact, 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 from the employer) depends on how much lower the ratio of job offers to applicants is for older applicants than for younger applicants. 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 an anti-age discrimination claim.[10] Thus, we can test the hypothesis that discriminating employers use ageist language in job ads by relating the measure of age discrimination in the resume-correspondence study (differences in callback rates for older versus young applicants) to the age stereotypes in the job-ad language.[11] This hypothesis does not necessarily distinguish between taste and

[10] 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.

[11] There is, though, a potential bias against finding evidence that job ads with ageist stereotypes lead to lower callback rates for older applicants, if the ageist language lowers the share of older applicants enough so that the employer does not have to discriminate much against older applicants to get the desired younger workforce. While this may seem implausible, it would only imply that our results would be stronger without this bias.

statistical discrimination,[12] but rather just tests whether employers who do not want to hire older workers use stereotypes in job ads to facilitate their discrimination.

A second hypothesis is more closely related to statistical discrimination. Different jobs may have different requirements, which could be stated in job ads. But 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 the heavy lifting that a job requires. 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, religion, sex (including pregnancy), national origin, age (40 or older), disability or genetic information."[13] This text does 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.

A somewhat different and more complicated question is whether job requirements reflected in stereotyped language in job ads, to the extent they result in less hiring of older workers, are legal, which generally requires an employer to show that the use of these requirements is based on a reasonable factor other than age (RFOA), even if that factor is correlated with age. 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."[14] 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

[12] Taste discrimination is discrimination that occurs because employers, employees, or customers having animus against, or a distate 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.

[13] See http://www1.eeoc.gov//laws/practices/index.cfm?renderforprint=1 (viewed September 15, 2019).

[14] 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).

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 in 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

A key task in this paper is to classify job ads by the age stereotypes that appear in their language. To do this, we scrape the text of the job ads and use language processing software to identify language that conveys or relates to age stereotypes. We then use this classification of job ads to test whether employers who use language in their job ads that is related to negative stereotypes of older workers are less likely to hire older workers – as captured in the experimental results. (And we also study the possibility that employers who use positive age-related stereotypes in their job ads are more likely to hire older workers.)

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).

Our steps are as follows: First, we identify common age stereotypes from the research literature in industrial psychology and related fields. Second, we use computer science methods on 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). The similarity of these job-ad phrases to the age stereotypes is measured by a "semantic similarity score."

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. This allows us to quantify the usage of stereotyped language across ads, for each stereotype. The measure of age stereotyped language we use is the 95th percentile of the distribution for each stereotype for each ad. For example, consider two job ads. The first has many phrases pertaining to physical ability (a negative stereotype applied to older workers), and the second does not. Then the 95th percentile of the distribution of similarity scores to the physical ability stereotype will be much higher for the first ad.

Finally, we regress a dummy variable for observing discrimination age discrimination, in our experiment, on the 95th percentiles of each ad's similarity score distribution – for all of the stereotypes simultaneously. If we find a positive effect of the 95th percentile for a particular stereotype, the implication is that job-ad language related to that stereotype predicts hiring discrimination against older workers.[15] These steps are explained 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 literature 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, as there may be differences in age stereotypes across cohorts (Gordon and Arvey, 2004); hence, we avoided studies published before the 1980s and studies that focused on non-Western countries. We reviewed an extensive set of both 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.

If a study met these inclusion criteria, we compiled the list of the stereotypes that the study identified as applying to older workers. We also noted how the stereotype was described or phrased. Since studies often

[15] 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.

have similar stereotypes but phrase them differently, we grouped the stereotypes that were very similar into aggregate categories in a similar manner to the literature review and meta-analysis papers (e.g., Posthuma and Campion, 2007).[16] To focus 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 a list of 17 stereotypes of older workers, listed in Tables 1-3, corresponding to stereotypes related to health, personality, and skills. Of these 17 stereotypes, 11 are negative, including all the health-related stereotypes (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 two pairs are contradictory: worse/better communication skills, 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, with the goal of capturing all the ways that the stereotypes could reasonably appear in job abs. The complication is that we do not expect age stereotypes to be expressed in the job ads exactly as they are in the research literature.[17] Rather, there are many words and phrases that could be related to these 17 stereotypes, so that the true number of stereotyped words and phrases in the job ads could be large, and the strength of their association with age stereotypes varied.

We use methods from computational linguistics to determine the semantic similarity between phrases, as explained below. This process includes 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

[16] 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).

[17] For example, Appendix Figure A1 gives an example of a job ad from NBB. The job ad contains phrases that, on the surface, could be related to these stereotypes, including, for example, "experience," "social skills," and "social networking."

to train a model using textual data from English-language Wikipedia.[18] 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.[19] Second, we use this Wikipedia model to calculate the similarity between the 17 stereotypes and phrases consisting of these words 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, to take what would otherwise appear to be a jumble of words from the job ads (as well as the age stereotypes) and to sort them such that words that are used in similar contexts, as measured by Wikipedia, are placed closer together.

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.[20] The *word2vec* algorithm employs a continuous "bag of words" algorithm 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 created by *word2vec*, and 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 (the entirety of English-language Wikipedia), and then uses a series of linear projection functions (also known as hidden layers because they are not observed by the researcher) 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,

[18] 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.

[19] 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.

[20] 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/#.XOthPIhKiUl (viewed September 15, 2019) for an overview of the implementation of the *word2vec* algorithm and alternative applications.

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 which transform the data, and a constant (known as the bias) which shifts the projection to improve prediction.[21]

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 like 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.

Figure 2 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 5×1 vector projected onto a 5×1 output vector. There is a total of five words and only one node (the second dimension of the output vector) to define the context.

Usually, 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 the precision in the measurement of semantic similarity.[22] Hence, the actual neural network we construct takes as its input an 885,424×1 vector containing all the words and projects it into an

---

[21] For simplicity, imagine a neural network that exists in 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: $y = w_1 x_1 + w_2 x_2 + w_3 x_3 + b$. The bias correction is (b), and the weights (w) are the coefficients of the model. Note that the bias correction is equivalent to the intercept in a regression.

[22] 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).

output matrix that is 885,424×200.[23] We 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.[24]

Once the vector space has been estimated, 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 our set of stereotypes.[25] However, to this point, our explanation (and the illustrative example in Figure 2) 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, called "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 these stopping words. We retain the stereotypes as the number of words in which they are expressed in the first column of Tables 1-3 after removing the words indicating the direction of the stereotype, such as "more" or "less." Then, for each of the stereotypes, we calculate the cosine similarity (CS) score between the stereotype and every trigram used in the entire set of job ads. This requires some explanation.

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 layer 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

[23] 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 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.

[24] For more details about cosine similarity and semantic similarity and these kinds of models, see Clark (2015) and Jurafsky and Martin (2017).

[25] Note that there are two pairs of stereotypes that are mirror images of each other: worse/better communication skills and warm/negative personality. For these pairs, we just combine the stereotypes into a single phrase. Worse/better communication skills becomes communication skills and negative/warm personality becomes personality. Thus, we end up looking at cosine similarity scores with these 15 stereotypes. When we discuss the results, below, we explicitly consider the evidence on these ambiguous stereotypes.

score.

The first step in this process is to estimate the vector corresponding to the three words in the trigram (or the words in a stereotype). To do this, we add the weights element-by-element for each word.[26] 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:

$$able\ lift\ lbs = \begin{bmatrix}0.3\\0.2\end{bmatrix} + \begin{bmatrix}0.4\\0.1\end{bmatrix} + \begin{bmatrix}0.5\\0.2\end{bmatrix} = \begin{bmatrix}1.2\\0.5\end{bmatrix} \quad [1]$$

Using these vectors for trigrams and stereotypes, our next step is to estimate the CS score between them; in particular, we estimate the CS score between every trigram and every stereotype. The CS score measures the similarity between two vectors of an inner product space. The similarity between the vector of weights of a trigram and the vector of weights of a stereotype is given by the following equation:

$$\mathrm{CS}(\mathrm{trigram}, \mathrm{stereotype}) = \frac{\mathrm{dot\ product}(\mathrm{trigram}, \mathrm{stereotype})}{\|\mathrm{trigram}\|\ \|\mathrm{stereotype}\|} \quad [2]$$

where "trigram" and "stereotype" in the equation refer to the vectors of weights.[27]

The CS score varies between −1 and 1. A CS 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, we begin to see them used in 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, Figure 3 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

[26] 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 ≈ king − man + woman.

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

"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 4 (an actual ad from NBB). Figure 5 displays the distributions of the CS scores for each phrase (trigram) in the ad, with the communication skills, physical ability, and technology stereotypes. In this job ad, there is a wide range of CS scores, 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.

For our analysis, we need a summary measure of the distribution of CS scores in each job ad for each stereotype, to quantitatively compare the usage of stereotyped language (for each stereotype) across job ads. In determining which percentile of the distribution to use, we plotted the distributions of CS scores at the median, the 75th percentile, the 95th percentile, and the maximum. Figure 6 shows these for the same three stereotypes used in Figure 5. The histograms in Figure 6 provide a good rationale for using the 95th percentile. The mass of the distributions is much lower using the median (or the 75th 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 across the language in job ads that is by and large unrelated to the stereotypes we are studying to see if – when reflected in job-ad language – they predict discrimination. In addition, Figure 6 shows that if we used the median (or the 75th percentile), we would not pick up much range in the language across ads. This is also 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 highly related to the stereotypes and have a higher range of observed values, which suggests real differences in the usage of language. 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. We therefore chose to use the 95th percentile of the distributions of the CS scores to measure how stereotyped the language in a job ad is to capture meaningful differences in the usage of stereotyped language across job ads.

In Figures 7A-7C, we display the distributions – for each stereotype – of the 95th percentiles of the CS score computed from each job ad.[28] Each figure displays this information for all occupations combined, and then each occupation separately. Figure 7A presents the distributions for stereotypes related to health. 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 fairly common. For this stereotype, the distributions in all occupations feature a large mass of ads with trigrams in the range of 0.6 or higher.

Figure 7B turns to stereotypes related to personality. For these stereotypes, the distributions of the 95th percentiles of the CS scores are more normally (or at least symmetrically) distributed than those in Figure 7A. The job ads do not include phrases highly related to many of the stereotypes, with the median of the 95th 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).

In Figure 7C, we display the results for stereotypes related to 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 as well as most of those in Figures 7A and 7B, although the upper tails are not as extreme as for physical ability. And 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

[28] 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 associate with these ambiguous stereotypes is associated with higher or lower hiring of older applicants.

strongly related to experience and productivity are less common, with medians of the distributions of the 95th percentile scores in the 0.2 to 0.3 ranges, and sometimes below 0.2.

Testing which Stereotypes Predict Callback Differences by Age

Our goal is to estimate the relationship between age stereotyped language in job ads and the likelihood that older or younger applicants received callbacks. We hypothesize that job ads with negative age stereotypes will have relatively lower callback rates for older workers (while job ads with positive age stereotypes will have relatively higher callback rates for older workers). Thus, our next step is to use our job-ad-level measures of the relationship between language in the ad and age stereotypes to identify which stereotyped job-ad language predicts differential treatment of older applicants.

Our data set includes all responses to the triplet of job applications sent in response to each job ad that could be matched to an employer and their job advertisement. It is possible to match 34,260 job applications to 11,420 job advertisements, corresponding to 22,840 observations for older and middle-aged applicants.[29] Our outcome (D_{ij}) is a dichotomous 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 – our experimental measure of age discrimination. That is, 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; in these cases, we code D_{ij} as zero.[30] 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

[29] There are 4,266 applications that cannot be matched to a saved job ad. This can be due to a number of reasons; the most common 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.

[30] 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.)

cases.[31]

We estimate probit models for our experimental measure of discrimination, for which the key independent variables of interest are the 95th 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* as in NBB.[32] Thus, our model is

$$Pr[D_{ij} = 1] = \alpha + \sum_s \beta_s P_{js}^{95} + X_{ij}\delta + \varepsilon_{ij} \qquad [3]$$

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. β_s then represents the effect of a one standard deviation increase in the 95th percentile of the CS score in a job ad for stereotype *s*.[33] The estimate of β_s then reflects both the effect of the stereotype on discrimination and how related the job-ad language is to the stereotype. A small β_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 what is of interest, whereas a comparison between the estimated effects of the same absolute change in the 95th 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. The null hypothesis that stereotyped language related to stereotype *s* does not predict discrimination against older workers implies that $\beta_s = 0$.

We estimate our models at the gender-age-occupation level (e.g., women aged 49 to 51 applying to administrative assistant positions), to allow the effects of stereotypes to vary across different employers and applicants. Because each job ad received two pairs of applicants – with one younger applicant and one either

[31] 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.

[32] Resume features include: city, order sent, skill level, unemployment status, template, and email domain.

[33] In Appendix Table A2, we provide the text of some trigrams to give a sense of how they differ one standard deviation higher in the distribution of CS scores at the 95th percentile. We report the mean 95th percentile and the five trigrams closest to one standard deviation higher than the mean 95th percentile. This corresponds to the interpretation of β_s, although identification of β_s comes from variation across the entire distribution, not just the mean to one standard deviation higher.

middle-aged or older applicant in each pair – sometimes there are two correlated observations (when the two non-young applicants were both middle-aged or both older). Thus, we cluster the standard errors (ε_{ij}) at the job-ad level.

Results

We begin by examining the effect of stereotypes on outcomes of older workers generally, before moving to the more informative results by occupation. Table 4 presents these results for estimates of Equation [3] by age, gender, and occupation.[34] Positive coefficients in Table 4 represent the marginal effect of a one standard deviation increase in the CS score at the 95th percentile of the job ads' distributions for that stereotype. For example, middle-aged men who applied for a job as janitors (column (5)) experienced discrimination 9.6% of the time ("baseline discrimination" in the first row of the table, measured as the percentage of job ads where there younger applicant received a callback but the older applicant did not). If the job ad featured language where the 95th percentile was one standard deviation more related to physical ability than the average ad, measured discrimination was 3.8 percentage points (or 40%) higher.

The top row of Table 4 highlights the distribution of discrimination against older applicants. On average, 11% of the time an older worker was discriminated against when applying for a job, but the rate of discrimination varies by occupation. We observed the highest levels of discrimination among retail sales positions, with older women experiencing significantly higher levels of discrimination in retail sales than older men (14.0% vs. 10.2%). For women, measured discrimination was lower for administrative assistant than for sales jobs, for both age groups, while for men this varied by occupation.

The results in Table 4 provide evidence that stereotyped language on 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 correlations, 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

[34] Estimates where we pool all occupations and genders together are similar, and are available upon request.

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. 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 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 largely 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). Note also that 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). This is what we expected to find 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 95th 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 95th percentile is associated with a 3.2 percentage point increase in discrimination against middle-aged men.[35]

For retail sales positions, we find that employers with job ads that used language more highly related

[35] In Appendix Table A2, we show that this one standard deviation increase in the CS score for physical ability is akin to changing the 95th percentile from "assistant position available" to a phrase similar in relatedness to "work preferred necessary," "fast paced fun," or "required flexibility required."

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 95th 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.[36]

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 is what we expected to find based on the industrial psychology literature indicating that employers view older workers as having worse hearing than younger workers (see Table 1). Job ads with phrases for which the 95th 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 95th 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.

[36] 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.

An example of this problem may be the stereotype "memory." 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. The correlation we are observing then is due to the fact that our algorithm is unable to finely parse words related to memory to only be human related.

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 is 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 95th 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 is in line with the industrial psychology literature, which indicates that employers view older workers as more careful than younger workers (see Table 2). On janitor job ads, when phrases for which the 95th percentile of the distribution of CS scores with careful were one standard deviation higher, measured discrimination was 4.2 percentage points lower. On sales associate job ads, a one standard deviation increase in this 95th 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. This literature finds that older workers were viewed as less creative than younger by employers (see Table 2), whereas 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 95th 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 literation, in which adaptability is viewed as a trait that older workers did not possess (see Table 2). Job ads with phrases for which the 95th 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 a number of age-occupation cells. There is a wider range of stereotyped language that predicts discrimination for middle-aged men, and we observe significant heterogeneity across occupations, while 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 empirically the relative importance of the positive or negative association 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 experience are less likely to discriminate against middle-aged men. Job ads with phrases for which the 95th 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 large 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 95th 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. The industrial psychology literature points to a positive correlation between age and experience on the part of employers, perhaps a bit tautologically. In contrast, we find that employers with job ads with language more highly related to experience are more likely to discriminate against middle-aged and older men. Job ads with phrases for which the 95th 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 the predictions of the industrial psychology literature are not borne out by our results – for the creative stereotype for middle-aged men applying to security guard positions, and for the experienced stereotype for middle-aged and older men applying for janitor positions; in these cases we find negative stereotypes associated with less hiring discrimination. In contrast, in 13 cases the evidence is consistent with the industrial psychology literature.

Supplemental Analyses

In this section, we discuss alternative ways to conduct our analyses, to test how robust our results are to the decisions we made when designing our study and analysis. The three choices we will explore are how to define discrimination, which percentile to use, and how large of a phrase to use.

Definition of Discrimination

Our main focus has been to understand discrimination against older applicants, and hence we defined discrimination 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 where we redefined the outcome variable to be one if the younger applicant was not called back but the older applicant was called back. 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 the older applicants entailing callbacks to *both* older applicants but not the younger applicant. In this analysis, we aggregate the comparisons of older workers to younger workers and ignore the variation that comes from differences in whether employers prefer middle-aged and older workers to the younger workers.[37]

Table 5 presents the results defining the outcome variable as discrimination against *younger* applicants. 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 clustered in the personality stereotypes (similar to what was observed in Table 4). The evidence points to more discrimination against younger workers when job ads use stereotyped language associated with adaptable, and less discrimination against younger workers when the job-ad language reflects the careful and dependable stereotypes. All of these associations point in the opposite direction to what we would have predicted based on the industrial psychology literature, where older workers are stereotyped as more careful and dependable, and less adaptable.

For men, we only find significant associations for janitor positions. These results are less at odds

[37] Results looking at younger versus middle-aged and younger versus older were qualitatively similar, and results are available upon request.

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.[38] For stereotyped language related to skills, we find significant effects that are in line 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 – 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 existing statistically significant results from Table 4, suggesting that this alternative definition of discrimination captures some discriminatory behavior that our original definition did not capture. In only one case (for dependability for older women applying to administrative jobs) do the results contradict, making it unclear how the dependable stereotype is related to age discrimination.

Construction of Phrases

Another choice we made that could influence our results was how many words to use in job-ad phrases. As our baseline, we chose to use trigrams (three-word phrases). Our *word2vec* model builds up our phrases using the single word CS scores. But using the same method, it is possible to construct similarity scores for any phrase length. Consistent with our strategy of specifying the relationships between job-ad language and age stereotypes ex ante, we chose to use trigrams before doing any analysis of the relationship between the selected phrases, stereotypes, and measured discrimination. This avoided the risk of cherry picking – choosing the number of words to use in phrases (three, or something else) to obtain a particular set of results.

Nonetheless, after the fact, we also created versions of the histograms in Figures 7A-7C using two- and four-word phrases to examine the sensitivity of the results to changes in the number of words in a phrase.

[38] 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).

Comparing the distributions of the 95th percentile as we increased the phrase size from two words to four words, there was an increase in the mean similarity score at the 95th percentile, and higher variance. The results in Tables 6A and 6B provide evidence that the results we obtained using three-word phrases is robust in many cases to using two- or four-word phrases. The signs of our estimates point in the same direction as we increase the size of our phrases. For women, varying the size of the phrase does not result in more significant correlations. We still observe a fairly small number of significant results, suggesting that the lack of strong correlations between stereotyped language and discrimination against older women is not an artifact of our phrase length. There is some heterogeneity in the stereotypes that significantly predict discrimination as we vary the phrase length, even though the estimates are often similar and sign and magnitude.

For men, we find that the results from Table 4 are remarkably robust to changing the length of the phrase. 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). The results are perhaps more robust when comparing trigrams to quadgrams and less robust when comparing bigrams to trigrams, with a higher incidence of results that are large and significant only for bigrams (e.g., in Table 6B, for older male applicants to janitor or security positions (for dependable)). This may be because two-word phrases less reliably reflect the underlying stereotype – consistent with the mean CS scores being lower for bigrams.[39]

Discussion and Conclusions

We develop new machine-learning techniques for analyzing complicated textual data, which we

[39] In addition to the analyses described in this section of the paper, in Appendix Tables A1A and A1B, we replicate our baseline results using as alternatives the median and the maximum rather than the 95th percentile. The results for women in Appendix Table A1A, and for men in Appendix Table A1B, 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 95th percentile, but 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 6, this is not surprising, since the shapes of the distributions are sometimes very different for the maximum than for the 95th percentile.

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 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.

A key contribution of our techniques is that they can be adapted to other contexts. In audit or correspondence studies of labor market discrimination, regardless of the group studied, textual data is or can be collected. It may also be possible to apply our methods to studies of 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.[40]

In our context of age discrimination, 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, our evidence points to age stereotypes about all three categories we consider – health, personality, and skill – predicting age discrimination, and for women, age stereotypes about personality. 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 larger role in generating observed hiring discrimination against older men than against older women, even though our correspondence study found stronger evidence, overall, of hiring discrimination against older women. A similar puzzle emerges in that the evidence points to a larger effect of stereotypes in job ads

[40] For a discussion of research on letters of recommendation, see Madera et al. (2009).

for middle-aged than for older men, despite the experimental evidence providing somewhat stronger evidence of stronger age discrimination in hiring against older men. Why might stereotyped language matter more older workers when measured discrimination is lower? We have a number of hypotheses that could explain these findings.

Recall that we had two hypotheses about why age stereotypes in job ads might be associated with lower hiring of older applicants. One hypothesis – that could be consistent with either taste discrimination or statistical discrimination – is that employers use stereotyped language to discourage applications from certain groups, and the same employers who use this language also hire fewer older applicants. The second hypothesis – more closely related to statistical discrimination – is that the ads list real requirements, but employers hold stereotypes about older workers' ability to fulfill these requirements.

These hypotheses suggest one simple explanation for our findings – namely, that taste discrimination is more relevant for older vs. middle-aged men, and for older women vs. older men, consistent with evidence NBB used to argue that the discrimination they measure is likely to be mostly taste-based discrimination (p. 933). If taste discrimination is weaker for older men in comparison to older women, and for middle-aged men in comparison to older men, then stereotypes may play more of a role, explaining why we find stronger links between stereotypes and discrimination for older men than for older women, and for middle-aged men than for older men.

A second potential explanation is more related to the motivation to use age stereotypes to discourage workers from applying, and how this may interact with the effectiveness of age discrimination laws. Consider a subgroup of older workers for which age discrimination laws are less effective (we will explain why effectiveness might vary below). 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, because it is not as important to shape the applicant pool to avoid detection of discriminatory behavior, weakening the linkage between ageist stereotypes in job ads and hiring for this subgroup. Age discrimination laws may be less effective for older women than for older men because of the difficulty of bringing to court discrimination claims based on being both female and older and claims that intersect age and sex (McLaughlin, 2019). And age discrimination laws may 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 (see, e.g., the evidence in NBB) implies that the same hiring rate difference relative to young workers is less likely to be statistically significant for the comparison with older applicants than for the comparison with middle-aged applicants.[41] Hence the linkages between ageist stereotypes in job ads and hiring may be stronger for older men than for older women, and for middle-aged men than for older men, precisely because for the former groups discriminatory employers have a stronger incentive to shape the applicant pool to avoid detection of age discrimination.

Finally, 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 on age stereotypes largely 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, it simply may be that stereotypes relevant in more-traditionally female jobs (like administrative assistants) may be those which are more difficult to express in job ads, which may lead to a weaker relationships between the underlying stereotypes and phrases from the job ads.

Our evidence that measured age discrimination is related to the use of ageist stereotypes in job adds suggests that the policy responses to the age discrimination in hiring documented in NBB and other resume-correspondence studies of age discrimination (such as Farber et al., 2017 and 2019) need to be more nuanced. For example, if older workers are aware of the relationship between ageist language in job ads and hiring discrimination, they may alter their job search behavior, complicating efforts to detect age discrimination by comparing job application and hiring rates. On the other hand, the evidence provided in this paper has potentially constructive implications for discrimination policy. Our results can provide

[41] 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.

guidance to the Equal Employment Opportunity Commission and state agencies that enforce age discrimination laws. If employers use ageist language to discourage older workers from applying to jobs, then applicant pools may be shaped to make age discrimination in hiring harder to detect. Barring such language may reduce employer efforts to shape the applicant pool, and testing for age stereotypes in job ads could be used to detect firms that may discriminate based on age in hiring decisions. And of course, the methods we develop could be applied to evidence on discrimination against other groups.

One limitation of our work is that we can only learn about the role of age stereotypes that appear in the job ads studied. This could imply that there are stereotypes employers have about older workers that affect hiring, but on which our evidence is silent. On the other hand, thinking back to our two key 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. Certainly, if age-related stereotypes in job ads are being used to shape the applicant pools, it is the stereotypes in job ads that are of interest. And 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.

References

AARP. 2000. *American Business and Older Employees*. AARP: Washington, DC.

Adafre, Sisa F., and Maarten de Rijke. 2006. "Finding Similar Sentences Across Multiple Languages in Wikipedia." In *Proceedings of the EACL Workshop on New Text*. Trento, Italy

Ameri, Mason, Sean Edmund Rogers, Lisa Schur, and Douglas Kruse. "No Room at the Inn? Disability Access in the New Sharing Economy." Forthcoming in *Academy of Management Discoveries.*

Armstrong-Stassen, Marjorie, and Francine Schlosser. 2008. "Benefits of a Supportive Development Climate for Older Workers." *Journal of Managerial Psychology* 23(4): 419–437.

Baert, Stijn, Jennifer Norga, Yannick Thuy, and Marieke Van Hecke. 2016. "Getting Grey Hairs in the Labour Market. An Alternative Experiment on Age Discrimination." *Journal of Economic Psychology* 57: 86–101.

Banfi, Stefano, and Benjamín Villena-Roldán. 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.

Bendick, Marc, Jr., Lauren E. Brown, and Kennington Wall. 1999. "No Foot in the Door: An Experimental Study of Employment Discrimination Against Older Workers." *Journal of Aging & Social Policy* 10(4): 5–23.

Bendick, Marc, Jr., Charles W. Jackson, and J. Horacio Romero. 1997. "Employment Discrimination Against Older Workers: An Experimental Study of Hiring Practices." *Journal of Aging & Social Policy* 8(4): 25–46.

Button, Patrick. 2019. "Population Aging, Age Discrimination, and Age Discrimination Protections at the 50th Anniversary of the Age Discrimination in Employment Act." In S. Czaja, J. Sharit, and J. James (Eds.), *Current and Emerging Trends in Aging and Work*, 163–88. New York, NY: Springer.

Carlsson, Magnus, and Stefan Eriksson. 2019. "The Effect of Age and Gender on Labor Demand – Evidence from a Field Experiment." *Labour Economics* 59: 173–83.

Clark, Stephen. 2015. "Vector Space Models of Lexical Meaning." In S. Lapin and C. Fox (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 Gregory H. Dobbins. 1991. "The Influence of Subordinate Age on Managerial Actions: An Attributional Analysis." *Journal of Organizational Behavior* 12(5): 367–77.

Deming, David, and Lisa B. Kahn. 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, Dan Silverman, and Till M. von Wachter. 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., Chris M. Herbst, Dan Silverman, and Till von Wachter. 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., Kelly D. Higgins, and Maggie Clancy. 2000. "Justifications for Ratings of Older and Young Job Applicants: An Exploratory Content Analysis." *Experimental Aging Research* 26(3):

263–83.

Finkelstein, Lisa M., Michael J. Burke, and Nanbury S. Raju. 1995. "Age Discrimination in Simulated Employment Contexts: An Integrative Analysis." *Journal of Applied Psychology* 80(6): 652–63.

Finkelstein, Lisa M., and Michael J. Burke. 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.

Finkelstein, Lisa M., Katherine M. Ryan, and Eden B. King. 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., Amy J.C. Cuddy, Peter Glick, and Jun Xu. 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.

Fix, Michael, and Raymond J. Struyk, Eds. 1993. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C.: The Urban Institute Press.

Gaddis, S. Michael. 2018. "An Introduction to Audit Studies in the Social Sciences." In Gaddis, S. M. (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*. New York: Springer.

Gordon, Randall A., and Richard D. Arvey. 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, Zachary Hawley, and Aryn Taylor. 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, Zachary Hawley, Hal Martin, and Bo Liu. 2016. "Discrimination in Mortgage Lending: Evidence from a Correspondence Experiment." *Journal of Urban Economics* 92: 48–65.

Hellester, Miguel D., Peter Kuhn, and Kailing Shen. 2020. "The Age Twist in Employers' Gender Requests." *Journal of Human Resources* 55(2): 482–69.

Hendrick, Jennifer J., V. Jane Knox, William L. Gekoski, and Kate J. Dyne. 1988. "Perceived Cognitive Ability of Young and Old Targets." *Canadian Journal on Aging* 7(3): 192–203.

Hummert, Mary Lee, Teri A. Garstka, Jaye L. Shaner, and Sharon Strahm, S. 1994. "Stereotypes of the Elderly Held by Young, Middle-aged, and Elderly Adults." *Journal of Gerontology* 49(5): P240–9.

Hummert, Mary Lee, Teri A. Garstka, and Jaye L. Shaner. 1995. "Beliefs About Language Performance: Adults' Perceptions About Self and Elderly Targets." *Journal of Language and Social Psychology* 14(3): 235–59.

Jaeger, David. A, John M. Nunley, R. Alan Seals, and Eric Wilbrandt. 2020. "The Demand for Interns." NBER Working Paper No. 26729.

Johnson, Richard W., Janette Kawachi, and Eric K. Lewis. 2009. "Older Workers on the Move: Recareering in Later Life." Washington, DC: AARP Public Policy Institute.

Jurafsky Daniel, and James H. Martin. 2017. "Vector Semantics." In *Speech and Language Processing, Third Edition* (draft), https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf.

Karpinska, Kasia, Kène Henkens, and Joop Schippers. 2013. "Retention of Older Workers: Impact of Managers' Age Norms and Stereotypes." *European Sociological Review* 29(6): 1323–35.

Kite, Mary E., Kay Deaux, and Margaret Miele. 1991. "Stereotypes of Young and Old: Does Age Outweigh

Gender?” *Psychology and Aging* 6(1): 19–27.

Krings, Franciska, Sabine Sczesny, and Annette Kluge. 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., Martine Van Selm, Claartje L. ter Hoeven, and Rens Vliegenthart. 2016. “Reliable and Unproductive? Stereotypes of Older Employees in Corporate and News Media.” *Ageing and Society* 38(1): 166–91.

Kugelmass, Heather. “Just the Type with Whom I Like to Work: Two Correspondence Field Experiments in an Online Mental Health Care Market.” Forthcoming in *Society and Mental Health.*

Kuhn, Peter, and Kailing Shen. 2013. “Gender Discrimination in Job Ads: Evidence from China.” *Quarterly Journal of Economics* 128(1): 287–336.

Kuhn, Peter, Kailing Shen, and Shuo Zhang. 2018. “Gender-Targeted Job Ads in the Recruitment Process: Evidence from China.” NBER Working Paper No. 25365.

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.

Levin, William C. 1988. “Age Stereotyping: College Student Evaluations.” *Research on Aging* 10(1): 134–48.

McCann, Robert M., and Shaughan A. Keaton. 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 Lance Gray. 2002. “Stereotypes and Older Workers: The New Zealand Experience.” *Social Policy Journal of New Zealand* 18: 163–77.

Madera, Juan M., Michelle R. Hebl, and Randi C. Martin. 2009. “Gender and Letters of Recommendation for Academia: Agentic and Communal Differences.” *Journal of Applied Psychology* 94(6): 1591–99.

Maestas, Nicole. 2010. “Back to Work: Expectations and Realizations of Work after Retirement.” *Journal of Human Resources* 45(3): 718–48.

Marinescu, Ioana Elena, and Ronald Wolthoff. 2019. “Opening the Black Box of the Matching Function: The Power of Words.” *Journal of Labor Economics* 38 (2).

Maurer, Todd J., Frank G. Barbeite, Elizabeth M. Weiss, and Micheal Lippstreu. 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, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. “Efficient Estimation of Word Representations in Vector Space.” Unpublished paper, ICLR Workshop.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013b. “Distributed Representations of Words and Phrases and their Compositionality.” In *Advances in Neural Information Processing Systems* 26: 3111–19.

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. 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, Daniel Shoag, and Joshua Ballance. 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, Ian Burn, and Patrick Button. 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, Ian Burn, and Patrick Button. 2017. “Age Discrimination and Hiring of Older Workers.” *Federal Reserve Board of San Francisco Economic Letter* #2017-06.

Neumark, David, Ian Burn, and Patrick Button. 2016. “Experimental Age Discrimination Evidence and the Heckman Critique.” *American Economic Review Papers & Proceedings* 106(5): 303–08.

Neumark, David, Ian Burn, Patrick Button, and Nanneh Chehras. 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.

Nunley, John M., Adam Pugh, Nicholas Romero, and R. Alan Seals. 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, Richard Socher, and Christopher D. Manning. 2014. “GloVe: Global Vectors for Word Representation.” In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP).

Pitt-Catsouphes, Marcie, Micheal A. Smyer, Christina Matz-Costa, and Katherine Kane. 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.

Posthuma, Richard A., and Michael A. Campion. 2007. “Age Stereotypes in the Workplace: Common Stereotypes, Moderators, and Future Research Directions.” *Journal of Management* 35(1): 158–88.

Riach, Peter A., and Judith Rich. 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, Sheree Kwong See, W. Bryan Meneer, and Diane Trovato. 1992. “Age-Based Perceptions of Language Performance Among Younger and Older Adults.” *Communication Research* 19(4): 423–43.

Ryan, E. B., and Sheree Kwong See. 1993. “Age-Based Beliefs about Memory Changes for Self and Others across Adulthood.” *Journals of Gerontology* 48(4): 199–201.

Schmidt, Daniel F., and Susan M. Boland. 1986. “Structure of Perceptions of Older Adults: Evidence for Multiple Stereotypes.” *Psychology and Aging* 1(3): 255–60.

Singer, M. S. 1986. “Age Stereotypes as a Function of Profession.” *Journal of Social Psychology* 126(5): 691–92.

Stewart, Mark A., and Ellen Bouchard Ryan. 1982. “Attitudes toward Younger and Older Adult Speakers: Effects of Varying Speech Rates.” *Journal of Language and Social Psychology* 1(2): 91–109.

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., Elizabeth A. McCune, Marilena Bertolino, and Franco Fraccaroli. 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.

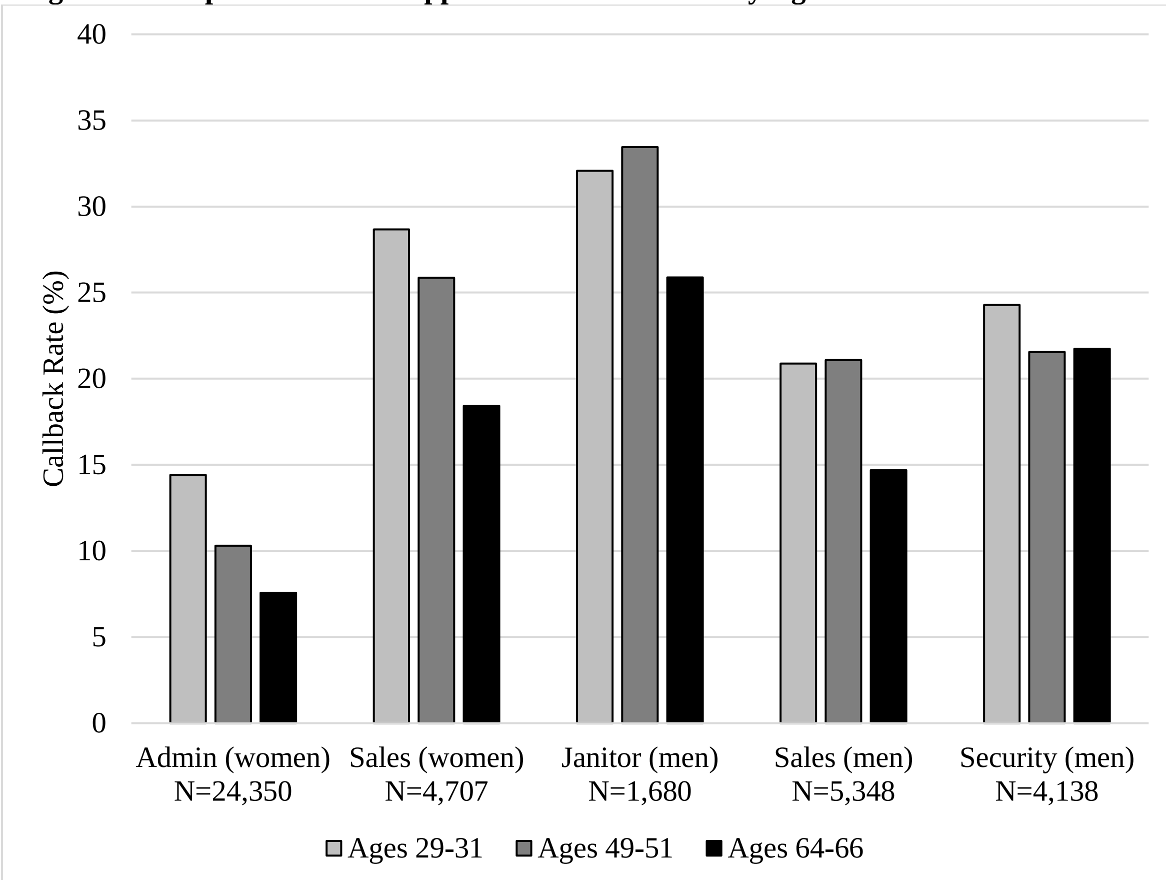
van Borm, Hannah, Ian Burn, and Stijn Baert. 2019. "What Does a Job Candidate's Age Signal to Employers?" IZA Discussion Paper No. 12849.

van Dalen, Harry, Kene Henkens, and J.J Schippers. 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 Janet Pennington. 1993. "Views about Age Discrimination and Older Workers." In P. Taylor, et al. (Eds.), *Age and Employment: Policies, Attitudes, and Practice*, 75–106. London: Institute of Personnel Management.

Weiss, Elizabeth M, and Todd J. Maurer. 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.

Figure 2: Visual Representation of a Hypothetical *Word2Vec* Neural Network

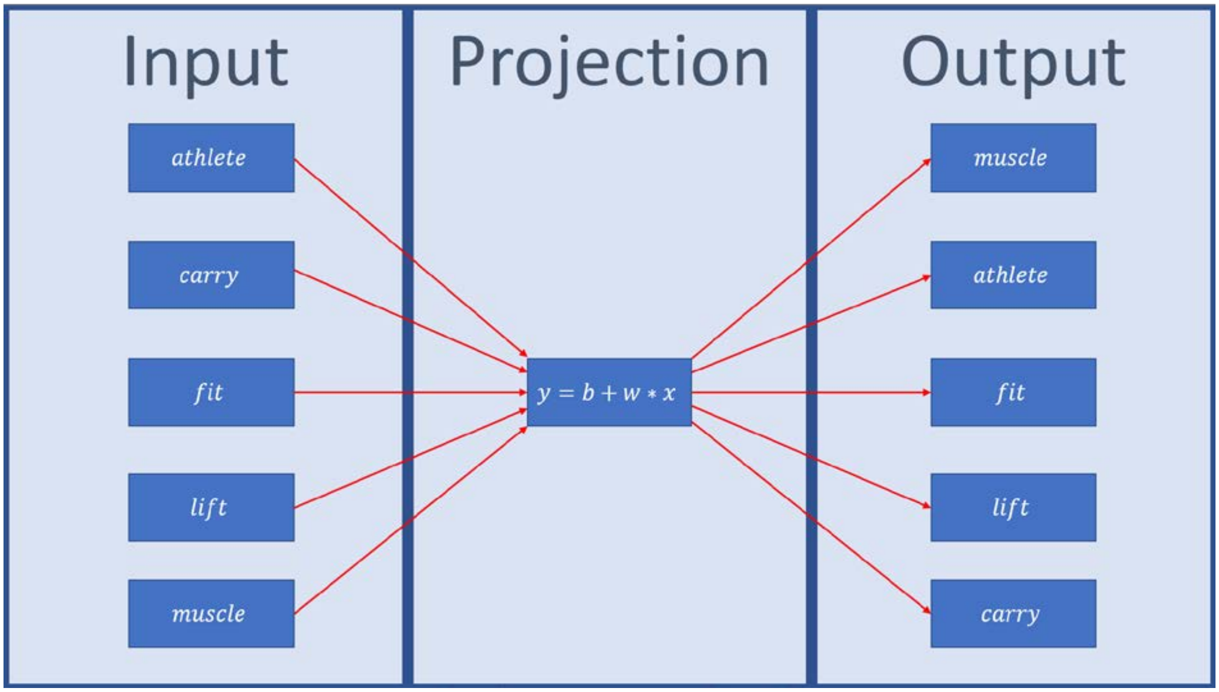
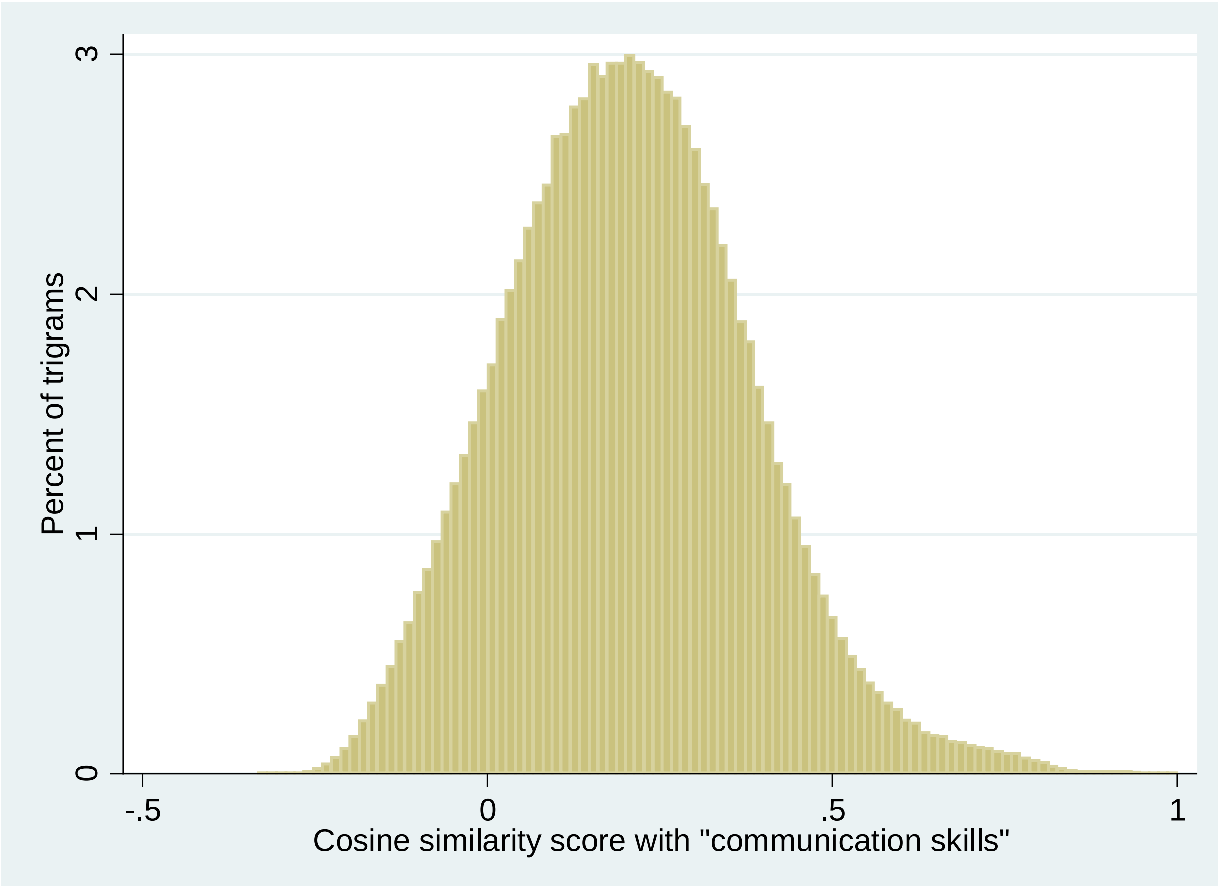


Figure 3: Example of the Distribution of Cosine Similarity (CS) Scores



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."

Figure 4: Text of a Job Ad

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.

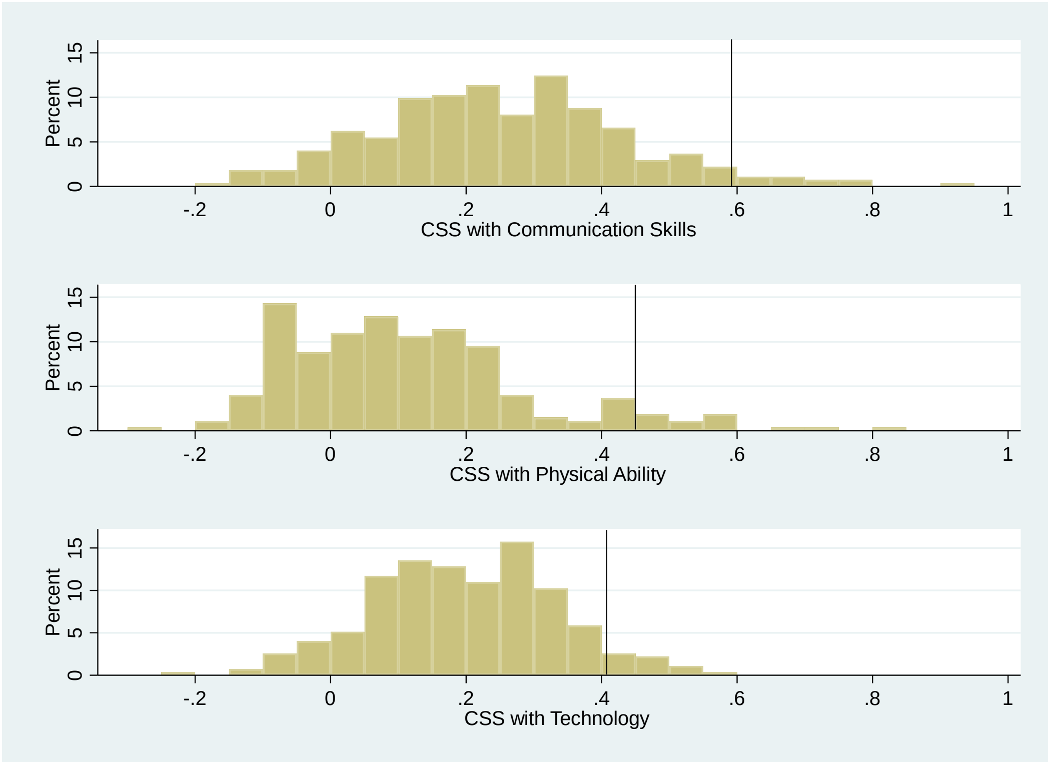
Requirements:

- 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.

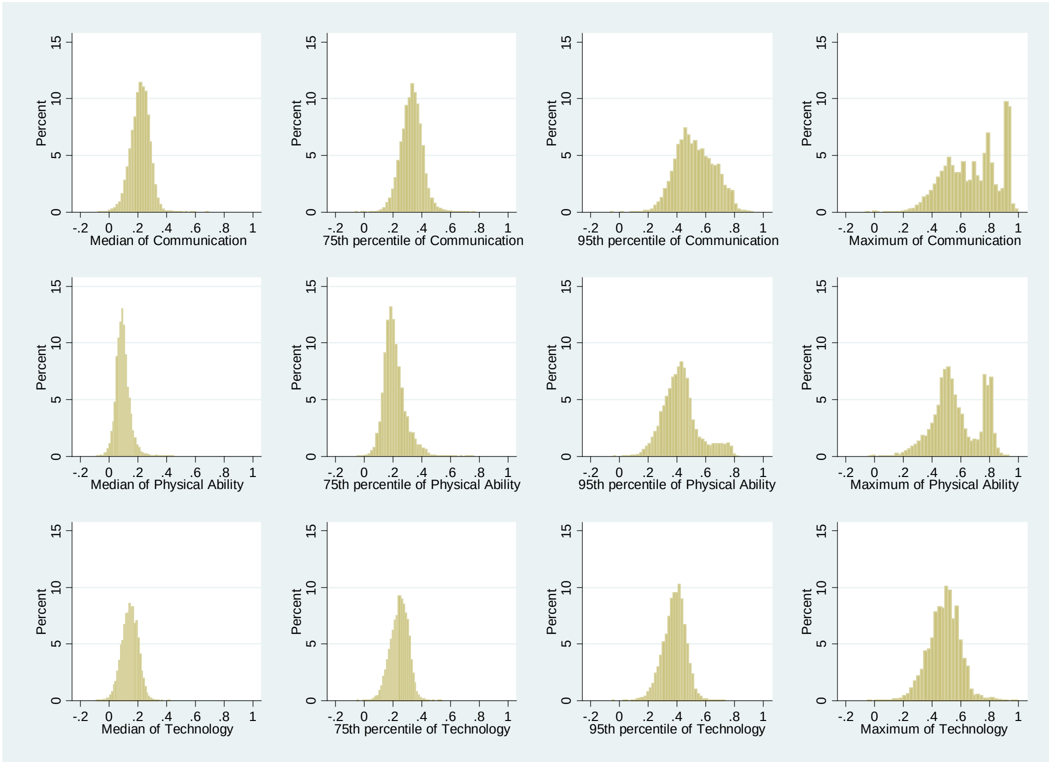
Note: Text drawn from actual job ad applied to in NBB, posted in June 2015. Company name and contact details have been removed.

Figure 5: Distributions of CS Scores within Example Ad (from Figure 4)



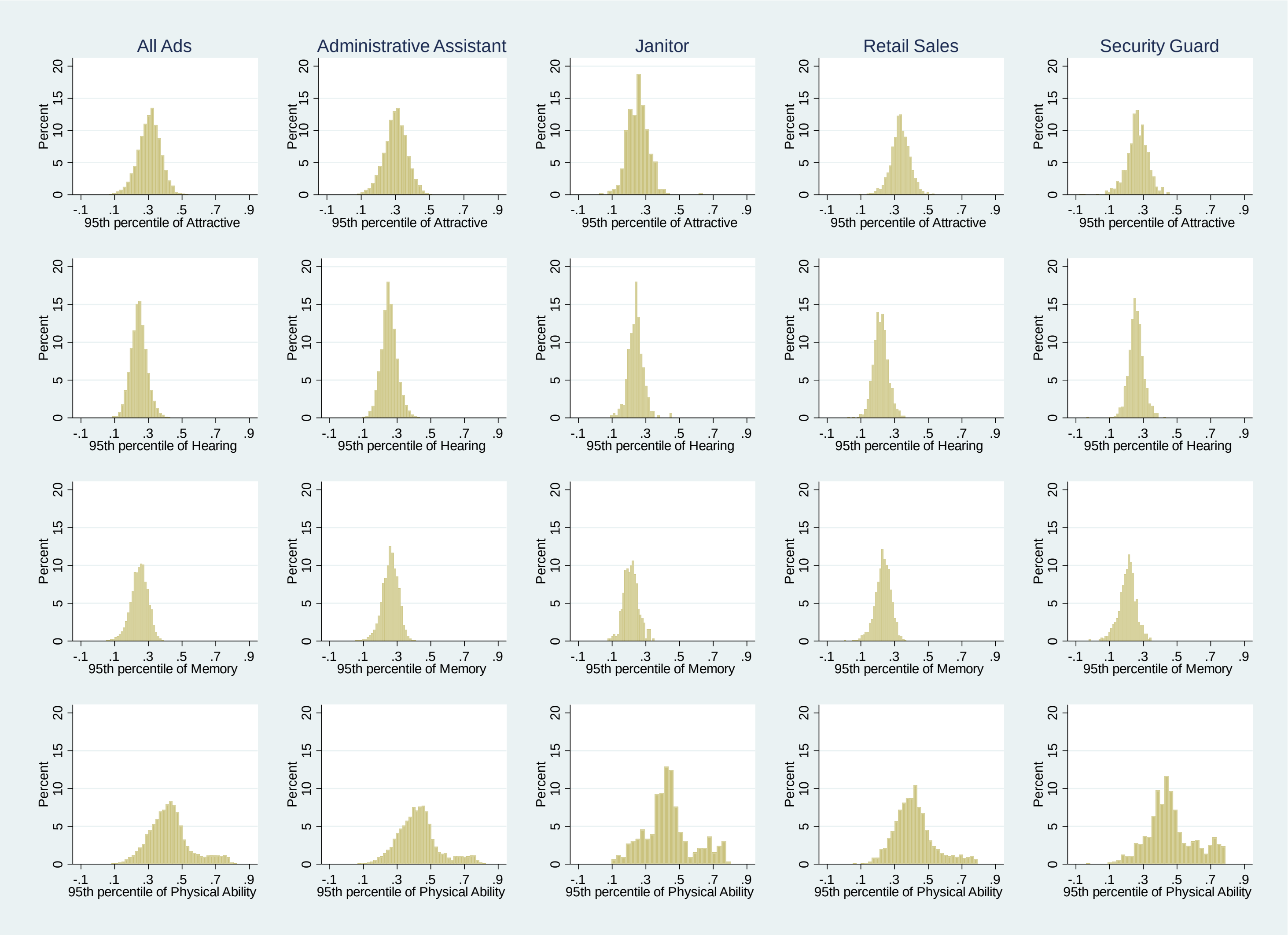
Note: Drawing on the text from the job ad in Figure 4, these histograms plot the distributions of CS scores for all trigrams in the example job ad, for three stereotypes. The solid line in each graph indicates the 95th percentile.

Figure 6: Alternative Percentile Distributions of CS Scores of Job-Ad Trigrams with Select Stereotypes



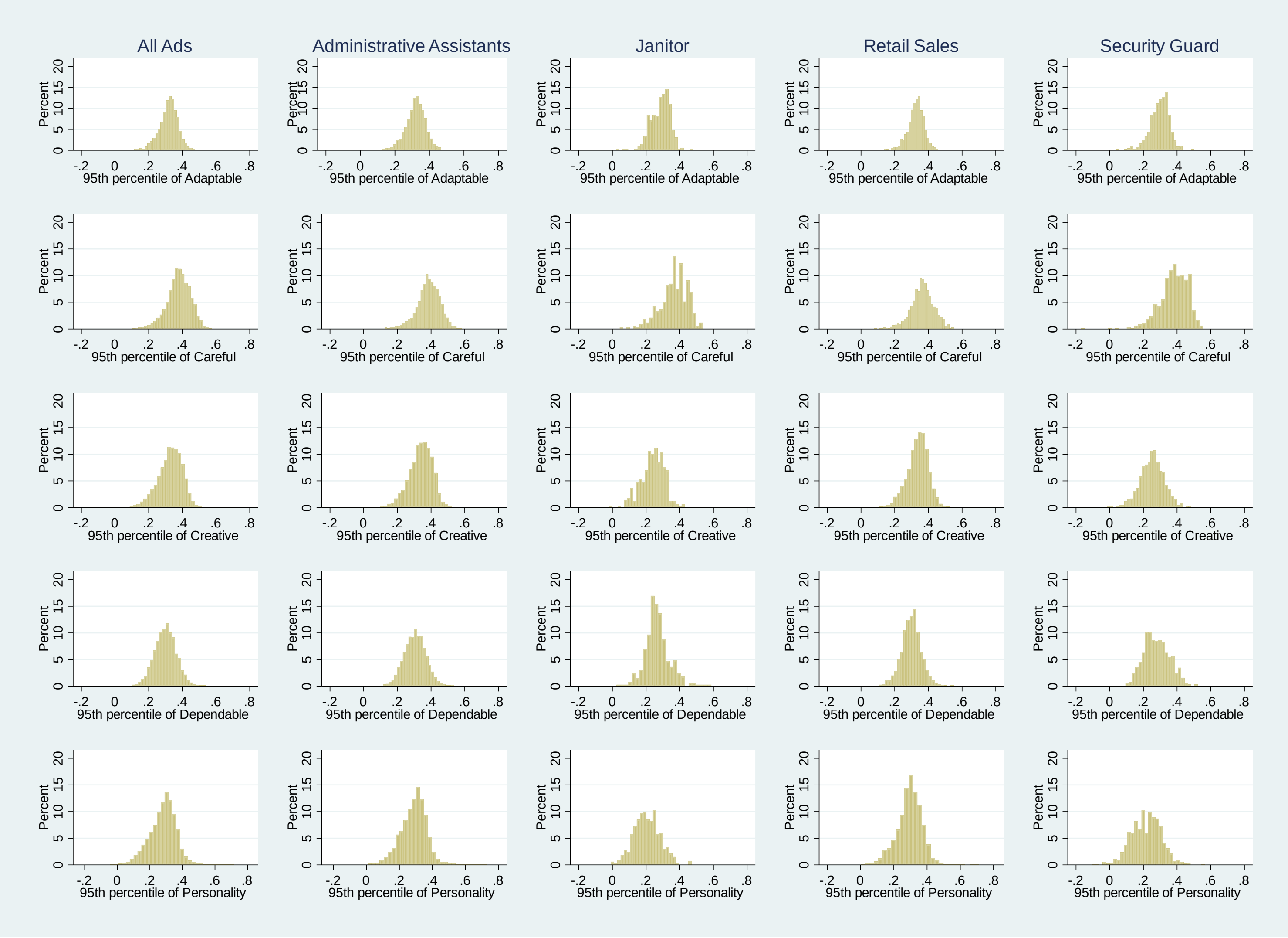
Note: Data come from the job ads collected in NBB. The distributions are for all the ads in our sample.

Figure 7A: Distributions of 95th Percentiles for Stereotypes Related to Health



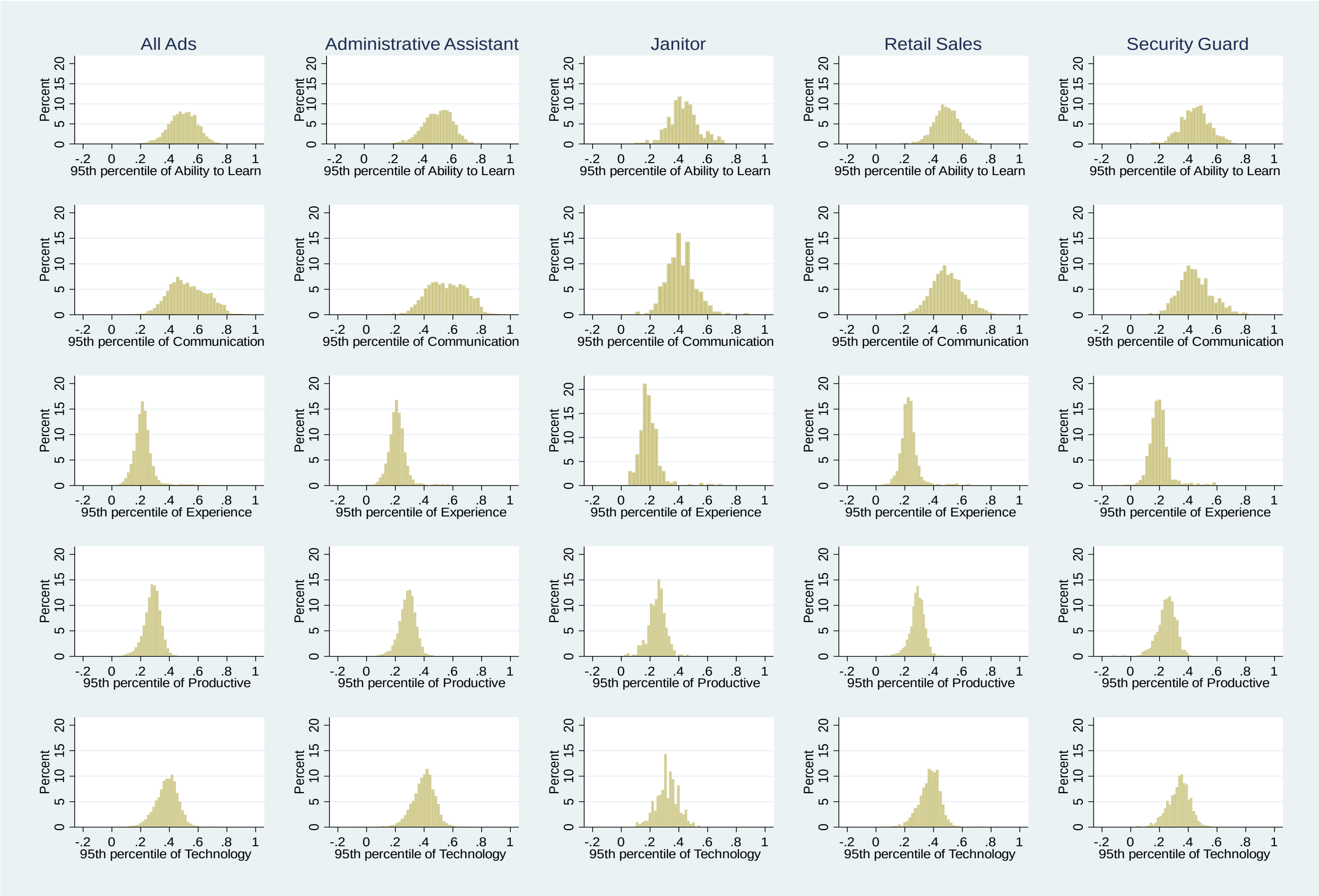
Note: Data come from the job ads collected in NBB. Each panel plots the distribution of CS scores at the 95th percentile for the job ads with each stereotype related to health. The first column contains the distribution of all the ads in our sample. The remaining columns disaggregate the job ads by occupation.

Figure 7B: Distributions of 95th Percentiles 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 95th 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.

Figure 7C: Distributions of 95th Percentiles 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 95th 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.

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)
	"talks of past," "focuses away from future toward past"	Kite et al. (1991)
	"less flexible," "more old-fashioned"	Stewart and Ryan (1982)
Careful	"think before they act"	Lyon and Pollard (1997) (p. 251)
	"older workers are more cautious than younger workers."	McCann and Keaton (2013)
	"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-Catsouphes et al. (2007) (p. 8)
	"more loyal to the organization" "more reliable"	Warr and Pennington (1993) (p. 89)
	"more stable"	Singer (1986)
	"more trustworthy"	Stewart and Ryan (1982)
Negative Personality	"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)
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)

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 at 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-Catsouphes 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)
	"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)

Table 4: Baseline Results by Gender, Age, and Occupation

		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)
Baseline discrimination		0.100	0.155	0.103	0.140	0.096	0.104	0.089	0.148	0.102	0.121
Health stereotypes	*Predicted sign*										
Attractive	Positive	-0.004	-0.008	-0.001	0.023	0.017	-0.017	0.019	0.018	-0.026	-0.029
		(0.007)	(0.021)	(0.007)	(0.015)	(0.022)	(0.017)	(0.012)	(0.034)	(0.014)	(0.020)
Hearing	Positive	0.002	0.003	0.001	0.021	-0.007	-0.010	**0.020***	-0.010	0.000	0.012
		(0.004)	(0.015)	(0.005)	(0.011)	(0.016)	(0.010)	**(0.010)**	(0.023)	(0.009)	(0.014)
Memory	Positive	-0.007	0.006	-0.009	0.010	-0.019	**0.030***	0.012	0.006	0.006	0.006
		(0.006)	(0.020)	(0.006)	(0.015)	(0.019)	**(0.012)**	(0.010)	(0.028)	(0.013)	(0.017)
Physical Ability	Positive	-0.010	-0.031	-0.001	0.013	**0.038***	**0.032***	0.022	0.002	0.026	-0.002
		(0.007)	(0.021)	(0.007)	(0.015)	**(0.018)**	**(0.015)**	(0.012)	(0.027)	(0.015)	(0.015)
Personality stereotypes	*Predicted sign*										
Adaptable	Positive	-0.012	-0.005	0.002	-0.006	-0.000	0.032	-0.021	-0.001	**0.031***	0.015
		(0.008)	(0.023)	(0.007)	(0.018)	(0.020)	(0.020)	(0.011)	(0.037)	**(0.016)**	(0.018)
Careful	Negative	-0.001	0.034	0.003	-0.025	**-0.042***	**-0.035***	-0.016	-0.014	**-0.031***	-0.001
		(0.007)	(0.023)	(0.007)	(0.017)	**(0.020)**	**(0.015)**	(0.013)	(0.029)	**(0.015)**	(0.018)
Creative	Positive	0.011	-0.029	0.003	0.013	0.015	-0.018	**-0.026***	0.035	-0.008	0.011
		(0.007)	(0.026)	(0.008)	(0.017)	(0.023)	(0.016)	**(0.013)**	(0.036)	(0.014)	(0.023)
Dependable	Negative	0.007	-0.001	**-0.016***	0.006	0.020	0.011	0.004	0.016	-0.001	0.006
		(0.007)	(0.019)	**(0.007)**	(0.014)	(0.016)	(0.015)	(0.009)	(0.024)	(0.014)	(0.014)
Personality	Positive/Negative	0.007	0.003	0.004	**-0.039****	0.019	0.021	-0.014	0.035	-0.002	-0.014
		(0.006)	(0.022)	(0.006)	**(0.015)**	(0.017)	(0.014)	(0.012)	(0.033)	(0.012)	(0.020)
Skills stereotypes	*Predicted sign*										
Ability to Learn	Positive	0.008	0.014	0.005	0.015	0.004	**-0.045***	-0.009	0.027	-0.022	-0.001
		(0.008)	(0.025)	(0.008)	(0.019)	(0.025)	**(0.018)**	(0.015)	(0.036)	(0.017)	(0.023)
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)
Experienced	Negative	-0.000	0.005	0.002	0.000	**0.033****	0.014	-0.012	**0.042***	-0.007	0.003
		(0.005)	(0.015)	(0.005)	(0.009)	**(0.012)**	(0.010)	(0.008)	**(0.019)**	(0.008)	(0.011)
Productive	Positive/Negative	0.002	0.010	0.006	0.007	**-0.049***	-0.022	0.014	-0.036	-0.002	0.005
		(0.007)	(0.020)	(0.008)	(0.015)	**(0.021)**	(0.016)	(0.013)	(0.037)	(0.012)	(0.020)
Technology	Negative	0.009	-0.018	0.005	-0.020	0.011	0.004	**0.021***	0.034	-0.009	-0.007
		(0.006)	(0.019)	(0.006)	(0.014)	(0.014)	(0.012)	**(0.009)**	(0.026)	(0.013)	(0.016)
N		6,822	986	7,321	1,856	311	1,612	954	318	1,680	932

Note: Table reports estimates of Equation [3]. 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.

Table 5: Analysis of Discrimination Against Younger Applicants

		Female		Male		
		Admin	Sales	Janitor	Sales	Security
		(1)	(2)	(3)	(4)	(5)
Baseline discrimination		0.048	0.074	0.115	0.077	0.090
Health stereotypes	*Predicted sign*					
Attractive	Negative	0.004	-0.000	0.007	-0.010	-0.000
		(0.003)	(0.010)	(0.022)	(0.009)	(0.012)
Hearing	Negative	-0.002	-0.009	**0.036***	-0.005	-0.005
		(0.002)	(0.007)	**(0.015)**	(0.006)	(0.008)
Memory	Negative	-0.001	0.011	0.024	0.001	0.009
		(0.003)	(0.008)	(0.020)	(0.008)	(0.009)
Physical Ability	Negative	-0.002	0.003	0.025	-0.006	0.001
		(0.003)	(0.009)	(0.022)	(0.009)	(0.009)
Personality stereotypes	*Predicted sign*					
Adaptable	Negative	-0.003	**0.028****	0.021	0.008	-0.002
		(0.003)	**(0.010)**	(0.026)	(0.009)	(0.011)
Careful	Positive	0.005	**-0.022***	-0.026	-0.004	0.002
		(0.003)	**(0.009)**	(0.022)	(0.009)	(0.010)
Creative	Negative	-0.003	0.016	-0.020	-0.015	0.010
		(0.003)	(0.008)	(0.023)	(0.008)	(0.012)
Dependable	Positive	**-0.006***	-0.001	0.003	-0.013	0.010
		(0.003)	(0.008)	(0.018)	(0.008)	(0.009)
Personality	Positive/Negative	-0.002	-0.010	0.018	-0.001	0.004
		(0.003)	(0.008)	(0.023)	(0.008)	(0.010)
Skills stereotypes	*Predicted sign*					
Ability to Learn	Negative	0.000	-0.007	**-0.056***	-0.004	-0.008
		(0.003)	(0.010)	**(0.027)**	(0.010)	(0.012)
Communication Skills	Positive/Negative	0.005	-0.007	**0.085***	0.014	-0.023
		(0.003)	(0.011)	**(0.037)**	(0.011)	(0.014)
Experienced	Positive	0.002	-0.003	-0.018	-0.008	0.005
		(0.002)	(0.006)	(0.015)	(0.006)	(0.006)
Productive	Positive/Negative	-0.004	-0.003	-0.011	0.015	-0.018
		(0.003)	(0.008)	(0.025)	(0.009)	(0.011)
Technology	Positive	0.002	-0.005	-0.035	-0.003	-0.001
		(0.003)	(0.008)	(0.018)	(0.007)	(0.009)
N		14136	2834	646	3274	1804

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.

Table 6A: Varying the Number of Words in Phrases, Women

	Middle-Admin			Middle-Sales			Old-Admin			Old-Sales		
# words in phrase	2	3	4	2	3	4	2	3	4	2	3	4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Health stereotypes												
Attractive	-0.006	-0.004	-0.007	-0.009	-0.008	-0.014	0.002	-0.001	-0.007	0.013	0.023	0.018
	(0.007)	(0.007)	(0.007)	(0.020)	(0.021)	(0.022)	(0.007)	(0.007)	(0.008)	(0.015)	(0.015)	(0.016)
Hearing	0.003	0.002	-0.001	0.009	0.003	0.001	-0.002	0.001	-0.003	0.023	0.021	**0.028***
	(0.005)	(0.004)	(0.005)	(0.015)	(0.015)	(0.014)	(0.005)	(0.005)	(0.005)	(0.012)	(0.011)	**(0.011)**
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
	(0.006)	(0.006)	(0.006)	(0.019)	(0.020)	(0.020)	(0.006)	(0.006)	(0.006)	(0.016)	(0.015)	(0.015)
Physical Ability	**-0.014***	-0.010	-0.010	**-0.051***	-0.031	-0.030	0.008	-0.001	-0.003	-0.001	0.013	0.016
	(0.007)	(0.007)	(0.007)	**(0.022)**	(0.021)	(0.020)	(0.007)	(0.007)	(0.007)	(0.017)	(0.015)	(0.015)
Personality stereotypes												
Adaptable	-0.008	-0.012	-0.008	0.007	-0.005	-0.002	-0.003	0.002	0.005	0.008	-0.006	-0.019
	(0.007)	(0.008)	(0.008)	(0.023)	(0.023)	(0.024)	(0.007)	(0.007)	(0.008)	(0.017)	(0.018)	(0.019)
Careful	-0.001	-0.001	-0.003	0.048*	0.034	0.026	-0.001	0.003	0.004	-0.006	-0.025	**-0.034***
	(0.007)	(0.007)	(0.007)	(0.022)	(0.023)	(0.023)	(0.007)	(0.007)	(0.007)	(0.018)	(0.017)	**(0.017)**
Creative	0.008	0.011	0.012	-0.027	-0.029	-0.015	-0.002	0.003	0.010	0.026	0.013	0.015
	(0.007)	(0.007)	(0.007)	(0.026)	(0.026)	(0.025)	(0.007)	(0.008)	(0.008)	(0.016)	(0.017)	(0.016)
Dependable	0.011	0.007	0.010	-0.005	-0.001	0.023	-0.014	**-0.016***	-0.003	-0.011	0.006	0.012
	(0.007)	(0.007)	(0.007)	(0.019)	(0.019)	(0.020)	(0.007)	**(0.007)**	(0.007)	(0.015)	(0.014)	(0.014)
Personality	0.006	0.007	0.007	0.012	0.003	0.016	0.012	0.004	0.003	**-0.035***	**-0.039****	**-0.033***
	(0.007)	(0.006)	(0.006)	(0.020)	(0.022)	(0.021)	(0.006)	(0.006)	(0.006)	**(0.015)**	**(0.015)**	**(0.014)**
Skills stereotypes												
Ability to Learn	0.014	0.008	0.012	0.028	0.014	0.028	-0.002	0.005	0.006	0.020	0.015	0.020
	(0.008)	(0.008)	(0.008)	(0.025)	(0.025)	(0.023)	(0.009)	(0.008)	(0.008)	(0.020)	(0.019)	(0.018)
Communication Skills	-0.015	-0.013	**-0.019***	-0.008	-0.012	-0.045	0.002	0.001	-0.007	0.005	0.007	-0.002
	(0.008)	(0.008)	**(0.008)**	(0.030)	(0.029)	(0.025)	(0.008)	(0.008)	(0.008)	(0.020)	(0.019)	(0.019)
Experienced	0.001	-0.000	-0.001	0.008	0.005	0.002	-0.003	0.002	0.005	-0.010	0.000	-0.007
	(0.005)	(0.005)	(0.005)	(0.014)	(0.015)	(0.015)	(0.005)	(0.005)	(0.005)	(0.009)	(0.009)	(0.010)
Productive	-0.002	0.002	0.002	-0.004	0.010	-0.020	0.006	0.006	-0.004	0.012	0.007	0.024
	(0.007)	(0.007)	(0.008)	(0.019)	(0.020)	(0.022)	(0.007)	(0.008)	(0.008)	(0.015)	(0.015)	(0.017)
Technology	0.009	0.009	0.007	-0.019	-0.018	0.001	0.008	0.005	0.003	**-0.030***	-0.020	-0.015
	(0.006)	(0.006)	(0.006)	(0.017)	(0.019)	(0.019)	(0.006)	(0.006)	(0.006)	**(0.014)**	(0.014)	(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.

Table 6B: Varying the Number of Words in Phrases, Men

	Middle-Janitor			Middle-Sales			Middle-Security			Old-Janitor			Old-Sales			Old-Security		
# words in phrase	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)
Health stereotypes																		
Attractive	-0.002	0.017	-0.007	-0.016	-0.017	-0.025	0.015	0.019	0.015	-0.021	0.018	0.032	**-0.037****	-0.026	-0.007	-0.019	-0.029	-0.020
	(0.019)	(0.022)	(0.024)	(0.017)	(0.017)	(0.017)	(0.012)	(0.012)	(0.014)	(0.035)	(0.034)	(0.037)	**(0.013)**	(0.014)	(0.013)	(0.018)	(0.020)	(0.021)
Hearing	-0.013	-0.007	-0.004	-0.012	-0.010	**-0.020***	**0.018***	**0.020***	**0.022***	-0.012	-0.010	-0.013	0.000	0.000	-0.003	**0.025***	0.012	0.003
	(0.013)	(0.016)	(0.014)	(0.011)	(0.010)	**(0.010)**	**(0.009)**	**(0.010)**	**(0.009)**	(0.024)	(0.023)	(0.020)	(0.010)	(0.009)	(0.009)	**(0.013)**	(0.014)	(0.015)
Memory	-0.011	-0.019	-0.025	0.024	**0.030***	**0.033****	0.009	0.012	0.004	-0.006	0.006	-0.016	-0.006	0.006	0.012	-0.008	0.006	0.020
	(0.017)	(0.019)	(0.022)	(0.012)	**(0.012)**	**(0.012)**	(0.009)	(0.010)	(0.010)	(0.030)	(0.028)	(0.031)	(0.014)	(0.013)	(0.013)	(0.017)	(0.017)	(0.016)
Physical Ability	**0.041***	**0.038***	0.025	0.017	**0.032***	0.028	0.023*	0.022	0.011	0.013	0.002	-0.012	0.005	0.026	0.006	0.001	-0.002	-0.014
	(0.017)	**(0.018)**	(0.020)	(0.016)	**(0.015)**	(0.016)	(0.011)	(0.012)	(0.012)	(0.024)	(0.027)	(0.030)	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)	(0.017)
Personality stereotypes																		
Adaptable	0.011	-0.000	0.010	0.030	0.032	0.033	-0.024	-0.021	-0.013	0.008	-0.001	-0.023	0.010	**0.031***	0.023	0.006	0.015	0.005
	(0.017)	(0.020)	(0.023)	(0.016)	(0.020)	(0.018)	(0.012)	(0.011)	(0.016)	(0.036)	(0.037)	(0.039)	(0.014)	**(0.016)**	(0.016)	(0.016)	(0.018)	(0.020)
Careful	**-0.051****	**-0.042***	-0.031	-0.026	**-0.035***	-0.033	-0.010	-0.016	-0.007	-0.028	-0.014	0.015	-0.016	**-0.031***	-0.007	-0.012	-0.001	0.011
	(0.019)	**(0.020)**	(0.019)	(0.017)	**(0.015)**	(0.018)	(0.012)	(0.013)	(0.014)	(0.030)	(0.029)	(0.030)	(0.015)	**(0.015)**	(0.015)	(0.016)	(0.018)	(0.019)
Creative	0.017	0.015	0.017	-0.023	-0.018	-0.027	-0.018	**-0.026***	**-0.030***	0.030	0.035	0.018	-0.008	-0.008	-0.013	0.014	0.011	-0.005
	(0.015)	(0.023)	(0.025)	(0.017)	(0.016)	(0.017)	(0.012)	**(0.013)**	**(0.013)**	(0.030)	(0.036)	(0.039)	(0.014)	(0.014)	(0.015)	(0.020)	(0.023)	(0.019)
Dependable	**0.038****	0.020	0.011	0.010	0.011	0.006	-0.001	0.004	0.003	**0.051***	0.016	0.015	0.007	-0.001	-0.011	**0.032***	0.006	0.006
	(0.014)	(0.016)	(0.017)	(0.015)	(0.015)	(0.015)	(0.010)	(0.009)	(0.008)	**(0.025)**	(0.024)	(0.024)	(0.012)	(0.014)	(0.013)	**(0.016)**	(0.014)	(0.014)
Personality	0.012	0.019	0.027	0.019	0.021	0.027*	-0.011	-0.014	-0.001	0.050	0.035	0.048	0.003	-0.002	-0.007	-0.024	-0.014	-0.008
	(0.014)	(0.017)	(0.021)	(0.014)	(0.014)	(0.014)	(0.012)	(0.012)	(0.014)	(0.032)	(0.033)	(0.036)	(0.011)	(0.012)	(0.011)	(0.017)	(0.020)	(0.019)
Skills stereotypes																		
Ability to Learn	0.007	0.004	-0.007	-0.022	**-0.045***	**-0.047****	-0.018	-0.009	-0.010	0.025	0.027	0.028	0.004	-0.022	-0.021	0.015	-0.001	-0.014
	(0.019)	(0.025)	(0.030)	(0.018)	**(0.018)**	**(0.017)**	(0.017)	(0.015)	(0.015)	(0.032)	(0.036)	(0.040)	(0.016)	(0.017)	(0.017)	(0.023)	(0.023)	(0.021)
Communication	0.003	0.002	0.010	-0.021	0.003	0.007	0.020	0.010	0.007	-0.067	-0.068	-0.044	0.001	0.015	0.012	0.011	0.004	0.015
Skills	(0.027)	(0.033)	(0.040)	(0.020)	(0.020)	(0.020)	(0.016)	(0.015)	(0.016)	(0.047)	(0.053)	(0.059)	(0.018)	(0.019)	(0.017)	(0.023)	(0.027)	(0.025)
Experienced	**0.021****	**0.033****	**0.029***	0.013	0.014	0.011	-0.010	-0.012	-0.011	**0.031***	**0.042***	**0.042***	-0.017	-0.007	-0.003	0.008	0.003	-0.003
	(0.008)	**(0.012)**	**(0.013)**	(0.010)	(0.010)	(0.010)	(0.008)	(0.008)	(0.009)	**(0.015)**	**(0.019)**	**(0.019)**	(0.009)	(0.008)	(0.008)	(0.010)	(0.011)	(0.013)
Productive	**-0.040***	**-0.049***	-0.022	-0.010	-0.022	-0.009	0.013	0.014	0.008	-0.024	-0.036	-0.047	0.015	-0.002	-0.001	-0.023	0.005	0.009
	(0.018)	**(0.021)**	(0.021)	(0.015)	(0.016)	(0.017)	(0.013)	(0.013)	(0.014)	(0.032)	(0.037)	(0.036)	(0.013)	(0.012)	(0.013)	(0.020)	(0.020)	(0.019)
Technology	-0.004	0.011	-0.004	0.009	0.004	-0.007	0.016	**0.021***	**0.022***	0.028	0.034	0.042	-0.004	-0.009	-0.009	-0.005	-0.007	-0.015
	(0.014)	(0.014)	(0.018)	(0.012)	(0.012)	(0.013)	(0.010)	**(0.009)**	**(0.010)**	(0.025)	(0.026)	(0.027)	(0.012)	(0.013)	(0.012)	(0.017)	(0.016)	(0.017)
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 Figure A1: Example Job Advertisement

◀ prev ▲ next ▶

Boutique Sales Associate (Larchmont Village in Hancock Park)

Beverly Blvd at Larchmont Blvd

(google map) (yahoo map)

compensation: **Commensurate with experience or performance.**

Larchmont Village/Hancock Park European-style boutique seeks a full-time sales person. Required: experience in retails sales. Above all, this is a sales position. Must have excellent social skills. A plus: experience with visual merchandising, administrative tasks, and social networking, including Facebook. Please email résumé in a PDF or Word document or in the email itself.

- Principals only. Recruiters, please don't contact this job poster.
- do NOT contact us with unsolicited services or offers

post id: 4959905041

posted: 2015-04-01 6:26pm

updated: 2015-04-01 6:26pm

Note: An example of a saved job advertisement from NBB.

Appendix Table A1A: Varying the Percentile, Women

	Middle-Admin			Middle-Sales			Old-Admin			Old-Sales		
Percentile	Median	95^{th}	Max	Median	95^{th}	Max	Median	95^{th}	Max	Median	95^{th}	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Health stereotypes												
Attractive	-0.011	-0.004	-0.006	0.043	-0.008	**-0.043***	-0.000	-0.001	0.005	0.029	0.023	-0.024
	(0.009)	(0.007)	(0.007)	(0.025)	(0.021)	**(0.019)**	(0.009)	(0.007)	(0.007)	(0.019)	(0.015)	(0.015)
Hearing	**-0.017***	0.002	0.000	0.008	0.003	0.011	-0.003	0.001	-0.004	0.010	0.021	0.006
	(0.007)	(0.004)	(0.004)	(0.023)	(0.015)	(0.012)	(0.007)	(0.005)	(0.005)	(0.016)	(0.011)	(0.009)
Memory	-0.005	-0.007	-0.005	-0.021	0.006	-0.010	**-0.016***	-0.009	0.000	-0.003	0.010	0.002
	(0.007)	(0.006)	(0.006)	(0.021)	(0.020)	(0.015)	**(0.007)**	(0.006)	(0.006)	(0.016)	(0.015)	(0.013)
Physical Ability	0.000	-0.010	0.001	0.032	-0.031	-0.024	-0.006	-0.001	0.011	0.005	0.013	-0.010
	(0.009)	(0.007)	(0.006)	(0.025)	(0.021)	(0.016)	(0.009)	(0.007)	(0.006)	(0.019)	(0.015)	(0.013)
Personality stereotypes												
Adaptable	-0.016	-0.012	-0.004	-0.033	-0.005	0.006	-0.009	0.002	-0.006	-0.004	-0.006	0.001
	(0.010)	(0.008)	(0.007)	(0.032)	(0.023)	(0.019)	(0.010)	(0.007)	(0.007)	(0.023)	(0.018)	(0.015)
Careful	0.002	-0.001	-0.007	-0.001	0.034	**0.038***	0.014	0.003	-0.005	**-0.045***	-0.025	0.002
	(0.009)	(0.007)	(0.006)	(0.031)	(0.023)	**(0.018)**	(0.010)	(0.007)	(0.006)	**(0.020)**	(0.017)	(0.014)
Creative	**0.030****	0.011	0.009	-0.007	-0.029	0.005	0.009	0.003	0.008	-0.022	0.013	0.015
	(0.010)	(0.007)	(0.006)	(0.033)	(0.026)	(0.018)	(0.009)	(0.008)	(0.007)	(0.024)	(0.017)	(0.012)
Dependable	0.008	0.007	0.012	-0.020	-0.001	0.023	0.001	**-0.016***	-0.002	-0.009	0.006	0.019
	(0.008)	(0.007)	(0.006)	(0.026)	(0.019)	(0.017)	(0.008)	**(0.007)**	(0.007)	(0.019)	(0.014)	(0.014)
Personality	-0.003	0.007	0.006	-0.007	0.003	0.013	-0.008	0.004	0.004	-0.016	**-0.039****	-0.017
	(0.009)	(0.006)	(0.006)	(0.029)	(0.022)	(0.014)	(0.008)	(0.006)	(0.005)	(0.020)	**(0.015)**	(0.012)
Skills stereotypes												
Ability to Learn	0.009	0.008	-0.003	-0.009	0.014	0.024	0.012	0.005	0.001	0.018	0.015	0.013
	(0.010)	(0.008)	(0.007)	(0.031)	(0.025)	(0.017)	(0.010)	(0.008)	(0.007)	(0.024)	(0.019)	(0.015)
Communication Skills	0.004	-0.013	-0.001	0.023	-0.012	**-0.052****	-0.006	0.001	0.002	0.028	0.007	0.000
	(0.011)	(0.008)	(0.007)	(0.036)	(0.029)	**(0.020)**	(0.011)	(0.008)	(0.007)	(0.027)	(0.019)	(0.016)
Experienced	0.009	-0.000	0.000	-0.039	0.005	-0.013	-0.002	0.002	-0.001	-0.005	0.000	-0.005
	(0.007)	(0.005)	(0.005)	(0.024)	(0.015)	(0.015)	(0.007)	(0.005)	(0.005)	(0.017)	(0.009)	(0.010)
Productive	**-0.028****	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	**-0.037****
	(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 A1B: Varying the Percentile, Men

	Middle-Janitor			Middle-Sales			Middle-Security			Old-Janitor			Old-Sales			Old-Security		
Percentile	Median	95th	Max	Median	95th	Max	Median	95th	Max	Median	95th	Max	Median	95th	Max	Median	95th	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Health stereotypes																		
Attractive	-0.018	0.017	0.004	0.002	-0.017	-0.020	-0.010	0.019	-0.015	**-0.097***	0.018	0.019	-0.022	-0.026	0.005	0.005	-0.029	-0.028
	(0.026)	(0.022)	(0.019)	(0.020)	(0.017)	(0.015)	(0.018)	(0.012)	(0.013)	**(0.044)**	(0.034)	(0.033)	(0.018)	(0.014)	(0.013)	(0.026)	(0.020)	(0.017)
Hearing	0.021	-0.007	**-0.028***	0.011	-0.010	-0.011	0.015	**0.020***	0.012	0.030	-0.010	-0.013	-0.008	0.000	0.006	0.002	0.012	0.015
	(0.019)	(0.016)	**(0.012)**	(0.017)	(0.010)	(0.008)	(0.013)	**(0.010)**	(0.007)	(0.029)	(0.023)	(0.019)	(0.014)	(0.009)	(0.009)	(0.020)	(0.014)	(0.014)
Memory	-0.028	-0.019	0.007	0.002	**0.030***	0.017	-0.022	0.012	0.005	-0.057	0.006	0.001	-0.009	0.006	**0.021***	-0.019	0.006	0.019
	(0.024)	(0.019)	(0.015)	(0.015)	**(0.012)**	(0.011)	(0.013)	(0.010)	(0.010)	(0.032)	(0.028)	(0.027)	(0.014)	(0.013)	**(0.010)**	(0.019)	(0.017)	(0.014)
Physical Ability	0.026	**0.038***	0.029	0.004	**0.032***	0.000	-0.003	0.022	0.003	0.054	0.002	-0.021	**0.040***	0.026	0.013	-0.008	-0.002	-0.021
	(0.021)	**(0.018)**	(0.018)	(0.017)	**(0.015)**	(0.012)	(0.013)	(0.012)	(0.011)	(0.034)	(0.027)	(0.028)	**(0.016)**	(0.015)	(0.011)	(0.016)	(0.015)	(0.014)
Personality stereotypes																		
Adaptable	0.049	-0.000	-0.003	-0.011	0.032	0.020	0.000	-0.021	-0.006	**0.113****	-0.001	-0.031	0.000	**0.031***	0.005	0.005	0.015	0.035
	(0.027)	(0.020)	(0.020)	(0.021)	(0.020)	(0.015)	(0.017)	(0.011)	(0.013)	**(0.039)**	(0.037)	(0.035)	(0.018)	**(0.016)**	(0.014)	(0.026)	(0.018)	(0.020)
Careful	-0.032	**-0.042***	-0.025	-0.020	**-0.035***	-0.007	0.023	-0.016	-0.009	-0.022	-0.014	0.017	-0.007	**-0.031***	-0.011	**0.050***	-0.001	-0.014
	(0.025)	**(0.020)**	(0.016)	(0.019)	**(0.015)**	(0.014)	(0.016)	(0.013)	(0.013)	(0.036)	(0.029)	(0.027)	(0.019)	**(0.015)**	(0.013)	**(0.025)**	(0.018)	(0.016)
Creative	-0.032	0.015	-0.026	-0.042	-0.018	-0.009	-0.008	**-0.026***	-0.020	-0.033	0.035	-0.040	-0.021	-0.008	-0.007	-0.044	0.011	-0.010
	(0.027)	(0.023)	(0.019)	(0.023)	(0.016)	(0.013)	(0.018)	**(0.013)**	(0.014)	(0.048)	(0.036)	(0.029)	(0.021)	(0.014)	(0.009)	(0.026)	(0.023)	(0.020)
Dependable	0.004	0.020	0.012	0.017	0.011	0.018	-0.004	0.004	0.000	0.004	0.016	-0.003	-0.003	-0.001	0.006	-0.011	0.006	0.017
	(0.024)	(0.016)	(0.013)	(0.017)	(0.015)	(0.013)	(0.015)	(0.009)	(0.009)	(0.031)	(0.024)	(0.023)	(0.018)	(0.014)	(0.011)	(0.021)	(0.014)	(0.014)
Personality	0.039	0.019	0.026	0.025	0.021	0.009	-0.003	-0.014	-0.021	0.071	0.035	0.055	0.020	-0.002	-0.008	0.034	-0.014	-0.032
	(0.026)	(0.017)	(0.018)	(0.017)	(0.014)	(0.009)	(0.006)	(0.006)	(0.006)	(0.040)	(0.033)	(0.029)	(0.017)	(0.012)	(0.009)	(0.026)	(0.020)	(0.019)
Skills stereotypes																		
Ability to Learn	-0.008	0.004	-0.013	0.006	**-0.045***	-0.015	-0.000	-0.009	0.008	-0.027	0.027	0.006	-0.009	-0.022	-0.017	-0.031	-0.001	0.012
	(0.025)	(0.025)	(0.019)	(0.023)	**(0.018)**	(0.014)	(0.019)	(0.015)	(0.014)	(0.037)	(0.036)	(0.030)	(0.019)	(0.017)	(0.013)	(0.028)	(0.023)	(0.019)
Communication	-0.028	0.002	**0.048***	-0.019	0.003	-0.001	-0.013	0.010	0.012	-0.009	-0.068	0.033	-0.007	0.015	0.016	-0.040	0.004	0.019
Skills	(0.031)	(0.033)	**(0.022)**	(0.022)	(0.020)	(0.013)	(0.022)	(0.015)	(0.014)	(0.053)	(0.053)	(0.040)	(0.018)	(0.019)	(0.013)	(0.029)	(0.027)	(0.021)
Experienced	0.039	**0.033****	**0.020***	-0.006	0.014	0.013	-0.002	-0.012	-0.009	0.015	**0.042***	**0.036***	-0.020	-0.007	-0.012	0.009	0.003	0.011
	(0.020)	**(0.012)**	**(0.009)**	(0.018)	(0.010)	(0.009)	(0.015)	(0.008)	(0.010)	(0.032)	**(0.019)**	**(0.015)**	(0.015)	(0.008)	(0.009)	(0.020)	(0.011)	(0.010)
Productive	-0.037	**-0.049***	-0.023	0.021	-0.022	-0.019	-0.005	0.014	**0.033***	-0.036	-0.036	0.009	0.011	-0.002	0.000	-0.002	0.005	-0.001
	(0.029)	**(0.021)**	(0.017)	(0.022)	(0.016)	(0.015)	(0.019)	(0.013)	**(0.013)**	(0.040)	(0.037)	(0.033)	(0.019)	(0.012)	(0.013)	(0.028)	(0.020)	(0.021)
Technology	0.022	0.011	-0.026	0.026	0.004	0.002	0.030	**0.021***	0.002	-0.001	0.034	0.000	0.014	-0.009	-0.012	0.037	-0.007	-0.028
	(0.024)	(0.014)	(0.017)	(0.016)	(0.012)	(0.011)	(0.016)	**(0.009)**	(0.012)	(0.037)	(0.026)	(0.027)	(0.015)	(0.013)	(0.011)	(0.021)	(0.016)	(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 Tables A1A. * indicates that the estimate is statistically significant at the 5% level. ** indicates that the estimate is statistically significant at the 1% level.

Appendix Table A2: Text of Trigrams at the 95th Percentile of CS Scores within Job Ads

		Trigrams closest to 1 standard deviation above mean				
	Mean 95th 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
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 95th percentile of all trigrams in a job ad. We identify the trigram that is closest to the mean 95th 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 95th percentile (columns 2 and 3) and the two trigrams immediately above 1 standard deviation above the mean 95th 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.