

Stereotyped Language in Job Ads and Decisions by Older Workers to Apply for a Job: A Field Experiment*

Ian Burn
University of Liverpool

Daniel Firoozi
University of California-Irvine

Daniel Ladd
University of California-Irvine

David Neumark
University of California-Irvine, NBER, and IZA

May 2022

Abstract

In this study, we create a bank of job ads for administrative assistants/secretaries, retail sales associates/cashiers, or security guards. We construct these job ads using language from real job ads collected in Neumark et al. (2019a). We build on prior work by Burn et al. (forthcoming), who identify common age stereotypes from the research literature in industrial psychology, use machine learning to calculate the relationship between the text of the job ads and specific age stereotypes, and then test whether job-ad language related to the stereotypes predicts hiring discrimination against older workers in a correspondence study.

In our new experiment, we post these job ads in cities while randomly varying whether or not ageist language regarding age-related stereotypes is included in the job text. The innovation in this study is that the job ads are artificial, and we are studying the responses of real job searchers – in contrast to a correspondence study in which the job searchers are artificial, and we study the responses of real employers.

We find significant evidence that ageist language impacts the decisions of older workers to apply for jobs. Job ads that feature ageist language attract younger applicants on average than job ads that do not feature ageist language. The change in the age distribution of applicants is large.

Note: This draft is based on partial data collection through January 2022; the full data collection will be complete by June 2022 at which point the paper will be finalized.

* This research was funded by the Sloan Foundation. We received helpful comments from seminar participants at the University of Kentucky and Texas A&M. Any views or opinions expressed are solely those of the authors. The Pre-Analysis Plan for this project was registered on Open Science Framework on December 31, 2020. The research was approved by the UCI Office of Research Institutional Review Board on October 18, 2019: HS# 2015-2017, modification application #26404.

Introduction

Lengthening work lives for those able to work is a crucial part of the policy response to population aging. Because many seniors transition to “partial retirement” or “bridge jobs” at the end of their careers (Cahill et al., 2006; Johnson et al., 2009) or return to work after a period of retirement (Maestas, 2010), reducing age discrimination in hiring is critical to lengthening working lives. There is an extensive body of research that documents the extent to which employer discrimination against older workers in labor markets may impede the working lives of older workers using correspondence studies to test for discrimination in hiring (e.g., Bendick et al., 1997, 1999; Lahey, 2008; Farber et al., 2019; Neumark et al., 2019a, 2019b). This research focuses on measuring employer behavior – specifically, whether there is less hiring of qualified older workers – and generally finds evidence consistent with hiring discrimination against older workers. There is little work, however, that studies how workers respond to age discrimination in the labor market.

In this study, we create a bank of job ads for administrative assistants/secretaries, retail sales associates/cashiers, or security guards.¹ We construct these job ads using language from real job ads collected in Neumark et al. (2019a). We post these job ads in cities while randomly varying whether or not ageist language regarding age-related stereotypes is included in the job text. The innovation in this study is that the job ads are artificial, and we are studying the responses of real job searchers – in contrast to a correspondence study in which the job searchers are artificial, and we study the responses of real employers.²

The potential for age stereotypes or other language in job ads to deter applications from older job seekers is real. An extreme version of such language is stating maximum experience levels in job ads. This occurred recently in *Kleber v. Carefusion Corp.*, where the job ad requested “3 to 7 years (no more

¹ Neumark et al. (2019a) show that these jobs are fairly important for hiring of older workers..

² We obtained IRB approval to post these ads, subject to an IRB-approved protocol to do two things: (i) to quickly inform applicants that the job is not available, so as not to interrupt their job search; and (ii) subsequently, to inform those from whom we have collected data of their inclusion in an experiment, and allow them to opt out of their data being used. (This is the standard protocol in experiments involving real job searchers; see Krause et al., 2012).

than 7 years) of relevant legal experience,” language that will clearly act to exclude many older applicants.³ More generally, the U.S. 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). Beyond that, organizations like AARP suggest to members that “[d]espite protections by the Age Discrimination in Employment Act of 1967 (ADEA), employers have gotten cleverer in masking what is age discrimination” by using ageist phrases in job ads (Brenoff, 2019).

We find significant evidence that ageist language, even when it is not blatant or specifically age-related, impacts the decisions of older workers to apply for jobs. Job ads that feature ageist language attract younger applicants on average than job ads that do not feature ageist language. We test the effect of language related to stereotypes about three different skills: communication skills, physical ability, and technology skills. When we change the language in a job ad for a single stereotype, we find significant reductions in the age of applicants to job ads featuring ageist language related to communication skills or technology, with applicants being three years younger on average. We find somewhat weaker statistical evidence that changing only one phrase related to physical ability reduces the age of applicants. Our evidence suggests that there may be thresholds above which workers react, though. We find the most significant changes in the age distribution of applicants when we change all three job requirements. This induces a large and significant change in the behavior of older workers, with ads containing all three ageist phrases attracting job applicants that are between two-and-a-half and four-and-a-half years younger on average. This change in the age distribution of applicants is large. We also find evidence that ageist language in job ads significantly reduces the median and 75th percentile of age (among the distribution of

³ 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). Surprisingly, the court ruled in favor of the defense in this case, reaching a new interpretation that the ADEA does not authorize job applicants to bring a disparate impact claim. (See Button, 2019.)

applicants) age and the share of applicants over 40.

In our view, we add significant new evidence to the extensive research literature on discrimination. As the first paper to examine how discrimination in hiring impacts job search behavior using a field experiment, we provide evidence on the effect of discrimination on the behavior of workers, whereas the research literature on discrimination focuses on the behavior of *employers*. Building on the results from Neumark et al. (2019a), who showed that employers discriminate against older workers, and Burn et al. (forthcoming), who showed that the discriminating employers used ageist language in job ads, this paper answers the question of whether job seekers respond to these phenomena – which may also explain why employers use ageist stereotypes in job ads. Our experiment shows that workers respond to subtle shifts in the language of job ads that might signal that an employer holds ageist stereotypes about older workers or is otherwise less interested in hiring older workers.

Our experimental evidence that ageist stereotypes in job ads discourage older applicants from applying for jobs thus illustrates the potential for this potentially subtle form of age discrimination in the labor market.⁴ Age discrimination that deters older workers from applying for jobs has the same effect as direct age discrimination applied to job applicants; both reduce the employment of older workers. Strikingly, our evidence from this new experiment, combined with prior evidence on age discrimination in hiring from the correspondence study (Neumark et al., 2019a), suggests that the two forms of age discrimination could be of similar magnitudes – although the evidence from both experiments is specific to the experimental conditions, and may not generalize to actual incidence of age discrimination in hiring and age-related stereotypes in job ads in the broader labor market.

Our evidence has significant policy implications regarding age discrimination and its enforcement. Utilizing ageist language in job ads may be rational for employers, despite it being illegal to discriminate against older workers. Shaping the applicant pool by discouraging older applicants has a

⁴ Our evidence should be viewed as another dimension of age discrimination in hiring – one that has not been studied or detected in the research literature that tests for hiring discrimination, mainly using correspondence studies. These include Baert et al. (2016); Bendick et al. (1997, 1999); Carlsson and Eriksson (2019); Farber et al. (2017, 2019); Lahey (2008); Neumark et al. (2016, 2019a, 2019b); and Riach and Rich (2006, 2010).

potential benefit for discriminatory employers, because of the incentives created by age discrimination laws. A lower representation of older workers in their applicant pool can justify a lower representation of older workers among employees, making it easier to rebut an allegation of age discrimination in hiring. More generally, employers who do not want to hire older workers might, in order to avoid unnecessary search costs, discourage older workers from applying by signaling their ageism. To think about this another way, in the legal system, hiring discrimination cases based on age (or, similarly, other groups) typically hinge on shortfalls of older workers among hires relative to the applicant pool. But if job-ad language deters older workers from applying, these shortfalls may be obscured, and the courts may need to weight other evidence more heavily – including both job-ad language as the source of lower applications from older workers, and comparisons with other benchmarks to assess whether hiring of older workers is notably lower at the firm in question.

To address age discrimination from stereotyped job-ad language that discourages older workers from applying for jobs, there are two tools the EEOC could utilize. First, it could issue guidance to employers on language to avoid that might be interpreted as discouraging older workers from applying. There exists a duty of care for employers to knowingly avoid using language which may deter older workers from applying. Therefore, our evidence provides a basis for further guidance regarding the usage of ageist stereotypes in job ads which may shift employer behavior. Second, the EEOC might consider flagging for potential investigation firms that use age-stereotyped language in their job ads, recognizing that, for these firms, discrimination may be occurring even in the absence of shortfalls between the share of older applicants hired and the share of older workers who apply for jobs. Thus, rigorous evidence on the role of ageist language in job ads could potentially influence policy to reduce age discrimination in hiring and contribute to lengthening work lives.

Previous Related Literature

Very few studies in labor economics explore job ads, and fewer still focus on discrimination. Among studies of issues other than discrimination, Modestino et al. (2016) use text data from job ads to document that “downskilling” occurred during the recovery from the Great Recession, with firms

reducing skill requirements in their job ads. Deming and Kahn (2018) use text data in job ads to measure how ten different skills relate to wages. Marinescu and Wolthoff (2020) match text data from job ads to job application data to study the matching process between jobs and applicants. Kuhn and Shen (2013) and Kuhn et al. (2018) explore how gender preferences feature explicitly or implicitly in job ads in China, and Hellester et al. (2020) explore age and gender preferences in job ads in China and Mexico.

A small number of studies are closer in spirit to ours in that they run experiments manipulating job ads and study responses of job seekers. He et al. (2021) study how job flexibility conditions influence job application behavior. Flory et al. (2015) study job seeker responses to jobs with variation in competition and uncertainty about pay. And finally Flory et al. (2019) examine how signaling interest in employee diversity affects interest among minority candidates (as well as firm selection of candidates).

Two studies, to date, connect the text of job ads to measured discriminatory behavior of employers.⁵ Tilcsik (2011) identifies words in job ads related to masculine stereotypes (decisive, aggressive, assertive, and ambitious) and links those to hiring outcomes in a correspondence study of discrimination against gay men.⁶ And, in the most systematic approach, Burn et al. (forthcoming) identify common age stereotypes from the research literature in industrial psychology, use machine learning to calculate the relationship between the text of the job ads and specific age stereotypes, and then test whether job-ad language related to the stereotypes predicts hiring discrimination against older workers in a correspondence study. As already noted, the present paper builds on this prior work.

There has been no research on how the ageist language in job ads affects the decisions of older workers to apply. What is known about how job applicants read job ads for bias focuses exclusively on gender bias in job ads. Gaucher, Friesen, and Kay (2011) found that job ads for male-dominated

⁵ Though they did not focus on job ads, Hanson et al. (2011) and Hanson et al. (2016) study language used by mortgage originators and connect this language to their behavior. Hanson et al. (2011) 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.

⁶ In an early small study, Wax (1948) found that summer resorts in Ontario, Canada, were more likely to discriminate against Jewish customers (based on names) requesting accommodations if they used phrases like “restrictive clientele” in their advertising.

occupations used masculine wording (i.e., words associated with male stereotypes, such as leader, competitive, dominant) more frequently than advertisements for female-dominated occupations, and women find job advertisements less appealing when they contain more masculine than feminine wording (Bem and Bem, 1973; Gaucher et al., 2011). However, these findings are based on laboratory experiments that ask how subjects perceive job-ad language, whereas our research uses a field experiment that studies the behavior of actual job seekers.⁷

Conceptual Framework and Model

Why might employers use stereotyped language in job ads? One hypothesis is that employers who discriminate based on age use stereotyped language to try to shape the applicant pool. Using language that conveys positive stereotypes related to young workers might discourage older workers from applying (as might language conveying negative stereotypes related to older workers – although that seems less likely and is, in fact, less common in our data). Employers may introduce this language via job requirements that are correlated with age, natural to use in job ads, and not so blatant as to make the age discrimination clear.

This discouragement from applying would lead to the underrepresentation of older applicants in the applicant pool, and is potentially valuable to a discriminating employer because the probability of a hiring age discrimination claim and of an adverse outcome for the employer is smaller when the ratio of older applicants to younger applicants is lower.⁸ Employers could use job-ad language this way regardless

⁷ There is also research suggesting that, in other contexts, job seekers respond to job-ad language, including Belot et al. (2018) and Banfi et al. (2019) on posted wages; Hellester et al. (2020) on gender requests; and Ibanez and Riener (2018), Leibbrandt and List (2018), and Flory et al. (2021) on affirmative action or diversity statements in recruitment materials or job ads.

⁸ In legal cases, the most compelling data on hiring discrimination comes from comparing hiring rates of the group in question (e.g., older workers) relative to the applicant pool. Hiring charges under the U.S. Age Discrimination in Employment Act (ADEA) made up nearly 5% of total ADEA charges in 2020 – more than double the percentage under Title VII (protecting women, minorities, etc.) or the Americans with Disabilities Act. (This is based on authors' computations using EEOC statistics available at <https://www.eeoc.gov/statistics/statutes-issue-charges-filed-eeoc-fy-2010-fy-2020>, viewed January 18, 2022.) The representation of hires among applicants is important in anti-discrimination enforcement, as the EEOC uses a “4/5ths” rule (the ratio of the selection rate for the group in question to the group with the highest selection rate) as “a practical means of keeping the attention of the enforcement agencies on serious discrepancies in rates of hiring, promotion and other selection decisions” (U.S. Equal Employment Opportunity Commission, 1979).

of the nature of age discrimination, and, in the case of statistical discrimination, whether or not the language is related to the assumptions they make about older workers (e.g., they might assume older workers will leave the firm sooner). In either case, employers might use ageist language in job ads to deter older workers from applying.

A second hypothesis, which is more complex, is also 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, which may well be true on average but of course not of each applicant.

While social scientists are interested in the nature of discriminatory behavior, both statistical and taste discrimination are illegal under U.S. law. Not surprisingly, language in job ads that refers to age either explicitly or “mechanically” is illegal in the United States. The U.S. 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).⁹

The legality of less blatant job-ad language with job requirements that reflect age stereotypes and is associated with lower hiring of older workers is more complex. On the one hand, 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.” (See U.S. Equal Employment Opportunity Commission, n.d.(a).) On the other hand, job requirements that are based on factors related to age are not necessarily illegal. The legality of job requirements related to age generally requires an employer to show that the use of these requirements is

⁹ European Union law also bars age discrimination. To the best of our knowledge it is less explicit about the forms of discrimination barred, and it also differs in not protecting older workers per se, but rather barring discrimination based on age generally. See Lahey (2010) and European Commission (2000).

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.” (See Federal Register, n.d.) In other words, the law recognizes that characteristics of workers that are related to age can sometimes be legitimate for employers to consider.

Indeed, the law even goes further, as in some rare cases, employers can even use age as an explicit criterion if it is inherently related to a requirement for the job that is related to age but hard to assess independently. This requires that age can be shown to be a “bona fide occupational qualification” (BFOQ) that is “reasonably necessary to the normal operation of the business.” (U.S. Equal Employment Opportunity Commission, n.d.(b)). A key example is *Hodgson v. Greyhound Lines, Inc.*, where the company was sued for having a maximum hiring age. Greyhound prevailed by establishing that driving ability is essential to passenger safety, that older hires would be less safe drivers (because achieving maximum safety took 16-20 years of experience), that some abilities associated with safe driving deteriorate with age, and that these changes are not detectable by physical examination (which could otherwise be a substitute for an age criterion). (See U.S. Court of Appeals, 7th Circuit, 1974.)¹⁰

Our evidence cannot speak directly to the question of taste vs. statistical discrimination or whether the job requirements would be viewed as legal. Indeed, we do not study employer behavior in our experiment, although we do use job-ad language from real employers. Rather, in our experiment, we ask how job seekers respond to job-ad language containing ageist stereotypes. A response could mean either that the language is perceived as directly reflecting age bias – aversion to hiring older workers – or that the language is perceived as “biased” because it puts older workers at a disadvantage because they may be less likely to satisfy the stated job requirement.

Similarly, our evidence does not speak to whether a stated job requirement would be legal. What our evidence *does* address is whether age stereotypes expressed in job ads affect the likelihood that older

¹⁰ As discussed by Combs (1982), the issue of the rights of older workers vs. public safety have figured prominently in court decisions regarding age as a BFOQ under the ADEA.

job seekers apply for jobs, likely by signaling to job applicants that older workers are less likely to be hired. Thus, our evidence can reveal the potential for employers to use job-ad language to discriminate against older workers in hiring, and the potential adverse impact on older job applicants.

Model

Here we describe the behavior of job seekers that we use to interpret our evidence. When deciding whether to apply to a job, potential workers read the job ad and decide whether the potential benefit outweighs the costs of applying. Suppose the utility of job j to person i is $U_{ij} = \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0,1)$.¹¹ The cost of applying for a job is c . The potential benefit of a job is determined by posted wage and the probability of getting a job offer (callback).¹² For younger workers this is $p_y = b$ ($0 < b \leq 1$) if young, and $p_o = b(S)$ ($0 \leq b(S) \leq b$, $b'(S) < 0$) if old. That is, the probability of a job offer for an older worker is a function of how age-stereotyped (S) is the job ad. An example of a function with $0 \leq b(S) < b$, is:

$$(1) \quad b(S) = b \cdot e^{-\eta \cdot S} \quad (\eta > 0) \quad .$$

A young person applies if $b \cdot \varepsilon > c$ (dropping the i and j subscripts) or $\varepsilon < -c/b$. Given the distributional assumption, the probability of applying is $A_y = \Phi(-c/b)$, where ϕ denotes the standard normal density. An old person applies if $b(S) \cdot \varepsilon > c$ (dropping the i and j subscripts) or $\varepsilon < -c/b(S)$, so $A_o = \Phi(-c/b(S))$.¹³

In this paper, we are interested in estimating $\partial A_o / \partial S$. For old applicants,

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

Theoretically, since $b'(S) < 0$, $\partial A_o / \partial S < 0$, this establishes that there will be a negative response of older applicants to more stereotyped job ads.

¹¹ The variance can be fixed without loss of generality.

¹² Given our experimental setting, low-skilled jobs being posted on an internet job board, we can assume the wages of these jobs are all the minimum wage, and hence ignore wage variation across jobs.

¹³ We assume that $\partial A_y / \partial S = 0$, or young people do not respond to the stereotyped language. We could have the probability of an offer for a young applicant *increasing* in S – i.e., the opposite direction from old people – and the qualitative conclusion is the same.

Methods

To test whether older workers respond to ageist language in job ads (i.e., is $\partial A_o / \partial S < 0$?), we conduct an experiment where we manipulate S and observe how the applicant pool changes. In our experiment, we post job ads in three occupations in 15 U.S. cities, randomly varying the inclusion of age-related stereotypes in the text of the job ad. The job ads are artificial, and we study the responses of real job searchers. This allows us to test for differences in the applicant pool when otherwise similar ads use age-related stereotypes vs. age-neutral language.

Selecting the Cities and Occupations

Much of this project builds on the experiment conducted in Neumark, Burn, and Button (2019a, hereafter NBB). These cities were selected in NBB due to their large size, their geographic distribution across the U.S., and because they have different population age distributions. For this experiment, we added three more cities with a large online presence for the job board we use. For each city, we post our job ad on an online job board, setting the hiring firms' locations to the central business district. The cities in the experiment are shown in Figure 1 (which also provides additional information on the data collection in the experiment).

We use three of the four occupations from the original study: retail sales (mixed-gender), administrative assistant (female-dominated), and security guards (male-dominated). These occupations are low-skilled, with jobs often advertised using online job boards. These jobs are also very relevant for older workers seeking new employment. As shown in NBB, all three occupations were in the top decile of jobs in terms of the proportions of older people hired.^{14,15}

Selecting the Stereotypes

¹⁴ Looking at the distribution of the share of 62-70 year-olds hired recently (tenure less than five years) across all occupations, the percentiles for males in the occupations we use were 96.6 for retail salespersons and 93.3 for security guards and gaming surveillance officers. The percentiles for females were 100 for secretaries and administrative assistants, 96.4 for receptionists and information clerks, and 95.2 for retail salespersons.

¹⁵ We omit the janitor jobs also included in NBB, because for them the evidence of age discrimination was less clear-cut, and there are many fewer janitor job advertisements posted online.

To select the stereotypes we use in our experiment, we start with a list of ageist stereotypes from the industrial psychology literature (see Burn et al., forthcoming). These are listed in Table 1. Among these, we selected stereotypes that we believed met the following criteria. First, the stereotype is commonly expressed job-ad language about the ideal or preferred candidate skills or attributes; we did not want to focus on stereotypes that are not often included in job ads (e.g., hearing and memory), even if employers hold these stereotypes based on the industrial psychology literature.

Second, we focus on stereotypes for which we had evidence of a correlation between discrimination and the stereotype (from Burn et al., forthcoming) and evidence that the stereotype captures a skill that employers view older workers as less likely to possess (from van Borm et al., 2020).¹⁶

Third, older workers should be aware that employers held that stereotype. As evidence, we drew on various reports put out by AARP; see Brenoff (2019) and Terrell (2019).

Our final list of stereotypes is three skills or abilities for which older workers are stereotyped as deficient: communication skills, physical ability, and technological skills.

Designing the Job Ads

We created 15 templates per occupation using actual ads collected in NBB as a guide to creating our experimental job ads.¹⁷ We supplemented the sample of ads from NBB with recent real ads posted on job boards in the sample cities to capture contemporaneous patterns of behavior. We selected a handful of ads to use as our base to create a template. We copied the format of the ad (location of blocks of text, types of bullet points, and style of text) to ensure our template was similar in appearance to others on the website. The text of each ad was rewritten to give enough details about the company and the position to appear realistic, but not enough details to suggest a specific company. We stripped the ad of all identifying information, so there is no identifiable link between the ad posted and the template we created. We modified the requirements of the jobs to reduce the number of applicants they potentially exclude. All

¹⁶ We did not require this evidence for all three occupations or for both genders, but just for some subset.

¹⁷ This number describes the templates used to design the job ads. In the actual implementation of the experiment the number expands slightly because of additional cities added (as described below). But the different information in these additional templates refers only to location.

ads are written to have flexible hours, competitive pay, and the availability of part-time and full-time positions (at the employee's choice). For half of the templates, we include the requirement that applicants must have a high school diploma (randomized by template). Figure 2 provides an example of a job ad for each of our occupations.

The treatment and control ads differ in the job requirements (denoted in bold with asterisks in each template in Figure 2), with three sentences assigned to be either a treatment phrase (stereotyped) or a control phrase (not stereotyped). The requirements we manipulate have to do with a candidate's communication skills, physical ability, and technology skills. Our control phrases express job requirements that are also appropriate for the job but use age-neutral language not related to these age-stereotyped skills or abilities, while our treatment phrases use language highly related to these ageist stereotypes.

Creating Stereotyped Job Requirements

We use two methods to generate sentences highly related to ageist stereotypes, focusing on constructing sentences that were highly related to only one of the three stereotypes we use. The first uses measures of semantic similarity generated by machine learning methods. Drawing on Burn et al. (forthcoming), we calculate the semantic (cosine) similarity of thousands of phrases to communication skills, physical ability, and technology skills. From this list, we construct our treatment sentences. We iteratively edited the sentences to ensure that only the cosine similarity score of the manipulated stereotype substantively differed between the treatment and control phrases. For example, if the treatment language related to communication skills was also highly related to the stereotype about personality, we identified which words in the sentence were highly related to personality and selected synonyms that were less related to personality. Our control sentences were created to express requirements for similar jobs without referring to ageist stereotypes about skills or abilities. We iteratively removed phrases that were highly related to our stereotypes to minimize the semantic similarity. The sentences for the treatment and control groups are listed in columns (3) and (4) of Table 2.

Figure 3 illustrates the distribution of the treatment and control phrases in the distribution of all text from the job ads collected in Burn et al. (forthcoming). The key insight from this figure is that the control phrases are close to the median and thus should not be regarded as ageist by the average job-seeker reading the text, while the treatment phrases are coming from higher in the distribution, close to the 75th percentile.¹⁸

Our second treatment conveys bias by using ageist language identified by AARP as the text related to communication skills, physical ability, and technology skills. We select three AARP examples that correspond to our respective stereotypes: “cultural fit,” “energetic person,” and “digital native” (Brenoff, 2019; Terrell, 2019). We adapted the language to fit our job ads and created three sentences, one for each stereotype (Table 2, column 5). Using the text about cultural fit, we created the phrase “You must be up-to-date with current industry jargon and communicate with a dynamic workforce” to reflect stereotypes about communication skills, emphasizing the communication aspect of fitting in. Using the text about energetic persons, we created the sentence “You must be a fit and energetic person” to reflect stereotypes about physical ability. Using the text about digital natives, we created the sentence “You must be a digital native and have a background in social media” to reflect stereotypes about technology skills by emphasizing social media.

We vary the combination of treatment and control phrases used in a job ad to create six job ads from each template: one control ad and five treatment ads. In our control ad, we use all three control phrases to express the skill requirements in language unrelated to ageist stereotypes. Four of the treatment ads utilize machine learning derived stereotyped phrases. We have three ads where we use the stereotyped phrase for communication skills, physical ability, or technological skills (i.e., only one at a time) and the control phrases for the other two stereotypes, and there is one ad where we use all three treatment phrases. In the AARP treatment, we use all three treatment phrases.

¹⁸ This is slightly lower than the types of phrases analyzed in Burn et al. (forthcoming), which focused on phrases above the 90th percentile. But the usage of phrases closer to the 75th percentile provides greater insight into the types of phrases that are more common to observe in job ads.

Validating the Treatment vs. Control Differences

While the AARP language is quite blatant, an obvious question is how well the stereotyped vs. neutral phrases generated by the machine learning convey the intended stereotypes. In the language of epidemiology, we would like our treatment ads to have high “sensitivity” (conveying ageist stereotypes) and “specificity” (conveying information about the specific ageist stereotype intended). We assess this in two ways.

First, Figures 4a-4c illustrate how the semantic similarity differs across the templates for the treatment and control job ads, and they show that our treatment job ads do activate the intended stereotypes. In these figures, words have been aggregated up to three-word phrases to ensure that we measure semantic meaning more accurately. Information on the distribution of all phrases found in the ads in Burn et al. (forthcoming) is shown in grey, information for the treatment ads is shown with dashed black lines, and information for the neutral ads with solid black lines. The figures show the median to 99th percentile range and the average (with plotting symbols).

These figures indicate that biased (treatment) templates have considerably higher 99th percentiles than the control templates, as well as higher means (and medians, although less so). The implication of the differences in the means and especially the upper tails of the distributions is that the treatment ads we write using the stereotyped language do, in fact, create ads with notably stronger age stereotypes. In addition, we see – importantly – that our treatment ads only generate a shift in similarity for the stereotypes we are seeking to activate, hence isolating those stereotypes in the job ads. Finally, note that the actual “collected” ads are more similar to the treatment ads – reflecting the fact that actual job ads often use ageist stereotypes (as documented in NBB).

The second way we assessed the validity of the treatment vs. control ads as activating the intended stereotypes was to conduct a validation exercise using Amazon MTURK. We found that the control phrases were not perceived as ageist by applicants, and treatment phrases were perceived as more ageist than the control phrases. The AARP phrases were perceived as the most ageist, with the machine

learning phrases intermediate between the AARP and control phrases, as we would expect. The survey and results are detailed in Burn et al. (2021).

The summary table is provided in Figure 5, which provides a graphical depiction of the answers from the MTURK survey participants. Across the three blocks of the survey that solicited respondents' self-assessments of age bias, their predictions of previous respondents' answers, and their predictions of the answers of respondents over the age of 50, our results were consistent. The participants, on average, strongly disagreed with the notion that anyone would perceive the control phrases as biased against workers over the age of 50. Respondents rated the physical and technology-biased phrases derived from our cosine similarity score index as more biased than the control phrases, but viewed the communication skills stereotyped phrases as roughly identical to the controls. Opinions of the AARP-derived treatment phrases were starker, as all three were rated as far more age biased than their respective control counterparts. The absence of evidence for bias for the language related to communication skills may reflect the fact that older workers are not always stereotyped as having worse communication skills but are sometimes, as Table 1 showed, perceived as having better communication skills. In that sense, one might view the evidence of ageist ratings for the physical ability and technology-related stereotypes but not the communications stereotype as further confirmation that respondent perceptions accord with the industrial psychology literature. (Note that the cosine similarity scores from the machine learning do not detect positive vs. negative uses of the language.) The regression results reported in Burn et al. (2021) confirmed these results.

These results, like those in Figures 4a-4c, imply that our phrases capture real ageist sentiments and will be perceived as such by job applicants, so our results should be informative about the effect of ageist language on job ads on job applicants' behavior.

Posting the Job Ads

We have a total of 18 ads that we post in a city, six per occupation. We stagger the posting of ads to leave two weeks in between the taking down of one ad and the posting of the next within each city. To avoid p-hacking, we initially planned to run the experiment for 54 weeks, with the schedule of posting

pre-registered.¹⁹ The cities were split such that each week ads are posted in five of the cities, while the other ten are rested. To maximize the number of potential applicants, one ad was to be posted each weekday (Monday through Friday).²⁰ The rotation of the ads posted is staggered such that there are eight weeks between the same occupation's ad appearing in the same city with different treatment statuses.

This was a complicated process. The job board we used for the experiment makes money from fees for posting job ads, and hence is sensitive to fake ads, ads used for phishing, etc. In the course of the experiment, we encountered problems if we tried to use the same credit card to pay for ads in different cities, or used the same IP address for posting ads in different cities. In addition, there seem to be human “checkers” for each city on the job board, who monitor for highly similar ads or ads that appear to be from fictitious companies.

To get around the payments problem, we used a very large number of gift cards, so we would never use one more than four times.²¹ Even this required some workarounds, as the websites for some gift cards made it difficult or impossible to register a large number of cards from the same IP addresses over a short period of time – sometimes prompting impossible to resolve “are you a robot” questions or tasks. This is apparently because gift cards are used by those who steal credit card information,²² or others (like money launderers) who want to avoid detection.²³ We thus had to experiment to identify gift cards that did not have this constraint. To get around the problem with IP addresses, we purchased many cell phones and sim cards for each city, using pay-as-you-go plans that randomize the IP address each time the service is restarted. Figure 6 gives you an idea of what was involved.²⁴

¹⁹ As discussed below, we anticipated some difficulties in placing ads in some cities, and the PAP explicitly called for a period subsequent to the initial 54 weeks when we would re-try placing these ads, in the same order.

²⁰ Initially, all ads were randomized to either Monday or Tuesday, but we had to switch to five days a week to avoid triggering a moderator response. The job ad board suggests not to post ads more often than every 48 hours.

²¹ We used \$90 gift cards for the large cities that required \$45 to post, \$70 for the cities that required \$35, and \$100 gift cards for the cities that required \$25 to post.

²² See, e.g., <https://www.forbes.com/sites/laurengensler/2017/01/11/gift-cards-money-laundering/?sh=5498ac6f1449>.

²³ See, e.g., <https://www.fraud-magazine.com/article.aspx?id=4294967696>.

²⁴ Our university grants administrator was frequently puzzled by the receipts submitted for reimbursement.

There was no way around the human checkers. A number of our ads were flagged by the job board as spam and taken down before they had been active for a week, or our payment method was rejected, leading to a delayed or skipped job posting. If the ad was taken down before we received responses, we began to repost it at the end of the study, starting in week 55. For city-occupation cells where multiple ads have been taken down, we repost them in the order that they were originally meant to be posted, still leaving one week in between each ad. This draft does not cover the full set of reposted ads, but rather all ads from the initial 12 months plus a few more weeks; it will be updated when the data collection is complete. The responses from these additional ads are not included in this draft.

For two cities (Boston and Pittsburgh), we were unable to post many ads, due to flagging. Because of this, and because the budget allowed it, we replaced these two cities and added two additional cities (Seattle, Washington, D.C., Minneapolis/Saint Paul, and San Diego), early on in the experiment after we encountered problems. These cities were selected due to having higher numbers of job postings on the job board, which increased the likelihood that ads were being seen. Furthermore, we chose to add more than one city to replace these in case problems emerged in other cities, based on our early experience with posting ads in Boston. Because these cities were not specified in the Pre-Analysis Plan (PAP), in our final draft, we will report results both with and without these additional cities; here, we report results for all cities.

Consistent with our PAP, we collect responses to our ads that we received within one week of the posting. We found, early in the experiment (when we were testing our procedures), that very few responses arrived after one week. Additionally, with our design and schedule, no two ads based on the same template were ever concurrently available on the job posting board.

Collecting Responses

Usually, applicants sent us their resumes when they replied to our job ad. To reduce the cost of applying for our fake job, we informed applicants that they were not selected for an interview and that

we had decided to go with another candidate, within 24 hours of their application, via an email.²⁵ While it is rare for employers to inform their applicants of a negative outcome, we believe that this is important to reduce possible costs to participants.²⁶

Calculating Applicant Age

Our primary outcome of interest is the age of applicants. We calculate the age of our applicants based on the available information listed on the resume. The first method to calculate age is based on the year of high school graduation.²⁷ Assuming that an individual was approximately 18 years old when they graduated high school, age is calculated as

$$\text{Age} = \text{Date of Job Post} - \text{Year of HS Graduation} + 18.$$

If the applicant does not provide a year of high school graduation, we calculate age based on the earliest date of work experience listed on the resume.²⁸ Age is calculated as

$$\text{Age} = \text{Date of Job Post} - \text{Year of First Job} + 16.²⁹$$

Applicants were assigned the oldest age calculated across these methods.³⁰

A concern is that ageist language may cause older applicants to hide their age, so the above methods to calculate age may undercount the number of older workers applying to the job ads with ageist

²⁵ Emails read: “Thank you for your interest in this position. Unfortunately we will not be pursuing your application at this time.” If someone expressed interest in applying *and* had a question, we provided the same response. If someone responded and *only* had a question about the ad, we did not reply nor did we have resume data to include (nor an email address from the resume). We did not respond to recruiters. If a job placement agency (e.g., refugee resettlement) sent a resume on someone’s behalf, we sent the response addressed in the third person (“we will not be pursuing X’s application”).

²⁶ To try to avoid spam responses and to make sure that the applicants have read the job ad, we included a manipulation check. Each ad contained a cue that applicants were asked to respond to that does not appear out of place in a job ad, such as “Please indicate which days of the week you are available to work.” However, a majority (62.8%) did not respond to the cue.

²⁷ The date of college graduation is another viable way of measuring age. However, because individuals may go back to college somewhat later, it seems likely that inferring age based on, say, year of college graduation minus 22 would tend to understate age. Nonetheless, we will test how robust our results are to calculating age using college graduation date when that is all the information we have, perhaps supplemented by weeding out cases with a year of college graduation that post-dates substantial work experience.

²⁸ Additionally, we collected the earliest non-work listed year on their resume and calculated Age as being 18 in that year. If applicants explicitly listed their age or year of birth we would record their age as such.

²⁹ This is probably best interpreted as “minimal possible age” assuming one did not start working before age 16.

³⁰ E.g., for an applicant with a 2021 job posting date who had a high school graduation year of 2014, an earliest working year of 2007, and the earliest non-work year listed on their resume of 2010, we would assign them as being 30 years-old.

language. To address this concern, we use a binary indicator to record whether or not we can determine an applicant's age from the information on the resume. If ageist language causes older applicants to manipulate their resumes to obscure their age, we should be able to capture this effect by comparing the share of applicants whose age we cannot ascertain for job ads with ageist language and job ads without ageist language.

Empirical Analysis

To test whether ageist language changes the composition of the applicant pool, our primary outcome of interest is the age of applicants. We calculate three measures of the age of the applicant pool using the data on the age we have recorded from the applications. First, we calculate the average age of applicants (excluding those who have not provided enough information to approximate their age). Second, we calculate the distribution of the age of applicants and identify the median and 75th percentile. Third, we calculate the share of the applicants for whom we have an age who are aged 40 and over. Finally, we calculate the share of applicants to a job who do not provide enough information to approximate their age.

Baseline estimation

To estimate the effect of ageist language on whether older applicants apply, we estimate equations of the form for each stereotype S :

$$(3) \quad A = \alpha + \beta S + \delta X + \varepsilon.$$

Our primary outcome variable A is the average age of applicants to the job posting. We also test whether ageist language shifts the age distribution more broadly. We examine the effect of ageist language on the median and 75th percentile of the age of applicants, and the share of applicants over 40 years old.

Because workers may strategically mask their age in the face of ageist job ads, we test the effect of ageist language on the share of applicants providing no age-identifying information. This analysis can be viewed as studying selection into reporting age. We find no evidence of this kind of selection, and hence do not need to be concerned with this potential source of bias. In our view, the absence of any

effect on reporting of age also likely implies that applicants do not manipulate age-related information on their applications in response to job-ad language. Presumably the consequences of lying about age, upon being interviewed and considered for a job, would be more adverse than simply omitting age-related information.

We estimate the effect of the stereotyped language in an ad (S) on the age of applicants conditional on controls (X) for the city and occupation the ads were posted in.³¹ Because the job ads vary by city and occupation and the job search behavior of applicants in a city and occupation may be correlated, we cluster the standard errors at the occupation and city levels.³²

Our null hypothesis is that the presence of ageist language on a job ad has no effect on the share of older workers that apply to the job ad (i.e., $\partial A_o / \partial S = 0$). The alternative hypothesis is that the presence of ageist language in a job ad will reduce the share of older workers that apply to the job ad (i.e., $\partial A_o / \partial S < 0$). We do not think a two-sided hypothesis test is the most meaningful in our context, but we report test results from both one-sided tests and two-sided tests (because two-sided tests are more conventional).³³

For the one-sided tests, our null hypothesis is that stereotyped language in a job ad does not deter older applicants from applying for a job. If this hypothesis is true, then β will be greater than or equal to zero. If we find that β is negative, this is evidence in favor of the alternative hypothesis that stereotyped language in a job ad deters older applicants from applying for a job (or reporting age).

We also test for heterogeneous effects by treatment type. As can be seen in Figure 5, there is a significant difference in how ageist individuals view the AARP treatments and the machine learning treatments. Therefore, we estimate whether there is a difference in responses when using stereotyped language determined by the machine learning similarity scores from Burn et al. (forthcoming) and when using the stereotyped language provided by AARP (*AARP*). To do this, we modify Equation (3) to include an interaction between the dummy variable for stereotyped language in an ad and a dummy

³¹ As a robustness check, we will use the full list of controls from the experiment which in addition to city and occupation, include month posted, day of the week posted.

³² Our PAP calls for multiple hypothesis testing; later drafts will incorporate these results.

³³ Our PAP generally focused on the one-sided hypothesis.

variable for using the AARP language.

$$(4) A = \alpha + \beta_1 S + \beta_2 (S \times AARP) + \delta X + \varepsilon.$$

Then, we test whether workers respond more strongly to specific stereotypes. To do this, we interact a dummy variable for having stereotyped language related to communication skills (C), physical ability (P), or technology (T) with our dummy variable for treatment. The identification of the effect of the specific stereotypes will come from the machine learning generated phrases and not the AARP phrases because the AARP treatment only ever includes all three stereotypes, while we separately enter each stereotype for the machine learning generated phrases. We begin by entering each stereotype individually before entering all three simultaneously, as shown below. To better understand the mechanisms behind our results, we also compare the effect of getting all three treatments to the effects of the three separate treatments, as well as to the effect of the AARP treatment.

$$(5) A = \alpha + \beta_1 S + \beta_2 (S \times P) + \beta_3 (S \times T) + \beta_4 (S \times All3) + \beta_5 (S \times AARP) + \delta X + \varepsilon. \quad ^{34}$$

Robustness Checks

We report additional analyses to test how robust our estimates are to various choices we made when specifying our models. We test the effect of varying our definition of older workers by raising the threshold to 50 and then 65 years old. We also change how we measure the treatment to take into account the intensity of the ageist sentiment. To do this, we replace our binary indicator of treatment (S) with the cosine similarity scores and the ageist sentiments measure derived from our MTURK survey.

Results

Descriptive evidence

We begin by presenting the empirical cumulative distribution function of applicant ages for the treatment and control ads, aggregating across all of the treatment ads (Figure 7). The raw data clearly show that the treatment ads that include ageist stereotypes attract fewer older applicants than that control ads. In Figure 8, we disaggregate the different treatment arms. There is a good deal of heterogeneity, but

³⁴ The PAP also calls for estimating heterogeneous effects by occupation, gender, other demographics, and cities. This analysis will be provided in a later draft.

one can discern that the CDF for the control ads is lower throughout most of the distribution, consistent with all of the treatment arms involving ageist stereotypes leading to fewer older applicants. One can also see that the treatment ads with three ageist phrases together, and the AARP treatments, are most pronounced in attracting fewer older applicants. Finally, among the ads with one ageist phrase, ads with an ageist phrase related to communication skills or physical ability attract somewhat fewer older applicants than ads with an ageist phrase related to technology.

In Figure 9, we examine the empirical density functions of applicant ages by treatment arm. Displaying the data this way makes clear that the applicants who appear to be disappearing from the applicant pool when faced with ageist phrases are those between roughly 40 and 60 years of age. For most of our treatments, we observe a significant gap in the age distributions first appearing between 40 and 45, which is very close to the age at which legal protections against age discrimination begin to apply. In many of the treatment arms, the distribution features more young workers, spread evenly across lower age ranges, although for ads with ageist language related to communication skills, the mass has shifted just below age 40.

Regression results

We present the regression results from our baseline estimates of the effect of each stereotype on the age distribution of applicants in Tables 3a-3c. Here we compare the treatment arm with the single machine-learning generated stereotype against the control arm. As a preliminary analysis, in the last column of each table, we find no evidence that treatment is correlated with the share of applicants providing information from which to judge their age. Therefore, we do not need to be concerned with selection into reporting age-identifying information on one's resume.³⁵

³⁵ Nonetheless, we take the information on a resume as given and assume individuals are telling the truth, so it is possible that individuals do not remove age-identifying information but instead only report information that would indicate they were younger than 40. Therefore, our results may be a combination of individuals choosing not to apply and those who choose to lie about their age on their resume when faced with ageist stereotypes. We do not regard this as very likely, given that applicants will almost certainly have to reveal accurate age information prior to taking the job.

The key result is that, for every way that we measure the age of applicants (average age, median age, 75th percentile, and the share over 40), for all three stereotypes, the sign of the estimate indicates that the ageist stereotype reduces applicants from older job seekers. For ads featuring an ageist phrase related to communication skills (Table 3a), the average age of applicants is 3.5 years younger than the control ads. The median and 75th percentile are lower by a similar amount. Overall, the treatment ads saw a share of applicants over the age of 40 that was lower by 8.7 percentage points. The estimate for average age is statistically significant at the 5% level in one-sided and two-sided tests, the estimate for median age is significant at a 5% and 10% level in one-sided and two-sided tests, the difference for the 75th percentile is statistically significant at the 10% level or less in one-sided tests, and the change in the share over 40 is not statistically significant.

For ads featuring an ageist phrase related to physical ability (Table 3b), although again, the point estimates always indicate that fewer older applicants applied, the evidence is weaker statistically. The estimates for median age and the share over 40 are statistically significant, but only at the 10% level in one-sided tests. The magnitudes for average age, the median, and the 75th percentile are about one-quarter to over one-third lower than for the communication stereotype ads. Overall, the treatment ads saw a share of applicants over age 40 that was lower by 9.5 percentage points.

For ads featuring an ageist phrase related to technology (Table 3c), we again find that, no matter how we measure age, the treatment ad attracted fewer older applicants. The average age of applicants is 3.2 years younger than the control ads. The median also fell, by somewhat more. Overall, the treatment ads saw a share of applicants over age 40 that was lower by 9.1 percentage points. The strongest statistical evidence is for the average and median age, significant at the 5% level in the one-sided tests, and the 10% level in the two-sided tests. Like for the physical ability stereotypes, the point estimates are a bit smaller than for the communications skills stereotypes, although less so.

In Table 4, we estimate a single model with all of the different treatment arms. Note, first, that every single estimate in this table is negative, implying that no matter how we measure age, and for every treatment and stereotype, the age-stereotyped job ads attract fewer older applicants. The estimates for the

job ads with a single stereotype are similar to those in Tables 3a-3c; we would not expect any difference except because the city-occupation dummy coefficients are estimated from different (more) observations, and the residual variance changes. In particular, the estimates are larger and more strongly statistically significant for the communications stereotypes, but there are some estimates significant at the 10% level in one-sided tests for the other two stereotypes as well.

The treatment arms, including all three machine learning phrases or all three AARP phrases, generate large and more strongly significant reductions in the ages of job applicants. For ads that feature all three machine learning treatment phrases, job applicants are 2.4 years younger than the applicants to the control ad. The median age for these ads was 2.5 years younger, and the 75th percentile was 4.7 years younger. The share of applicants over 40 was lower by 14.2 percentage points. All of these estimates are statistically significant at the 10% level or less, in both one-sided and two-sided tests.

For ads that feature all three AARP treatment phrases, the estimates are even larger and more strongly statistically significant. For example, the average age of applicants is 4.8 years younger than applicants to the control ads, and the share of applicants over age 40 is lower by 18.7 percentage points.

These evidence in Table 4 for the “All 3” treatment gives some indication of a “dosage” response, with a job ad that reflects more than one stereotype more strongly signaling to job applicants an employer is less likely to hire older workers. However, this is not consistent; it is apparent for the 75th percentile and the proportion over 40, but not for the average or median age. Although we might have anticipated that a single stereotyped phrase related to one skill or characteristic would largely go unnoticed by job applicants, the consistent evidence of negative effects of a single stereotyped phrase are quite striking.

Overall, then, we find significant evidence that ageist stereotypes reduce the likelihood that older workers apply, and the effects are substantial. For example, when all three machine-learning generated stereotypes are used, average age across cities is lowered by about 2.4 years (on a mean of 32.1), and the share of applicants over age 40 was lowered by 0.14 (on a mean of 0.17). Note that these estimates are based on unweighted city observations. Looking at the cdfs in Figure 8, a large effect (compare the blue vs. red lines) is apparent, but not this large. Of course the cdf is based on individual data, implying that

cities are not weighted equally but cities with more observations are weighted more heavily (and there are no controls).³⁶

Robustness analyses

Tables 5 and 6 present two robustness checks for the results in Table 4. First, we present results in Table 5 for different age thresholds for defining the share of applicants who are older – using 50 and 65, as well an alternative definition of over 40 treating applicants with no age information as being under age 40.³⁷ We also add fixed effects for city and occupation. We find very similar results for the alternative definition of over age 40 – compared to Table 4. The estimates for the age 50 threshold are quite similar, while those for age 65 are very small and statistically insignificant – although there were very few applicants above age 65, as Figure 7 shows. Table 6 replicates Table 4 but adds fixed effects for the month and the day of the week. The estimates are little changed from Table 4.

Comparing the potential effects of age-stereotyped job-ad language vs. direct age discrimination in hiring

As noted in the Introduction, age discrimination that deters older workers from applying for jobs has the same effect as direct age discrimination applied to job applicants; both reduce the employment of older workers. We can compare the estimated effects on the share of older job seekers hired from the discouragement of older applicants from age-stereotyped job-ad language, estimated in this paper, and the direct impact of age discrimination in hiring in Neumark et al. (2019a).

It is important to keep in mind, though, that the evidence from both experiments is specific to the experimental conditions, and may not generalize to actual incidence of age discrimination in hiring and age-related stereotypes in job ads in the broader labor market. Still, this evidence suggests that the two influences on hiring of older workers in the labor market could be of similar empirical importance.

³⁶ Our PAP did not specify that we would weight the estimates, but we will do so as a robustness check.

³⁷ This robustness check to an alternative definition comes from our PAP and treats all job applicants who did not provide any age identification as under the age of 40, rather than leaving their age imputation blank.

In our experiment in this paper, the share of applicants over 40 in the control group is 21.29%; the use of ageist language reduces the share of job applicants over age 40 by 5.54 percentage points.³⁸ In the correspondence study, the overall callback rate for the over 40 group (averaging across those near age 50 and near age 65) was 13.78%, compared with 18.69% for those under 40, a shortfall of 4.91 percentage points. Clearly these effects are of a similar magnitude. However, it is more instructive to calculate the implied effects on the share of older “hires” among all “hires.”³⁹

- If there were no age discrimination in hiring and no discouragement of applications from older job seekers, the percentage of older workers among all hires would be the same as this percentage in the control group, or 21.29%.
- Age discrimination reduces the percentage of older workers among hires to 16.63%.⁴⁰
- The discouragement of older applicants reduces the percentage of older workers among hires to 16.67%.⁴¹

Thus, the two effects are very similar.⁴² If these numbers roughly generalize to the actual labor market, the implication is that enforcement that focuses only on hiring shortfalls could conceivably miss half of age discrimination – subject also to the caveat discussed earlier that job-ad language that reflects age-related stereotypes may not solely reflect age discrimination.

³⁸ We use the individual-level data to be comparable across the two experiments.

³⁹ We equate “hiring” with “callback” for these calculations.

⁴⁰ This is computed by applying the actual hiring rates for older and younger applicants to the shares of applicants in each age group in the control group, to eliminate the discouragement effect:

$$= (0.2129 \times 0.1378) / [(0.2129 \times 0.1378) + (1 - 0.2129) \times 0.1869].$$

⁴¹ Since the hiring rate for older and younger applicants would be the same, this is computed by simply adjusting downward the proportion of older applicants and recomputing the share of older applicants:

$$= (0.2129 - 0.0554) / [(0.2129 - 0.0554) + (1 - 0.2129)].$$

⁴² Note that the summed effects exceed the overall effect by a little bit. That is because there is a negative interaction from the lower callback rate for older applicants being applied to a reduced number of older applicants. The percentage of older workers among all hires resulting from both effects is 12.86%, computed as $(0.1667 \times 0.1378) / [(0.1667 \times 0.1378) + (1 - 0.1667) \times 0.1869]$. This is a reduction of 8.43 percentage points, vs. the sum of the two effects adding to a 9.28 percentage point reduction, or $[(0.2129 - 0.1663) + (0.2129 - 0.1667)]$.

We also have to be cautious in this interpretation, because it is possible that age-stereotyped language in job ads may signal job characteristics which older applicants dislike or do not think themselves capable of fulfilling. While it is not possible to fully disentangle job applicants' thought processes, we are quite confident these interpretations do not fully explain our findings for three primary reasons. First, we observe in Burn et al. (2021) that both older and younger respondents perceive that the machine learning and AARP job requirements are biased against older workers. Moreover, the AARP requirements were explicitly billed as “phrases employers use to mask ageist discrimination” (Brenoff 2019). Second, the evidence from Burn et al. (forthcoming) indicates that employers who used physical and technologically-biased language sometimes discriminated against older men. If older applicants learn from experience that callback rates are lower for ads that include biased language, then we may expect they will be less likely to apply to such ads.

It is also the case that the older workers deterred by our treatment phrases are largely between the ages of 40 and 60. Previous research indicate that age discrimination begins in one's early 40s (Carlsson and Eriksson 2019), which suggests that discrimination begins to appear before an age group becomes obviously less qualified to fulfill the job requirements. The AARP phrases, for example, do not convey any specific or objective skill requirements. In addition, the computer programs required have been available for over 30 years, so many of our deterred applicants have been familiar with these programs for much of their working lives.

Nonetheless, the aforementioned possibilities imply that the use of ageist language, per se, in job ads does not necessarily imply discrimination, which parallels what we said earlier about the relationship between such language and RFOAs. Job ads which feature bona fide job requirements related to ageist stereotypes may be less attractive to older workers. Thus, were job-ad language to be added to the tools of anti-discrimination enforcement, it should be used only as a potential flag for discriminatory behavior – prompting further investigation, including whether employers who use such language are still less likely to hire older job applicants.

Discussion and Conclusion

In this paper, we conducted the first field experiment that examines how older job seekers respond to the presence of ageist language in job ads. We manipulated the language of online job ads to feature control phrases that had low relatedness to ageist stereotypes or treatment phrases that were highly related to ageist stereotypes or flagged as such by AARP. The treatment and control phrases were validated using two methods. The first shows that the machine learning phrases are only related to the specific stereotypes and are not related to any of the other stereotypes about older workers. The second method showed the phrases to individuals on MTURK and asked them to rate how ageist they perceived them to be. We found that the control ads were viewed as significantly less ageist than the treatment phrases, and the perceived ageism increased with increases in the relatedness (semantic similarity) to ageist stereotypes.

We posted job ads in three occupations in 14 cities, with six job postings in each city-occupation cell. The results indicate that older workers, when faced with ageist language in job ads, are less likely to apply for jobs, with measures like average age and the share of applicants over the age of 40 (as well as other measures) falling. The results indicate that there may be additive or “dose-response” effects of ageist language in job ads. As the number of ageist phrases increases in the job ads, the effects grow stronger. This suggests that one ageist phrase may shift behavior, but the effect is sometimes smaller and less likely to be statistically significant. However, job ads with multiple ageist phrases lead to strong declines in applications from older job seekers. For example, when job ads included three machine-learning generated phrases with ageist stereotypes related to communication skills, physical ability, and technology skills, the share of applicants over 40 declined by 14.2 percentage points. And the decline is particularly sharp in the upper parts of the age distribution, with the 75th percentile falling by 4.7 years.

Our evidence has significant policy implications regarding age discrimination. We show that ageist stereotypes in job ads discourages older applicants from applying for jobs. The effects of this discouragement of applications from older job seekers can have as deleterious an impact on the hiring of older workers as can direct age discrimination in hiring. As a result, these results suggest the need for

further guidance from the EEOC to employers to avoid age-stereotyped job-ad language that deters older workers from applying for jobs. Using language that explicitly deters older workers from applying is already illegal under the ADEA, but the subtler usage of ageist language that we study suggests that job-ad language that would not be flagged as explicitly illegal can still have pernicious effects on older workers in the labor market, and possibly facilitate age discrimination. Moreover, the EEOC might consider flagging for potential investigation firms that use age-stereotyped language in their job ads, recognizing that, for these firms, discrimination may be occurring even in the absence of shortfalls between the share of older applicants hired and the share of older workers who apply for jobs. Third, in assessing evidence of age discrimination in hiring, the courts may need to put more weight on evidence aside from differences between the shares of older workers among hires and among job applicants, as the share of older workers among job applicants may itself reflect the discrimination that occurs through job-ad language. Finally, of course, these same considerations may apply to discrimination against other protected groups, but such an assessment awaits research on these groups using methods similar to ours.

References

- Adams, Scott. 2004. "Age Discrimination Legislation and the Employment of Older Workers." *Labour Economics* 11: 219-41.
- 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, S. and Villena-Roldán, B. 2019. "Do High-Wage Jobs Attract More Applicants? Directed Search Evidence from the Online Labor Market." *Journal of Labor Economics* 37: 715-746.
- Belot, Michele, Philipp Kircher, and Paul Muller. 2018. "How Wage Announcements Affect Job Search – A Field Experiment." IZA Discussion Paper No. 11814.
- Bem, S.L., and Bem, D.J., 1973. Does Sex-biased Job Advertising "Aid and Abet" Sex Discrimination? *Journal of Applied Social Psychology* 3: 6-18.
- 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: 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: 25-46.
- Burn, Ian, Patrick Button, Luis Felipe Munguia Corella, David Neumark. "Does Ageist Language in Job Ads Predict Age Discrimination in Hiring?" Forthcoming in *Journal of Labor Economics*.
- Burn, Ian, Daniel Firoozi, Daniel Ladd, and David Neumark. 2021. "Machine Learning and Perceived Age Stereotypes in Job Ads: Evidence from an Experiment." NBER Working Paper No. 28328.
- Brenoff, Ann. 2019. "5 Ageist Phrases to be Aware of." AARP: Washington, DC.
- Cahill, Kevin E, Michael D. Giandrea, and Joseph F. Quinn. 2006. "Retirement Patterns from Career Employment." *The Gerontologist* 46: 514-23.
- Carlsson, Magnus, and Stefan Eriksson. 2019. "Age Discrimination in Hiring Decisions: Evidence from a Field Experiment in the Labor Market." *Labour Economics* 59: 173-183.
- Combs, Deborah H. 1982. "Striking a Balance between the Interests of Public Safety and the Rights of Older Workers: The Age BFOQ Defense." *Washington and Lee Law Review*, 39, 1371-95.
- 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.
- European Commission. 2000. *Council Directive 2000/78/EC of 27 November 2000 Establishing a General Framework for Equal Treatment in Employment and Occupation*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32000L0078>.
- Farber, Henry S., Silverman, Daniel., and Till 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: 168-201.
- Farber, Henry S., Herbst, Chris M., Silverman, Daniel, and Till von Wachter, T. 2019. "Whom Do Employers Want? The Role of Recent Employment and Unemployment Status and Age." *Journal of Labor Economics* 37(2): 323-49.
- Federal Register. (n.d.). *Disparate Impact and Reasonable Factors Other Than Age Under the Age*

- Discrimination in Employment Act*. <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).
- Flory, Jeffrey A., Andreas Leibbrandt, and John A. List. 2015. "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions." *Review of Economic Studies* 82: 122-55.
- Flory, Jeffrey A., Andreas Leibbrandt, Christina Rott, and Olga Stoddard. 2021. "Increasing Workplace Diversity: Evidence from a Recruiting Experiment at a Fortune 500 Company." *Journal of Human Resources* 57: 73-92.
- Gaucher, Danielle, Justin Friesen, and Aaron C. Kay. 2011. "Evidence that Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality." *Journal of Personality and Social Psychology* 101: 109-28.
- 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: 468-92.
- Hanson, Andrew, Zachary Hawley, and Aryn Taylor. 2011. "Subtle Discrimination in the Rental Housing Market: Evidence from email Correspondence with Landlords." *Journal of Housing Economics* 20: 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.
- He, Haoran, David Neumark, and Qian Weng. 2021. "Do Workers Value Flexible Jobs? A Field Experiment." *Journal of Labor Economics* 39: 709-38.
- Hellerstein, Judith K., David Neumark, and Kenneth Troske. 1999. "Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations." *Journal of Labor Economics* 17: 409-46.
- Hellester, Miguel D., Peter Kuhn, and Kailing Shen. 2020. "The Age Twist in Employers' Gender Requests." *Journal of Human Resources* 55: 482-69.
- Ibanez, Marcela, and Gerhard Riener. 2018. "Sorting through Affirmative Action: Three Field Experiments." *Journal of Labor Economics* 36:437-78.
- Johnson, Richard W., Janette Kawachi, and Eric K. Lewis. 2009. "Older Workers on the Move: Recareering in Later Life." AARP Public Policy Institute. Washington, DC: AARP.
- Krause, Annabelle, Ulf Rinne, and Klaus Zimmerman. 2012. "Anonymous Job Applications of Fresh Ph.D. Economists." *Economics Letters* 117: 441-4.
- Kuhn, Peter, and Kailing Shen. 2013. "Gender Discrimination in Job Ads: Evidence from China." *Quarterly Journal of Economics* 128: 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. 2010. "International Comparisons of Age Discrimination Laws." *Research on Aging* 32: 679-97.
- Lahey, Joanna. 2008. "Age, Women, and Hiring: An Experimental Study." *Journal of Human Resources* 43: 30-56.
- Leibbrandt, Andreas, and John List. 2018. "Do Equal Employment Opportunity Statements Backfire? Evidence from a Natural Field Experiment on Job-Entry Decisions." NBER Working Paper No. 25035.

- Maestas, Nicole. 2010. "Back to Work: Expectations and Realizations of Work after Retirement." *Journal of Human Resources* 45: 718-48.
- Marinescu, Ioana Elena, and Ronald Wolthoff. 2020. "Opening the Black Box of the Matching Function: The Power of Words." *Journal of Labor Economics* 38: 535-568.
- 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, Ian Burn, and Patrick Button. 2016. "Experimental Age Discrimination Evidence and the Heckman Critique." *American Economic Review Papers and Proceedings* 106: 303-8.
- Neumark, David, Ian Burn, and Patrick Button. 2019a. "Is It Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment," *Journal of Political Economy* 127: 922-70.
- Neumark, David, Ian Burn, Patrick Button, and Nanneh Chehras. 2019b. "Do State Laws Protecting Older Workers from Discrimination Reduce Age Discrimination in Hiring? Evidence from a Field Experiment." *Journal of Law and Economics* 62: 373-402.
- Riach, Peter A., and Judith Rich. 2006. "An Experimental Investigation of Age Discrimination in the French Labour Market." IZA Discussion Paper No. 2522.
- Riach, Peter A., and Judith Rich. 2010. "An Experimental Investigation of Age Discrimination in the English Labor Market." *Annals of Economics and Statistics* 99/100: 169-85.
- 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: 423-43.
- Terrell, Kenneth. 2019. "Age Bias That's Barred by Law Appears in Thousands of Job Listings." AARP: Washington, DC.
- Tilcsik, András. 2011. "Pride and Prejudice: Employment Discrimination against Openly Gay Men in the United States." *American Journal of Sociology* 117: 586-626.
- van Borm, Hannah, Ian Burn, and Stijn Baert. 2021. "What Does a Job Candidate's Age Signal to Employers?" *Labour Economics* 17: 102003.
- Wax, S. L. 1948. "A Survey of Restrictive Advertising and Discrimination by Summer Resorts in the Province of Ontario." *Canadian Jewish Congress*.

Figure 1: Map of Cities in Experiment



Note: This map shows cities in the experiment. The relative size of the symbol corresponds to the total number of applicants in each city, ranging from 25 in Salt Lake City to 526 in New York City. Note that, as explained in the text, different cities have had different numbers of ads posted thus far (and some differences will likely remain at the completion of data collection).

Figure 2: Examples of Job Ad Templates

Administrative Assistants Template 1 (Admin Assistant)

Psychiatric office is in need of a full or part time Administrative Assistant to assist in front/back office general clerical duties. This individual will work on a several tasks and stay on course at all times. The Administrative Assistant we hire will be trained in various duties that cover the entire office.

This individual MUST possess the following:

- Exceptional customer service background to greet and register patients, answer phones, schedule appointments.
 - Can multitask.
 - High School diploma or GED.
 - Professional attitude.
 - *Communication Skill Requirement***.
 - *Technology Requirement***
 - *Physical Requirement***
 - Available for flexible hours.
- (Schedule hours and days will alternate every other week)

Please email us a CV or resume and put “full-time” or “part-time” in the subject line.

Retail Sales Associate Template 1 (Retail Sales Job)

Our women’s clothing store in ***City*** is looking for a sales associate to help us out weekday afternoons. We are pretty busy store and you must ***Physical Requirement***. We are looking for someone with open to working in retail, who ***Communication Skill Requirement***. We need you to ***Technology Requirement***. So if this sounds like you, send us your resume and your earliest possible starting date and we will be in touch.

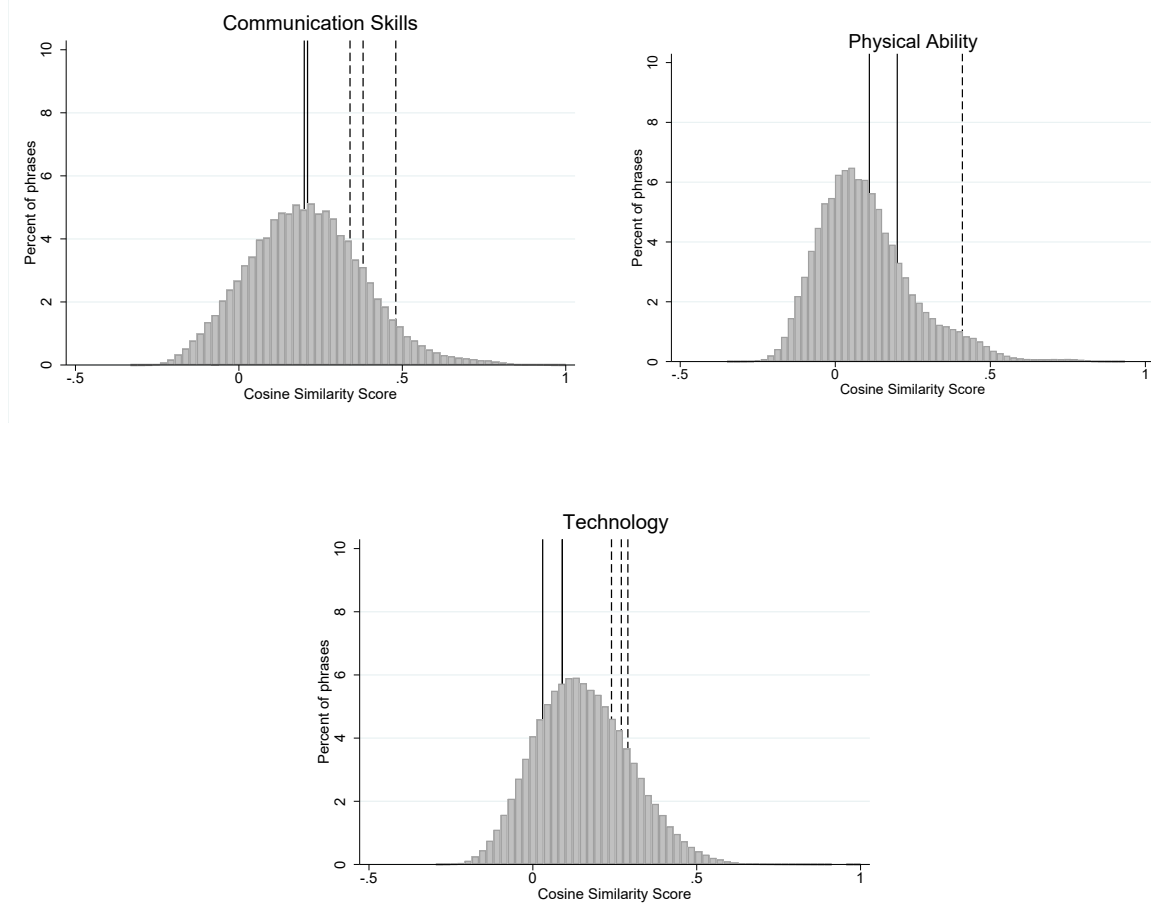
Security Guard Template 1 (HIRING UNARMED SECURITY GUARDS)

We currently have a position for a full-time or part-time security officer available. Training and uniforms will provided. We offer flexible working hours and have shifts any day of the week. Our pay scale is competitive. Email your resume and potential work hours to apply.

Requirements

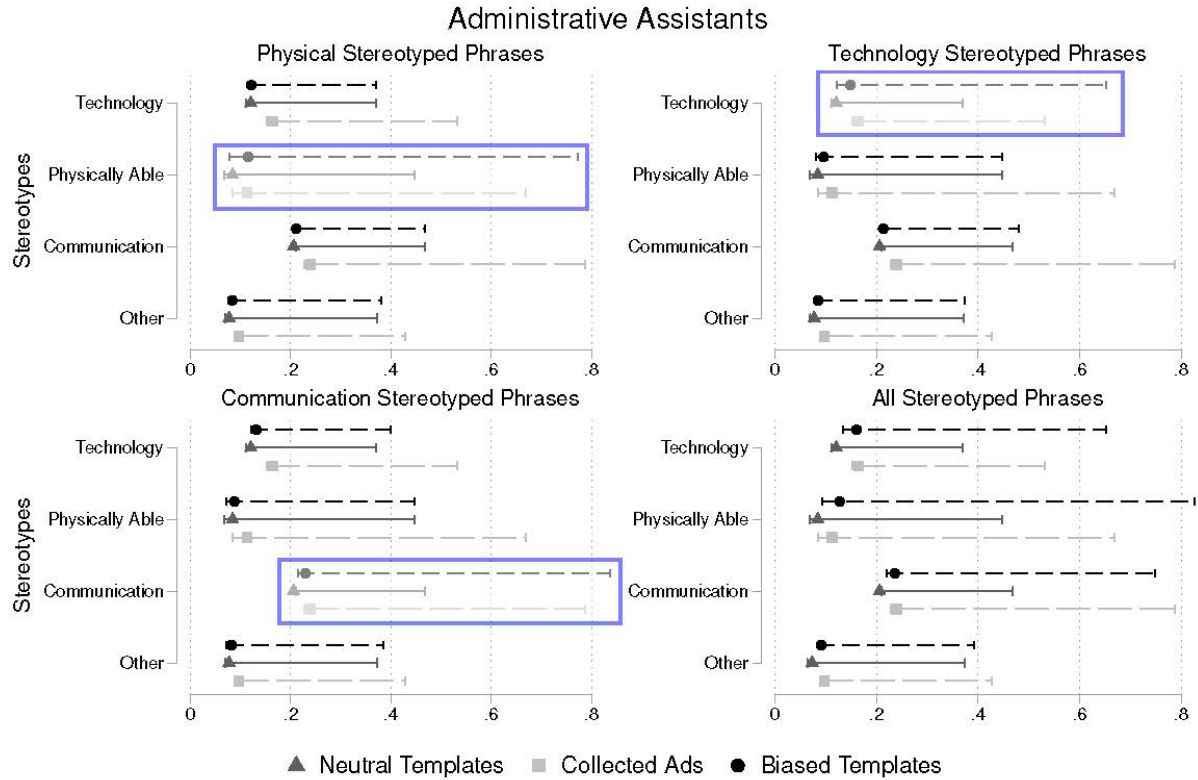
- Professional appearance & attitude
- Detail oriented
- *Communication Skill Requirement***
- *Physical Requirement***
- *Technology Requirement***
- At least 18 years of age
- Access to transportation

Figure 3: Locations of Treatment and Control Phrases in the Cosine Similarity Score Distribution of Job Ad Phrases



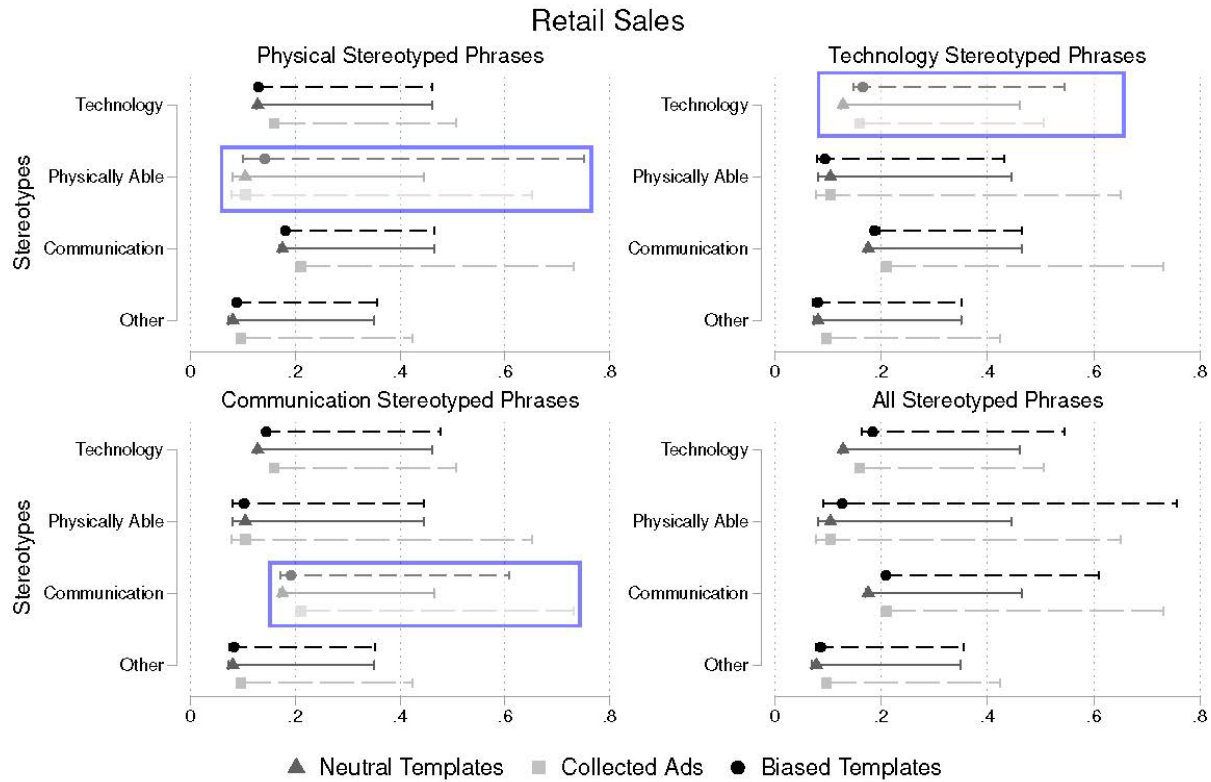
Note: Figure reports the distribution of cosine similarity scores for all trigrams from the job ads with the indicated stereotypes. The higher the cosine similarity score, the more related the trigram is to the stereotype, with a minimum of -1 and a maximum of 1 . Solid lines indicate the location of a control sentence in the cosine similarity score distribution. Dashed lines indicate the location of a treatment phrase (for the Machine Learning Treatments shown in Table 2).

Figure 4a: Cosine Similarity Score of Administrative Assistant Templates



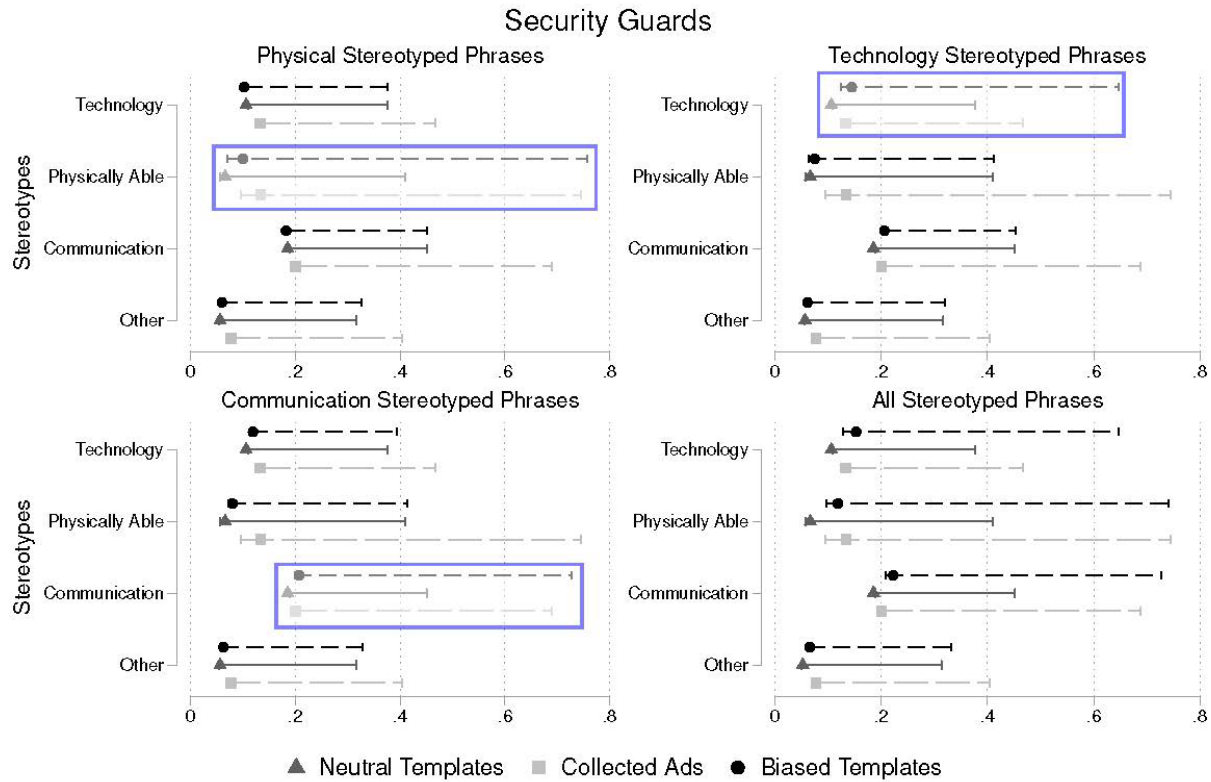
Note: Graphs display median to 99th percentile range of trigram semantic similarity scores for each stereotype for Administrative Assistant ads. The average trigram semantic similarity score for each stereotype is represented by the respective shape for each template. The category “Other” shows the averages for the remaining stereotypes listed in Table 1. Control (“neutral”) templates contain trigrams from the created ad templates with only non-stereotyped phrases included. Collected ads comprise trigrams from all Administrative Assistant job ads. Treatment templates contain trigrams from the created ad templates with the respective stereotyped phrase or phrases included.

Figure 4b: Cosine Similarity Score of Retail Sales Templates



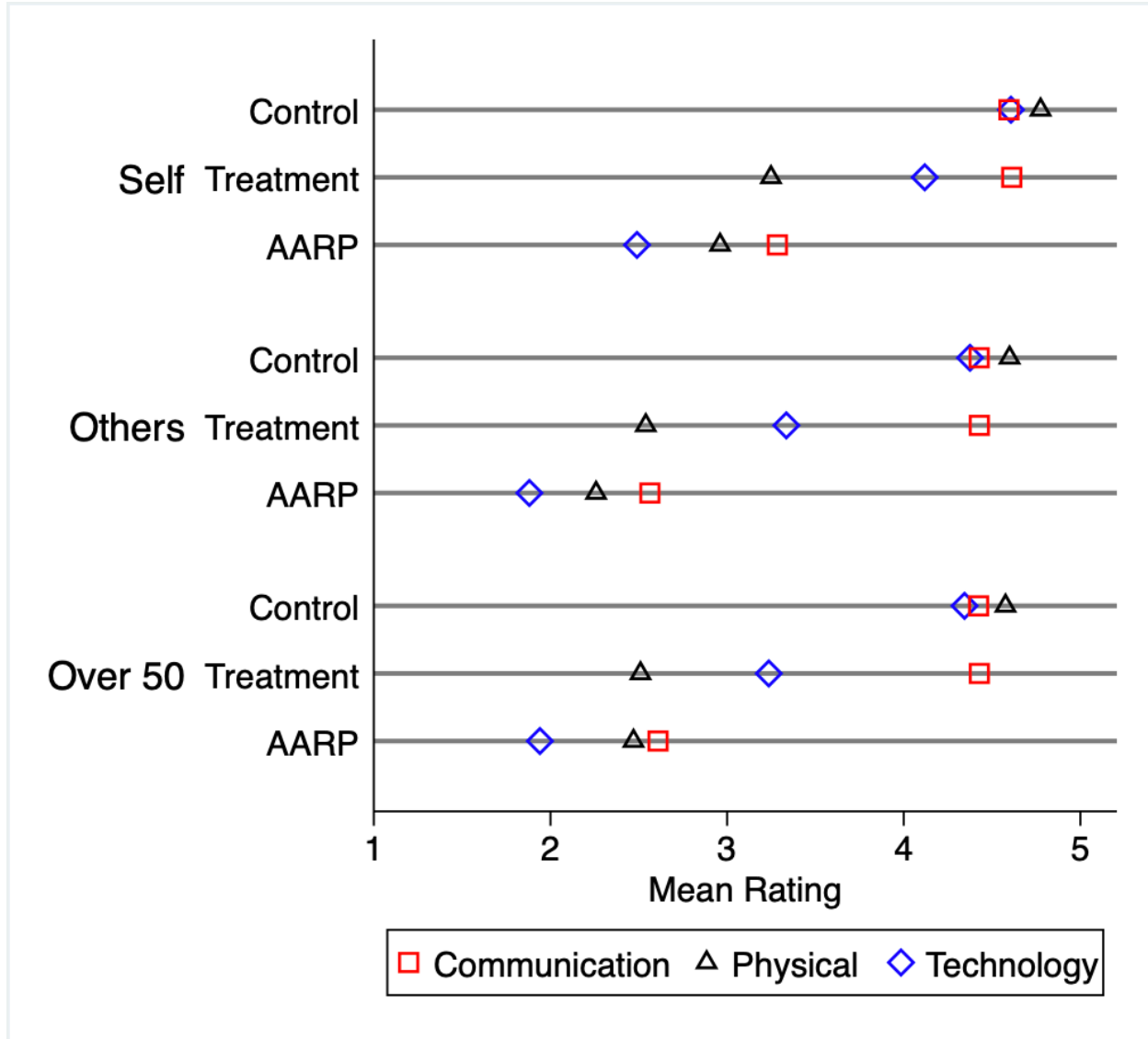
Note: Graphs display median to 99th percentile range of trigram semantic similarity scores for each stereotype for Retail Sales ads. The average trigram semantic similarity score for each stereotype is represented by the respective shape for each template. The category “Other” shows the averages for the remaining stereotypes listed in Table 1. Control (“neutral”) templates contain trigrams from the created ad templates with only non-stereotyped phrases included. Collected ads comprise trigrams from all Retail Sales job ads. Treatment templates contain trigrams from the created ad templates with the respective stereotyped phrase or phrases included.

Figure 4c: Cosine Similarity Score of Security Guard Templates



Note: Graphs display median to 99th percentile range of trigram semantic similarity scores for each stereotype for Security Guard ads. The average trigram semantic similarity score for each stereotype is represented by the respective shape for each template. The category “Other” shows the averages for the remaining stereotypes listed in Table 1. Control (“neutral”) templates contain trigrams from the created ad templates with only non-stereotyped phrases included. Collected ads comprise trigrams from all Security Guard job ads. Treatment templates contain trigrams from the created ad templates with the respective stereotyped phrase or phrases included.

Figure 5: Survey Results

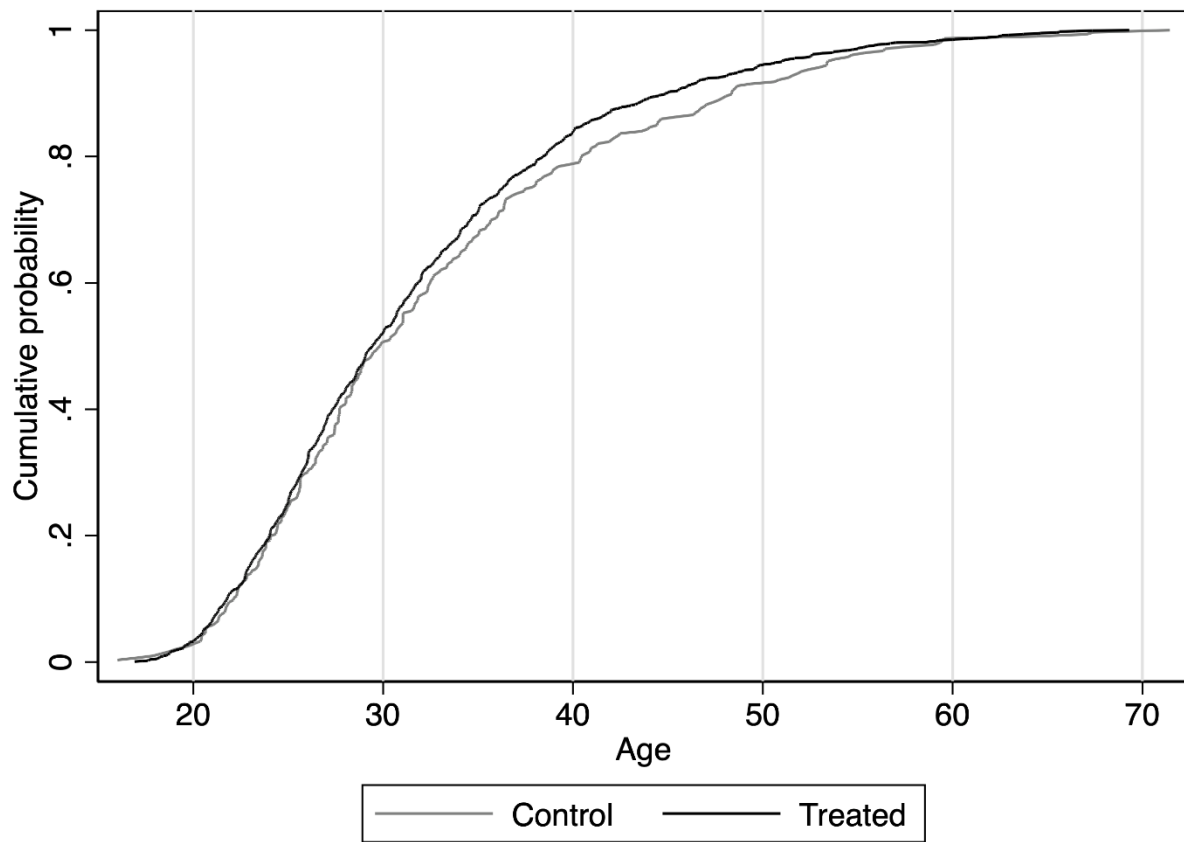


Note: These numerical ratings reflect the degree to which survey respondents rated phrases as age-biased or not age-biased, with lower numbers indicating a greater bias against older workers. Likert Scale ratings were translated to a numerical value such that “Strongly Agree” mapped to 1, “Somewhat Agree” mapped to 2, “Neither agree nor disagree” mapped to 3, “Somewhat Disagree” mapped to 4, and “Strongly Disagree” mapped to 5. The three categories “Self,” “Others,” and “Over 50” refer to which group’s opinions the MTURK respondents were asked to provide or predict in a given survey block. The average bias rating was collapsed on the treatment status of phrases (control, treatment, and AARP) as well as the category of the stereotype (communication, physical, or technology). Hence, each point in the figure above reflects the average bias rating MTURK respondents gave to a given treatment status for a specific stereotype from the perspective of a given group of people. For example, the triangle in the first row of the above figure indicates that when respondents were asked for their self-assessment of whether or not the physical stereotype control phrases were age-biased, they, on average, stated that they strongly disagreed.

Figure 6: Posting Job Ads Was Not Easy!

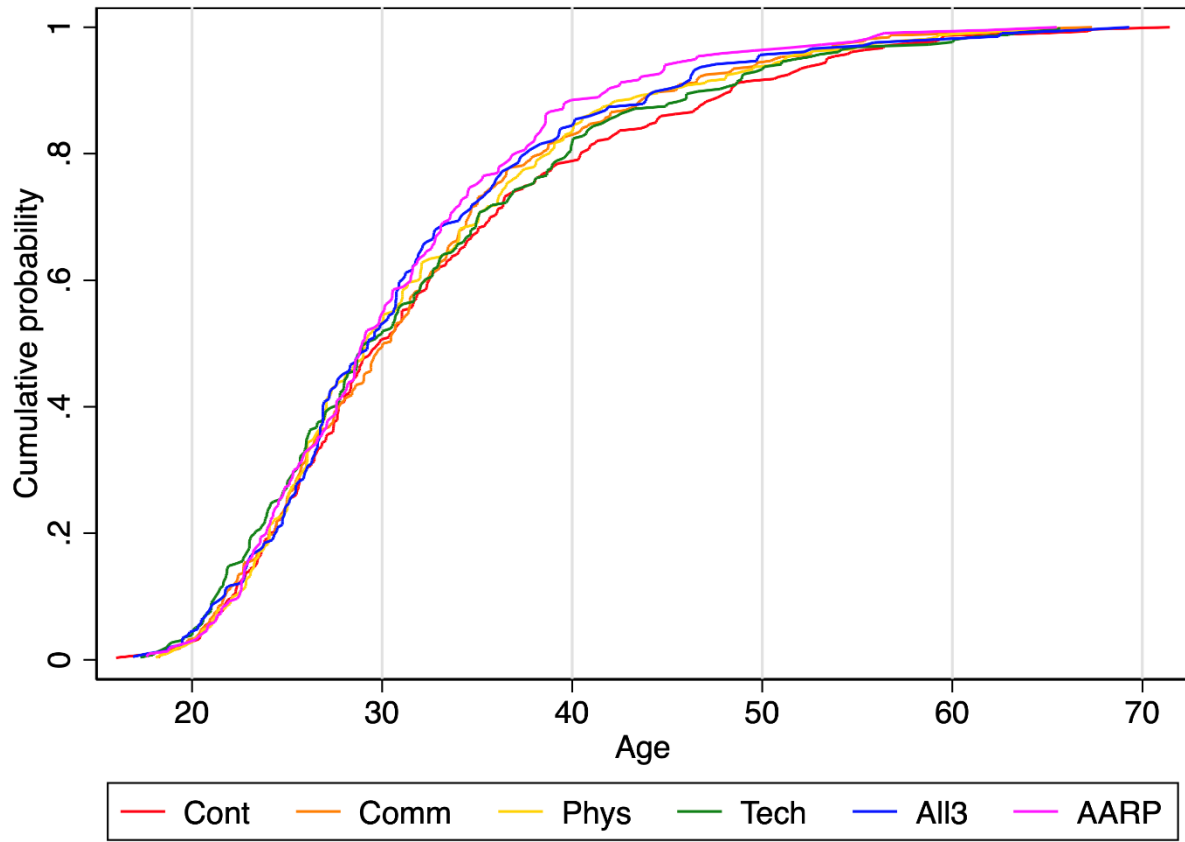


Figure 7: Empirical Cumulative Density Functions, Any Treatment



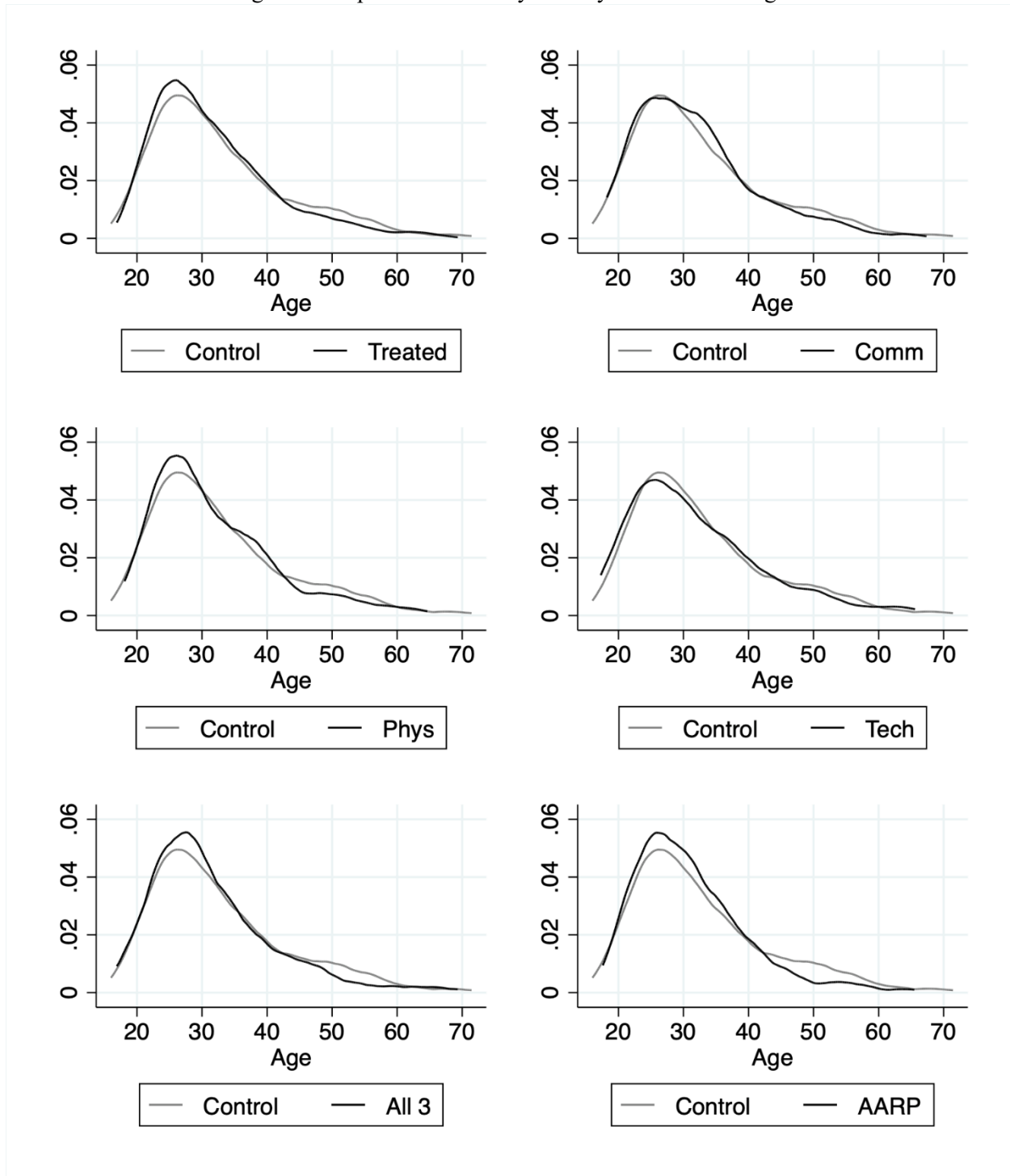
Note: “Treated” refers to any treatment (individual stereotypes, all 3, or AARP).

Figure 8: Empirical Cumulative Density Functions by Ad Type



Note: “Cont” refers to controls; “Comm” to communications skills stereotypes; “Phys” to physical ability stereotypes; “Tech” to technology stereotypes; “All3” to the ads with all three stereotypes reflected in the text; and “AARP” to the ads with AARP ageist language/stereotypes.

Figure 9: Empirical Probability Density Functions for Age



Note: In the upper-left panel, “Treated” refers to any treatment (individual stereotypes, all 3, or AARP). The other labels are explained in the notes to Figure 8.

Table 1: Age Stereotypes from Industrial Psychology Literature

Health	Personality	Skills
Less Attractive	Less Adaptable	Lower Ability to Learn
Hard of Hearing	Careful	Better Communication Skills
Worse Memory	Less Creative	Worse Communication Skills
Less Physically Able	Dependable	More Experienced
	Negative Personality	More Productive
	Warm Personality	Less Productive
		Worse with Technology

Note: See Burn et al. (forthcoming).

Table 2: Control and Treatment Phrases by Occupation

Occupation	Stereotype	Control	Machine Learning Treatment	AARP Treatment
(1)	(2)	(3)	(4)	(5)
Administrative Assistants	Communication skills	You must be good at working without supervision	You must have good communication and teamwork on tasks	You must be up-to-date with current industry jargon and communicate with a dynamic workforce
Administrative Assistants	Physical ability	You must enter bills and keep track of invoices	You must be able to lift 40 pounds	You must be a fit and energetic person
Administrative Assistants	Technical skills	You must produce and distribute documents such as correspondence memos, faxes and forms	You must use accounting software systems like Netsuite, Freshbook, and QuickBooks	You must be a digital native and have a background in social media
Retail sales	Communication skills	You must be good at working without supervision	You must have good communication with customers and staff	You must be up-to-date with current industry jargon and communicate with a dynamic workforce
Retail sales	Physical ability	You must enter bills and keep track of invoices	You must be able to lift 40 pounds	You must be a fit and energetic person
Retail sales	Technical skills	You must help to clean and organize the store	You must use software such as Microsoft Office/Excel or Google Sheets	You must be a digital native and have a background in social media
Security guard	Communication skills	You must follow instruction from supervisors	You must maintain communication about tasks with supervisors	You must be up-to-date with current industry jargon and communicate with a dynamic workforce
Security guard	Physical ability	You need to carry a flashlight	You must be able to lift 50 pounds	You must be a fit and energetic person
Security guard	Technical skills	You must write patrol records in journal notebook	You must type patrol entries into a journal application on a computer system	You must be a digital native and have a background in social media

Note: See text for a description of how each sentence was created.

Table 3a: Estimated Effects on Age Composition of Applicants, Communications Stereotypes, All Cities

	Average Age	Median Age	75 th Percentile	Over 40	No Age Information
<i>Communication</i>	-3.540 ^{***†}	-3.656 ^{***†}	-3.789 [†]	-0.087	-0.009
	(1.711)	(2.023)	(2.766)	(0.069)	(0.049)
N	60	60	60	60	62

Note: This table covers all cities. The regression includes the single machine learning stereotype arm and the control arm. All specifications include fixed effects for both city and occupation. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level.

***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or † indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test.

Table 3b: Estimated Effects on Age Composition of Applicants, Physical Stereotypes, All Cities

	Average Age	Median Age	75 th Percentile	Over 40	No Age Information
<i>Physical</i>	-1.907	-2.391 [†]	-3.019	-0.095 [†]	0.005
	(2.005)	(1.998)	(3.172)	(0.070)	(0.070)
N	59	59	59	59	61

Note: This table covers all cities. The regression includes the single machine learning stereotype arm and the control arm. All specifications include fixed effects for both city and occupation. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level.

***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or † indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test.

Table 3c: Estimated Effects on Age Composition of Applicants, Technology Stereotypes, All Cities

	Average Age	Median Age	75 th Percentile	Over 40	No Age Information
<i>Technology</i>	-3.155 ^{***†}	-3.486 ^{***†}	-2.574	-0.091 [†]	0.002
	(1.742)	(1.856)	(2.808)	(0.061)	(0.055)
N	58	58	58	58	60

Note: This table covers all cities. The regression includes the single machine learning stereotype arm and the control arm. All specifications include fixed effects for both city and occupation. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level.

***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or † indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test.

Table 4: Estimated Effects on Age Composition of Applicants, All Treatment Arms, All Cities

	Average Age	Median Age	75 th Percentile	Over 40	No Age Information
<i>Communication</i>	-3.104***†	-3.256**†	-3.265†	-0.088†	-0.002
	(1.533)	(1.759)	(2.469)	(0.060)	(0.047)
<i>Physical</i>	-1.870	-2.327†	-2.522	-0.091†	0.013
	(1.730)	(1.691)	(2.827)	(0.061)	(0.062)
<i>Technology</i>	-2.273†	-2.462†	-1.169	-0.070†	0.022
	(1.510)	(1.576)	(2.431)	(0.054)	(0.049)
<i>All 3</i>	-2.435***†	-2.470***†	-4.666***†	-0.142***††	-0.014
	(1.402)	(1.408)	(2.313)	(0.050)	(0.051)
<i>AARP</i>	-4.830***††	-4.320***††	-6.567***††	-0.187***††	0.021
	(1.658)	(1.62)	(2.518)	(0.050)	(0.076)
N	167	167	167	167	173

Note: This table covers all cities. All specifications include fixed effects for both city and occupation. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level. ***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or † indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test.

Table 5: Estimated Effects on Age Composition of Applicants, Different Age Cutoffs, All Treatment Arms, All Cities

	Over 40	Over 50	Over 65
<i>Communication</i>	-.077 ^{*††}	-0.082 [†]	0.0003
	(0.043)	(0.050)	(0.012)
<i>Physical</i>	-0.075 [†]	-0.043	-0.008
	(0.046)	(0.054)	(0.006)
<i>Technology</i>	-0.060 [†]	-0.048	-0.0002
	(0.0417)	(0.049)	(0.010)
<i>All 3</i>	-0.112 ^{***†††}	-0.096 ^{***††}	-0.004
	(0.039)	(0.042)	(0.007)
<i>AARP</i>	-0.154 ^{***†††}	-0.107 ^{***††}	-0.004
	(0.037)	(0.045)	(0.007)
N	173	167	167

Note: This table covers all cities. All specifications include fixed-effects for both city and occupation. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level. ***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or † indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test. Over 40 treats applicants with no age information available as being under the age of 40.

Table 6: Estimated Effects on Age Composition of Applicants, with Month and Day-of-Week Fixed Effects, All Treatment Arms, All Cities

	Average Age	Median Age	75 th Percentile	Over 40	No Age Information
<i>Communication</i>	-3.000 ^{*††}	-2.811 [†]	-2.818	-0.074	-0.017
	(1.559)	(1.758)	(2.659)	(0.058)	(0.057)
<i>Physical</i>	-1.150	-1.233	-1.791	-0.079	0.015
	(1.836)	(1.772)	(2.913)	(0.063)	(0.062)
<i>Technology</i>	-2.444 [†]	-1.966	-1.112	-0.073	0.026
	(1.672)	(1.683)	(2.830)	(0.064)	(0.051)
<i>All 3</i>	-1.647	-1.336	-3.424 [†]	-0.108 ^{*††}	-0.005
	(1.520)	(1.565)	(2.569)	(0.055)	(0.061)
<i>AARP</i>	-3.903 ^{***†††}	-3.070 ^{*††}	-5.679 ^{*††}	-0.174 ^{***†††}	0.020
	(1.812)	(1.739)	(2.822)	(0.060)	(0.071)
N	167	167	167	167	173

Note: This table covers all cities. All specifications include fixed effects for city, occupation, month of posting, and day-of-week of posting. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level. ***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or † indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test.