Stereotyped Language in Job Ads and Decisions by Older Workers to Apply for a Job: Extended Abstract/Partial Draft

Ian Burn
University of Liverpool

Daniel Firoozi
University of California-Irvine

Daniel Ladd
University of California-Irvine

David Neumark
University of California-Irvine

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 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 are 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

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

correspondence study in which the job searchers are artificial and we study the responses of real employers.[2]

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 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.

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 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). Thus, our research asking whether older workers are deterred from applying to job ads that feature ageist language can provide information and tools to parties that enforce age discrimination laws by helping to identify job-ad language that may predict intent to discriminate on the basis of age in hiring and adverse impact on older job applicants.

[2] We have 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).

[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.)

Previous Literature

Very few studies 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. And He et al. (forthcoming) study how job flexibility conditions influence job application behavior. 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.

Two studies, to date, connect the text of job ads to measured discriminatory behavior of employers.[4] 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.[5] And, in the most systematic approach, Burn et al. (2020) 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 whether the ageist language in job ads is perceived as ageist by potential job applicants. Obviously, the use of such language in job ads is much more troublesome if it is perceived as ageist and thus discourages older workers from applying for jobs. If this happens, it should

[4] 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.

[5] 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.

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.[6]

What is known about how job applicants read job ads focuses exclusively on gender bias. Gaucher et al. (2011) found that job ads for male-dominated occupations used masculine wording (words associated with male stereotypes, such as "leader," "competitive," or "dominant") more frequently than advertisements for female-dominated occupations, and women find job advertisements less appealing when they contained more masculine than feminine wording (Bem and Bem, 1973; Gaucher et al., 2011).

Methods

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 Cites 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 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.

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

[6] 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).

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.[7,8]

Selecting the Stereotypes

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., 2020). 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., 2020) and evidence that the stereotype captures a skill that employers view older workers as less likely to possess (from van Borm et al., 2020).[9]

Third, older workers should be aware that employers held that stereotype. As evidence, we drew on various reports put out by the American Association of Retired Persons (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. 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,

[7] 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.

[8] 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.

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

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 ads are written to have flexible hours, competitive pay, and the availability of part-time and full-time positions. For half of the templates, we include the requirement that applicants must have a high school diploma (randomized by template).

The treatment and control ads differ in the job requirements (denoted in bold with asterisks in each template), 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. (2020), 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 1.

Our second treatment conveys bias by using ageist language identified by the 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 1, 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.

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 generate ads that 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 1a-1c 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. (2020) 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 display a few key results. First, the 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. 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.

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. Appendix A details the survey and results. These results, like those in Figures 1a-1c, imply that our phrases capture real ageist sentiments and will be perceived as such by job applicants, so that 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 plan to run the experiment for 54 weeks, with the schedule of posting pre-registered (see Table 2 for an example of the schedule for some cities). The cities have been split such that each

week ads are posted in five of the cities, while the other 10 are rested. To maximize the number of potential applicants, one ad is posted each weekday (Monday through Friday). 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.

A number of our ads were flagged by the job board as spam and taken down before they had been active for a week. If the ad was taken down before we receive responses, we 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.

Collecting Responses

Applicants sent us their resumes as an attachment when they replied to our job ad. To avoid spam responses and to make sure that the applicants have read the job ad, we have included a manipulation check. Each ad contains a cue that applicants are 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." (See Table 3 for a list of these questions.) To reduce the cost of applying for our fake job, we informed applicants that they were not selected for an interview and that we have decided to go with another candidate within one week of their application via an email (and we hope to send these emails within 72 hours or less). 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.

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.[10] Assuming that an individual was approximately 18 years old when they graduated high school, age is calculated as

$$Age = 2021 - YearofHSGraduation + 18.$$

If the applicant does not provide a year of high school graduation but includes a sufficiently long work history (i.e., 25 years or more of experience), we calculate age based on the total number of years of experience listed on the resume. Age is calculated as

$$Age = 2021 - YearofFirstJob + 16.$$[11]

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 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

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. The alternative hypothesis is that the presence of ageist language in a job ad will change the share of older workers that apply to the job ad. We are also interested in hypotheses related to whether it is the number of ageist words, the severity of the ageist sentiment, or particular ageist stereotypes that affect the application choices of older applicants.

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

[10] 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.

[11] This is probably best interpreted as "minimal possible age" assuming one did not start working before age 16.

using the data on 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.[12]

Baseline estimation

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

$$A = \alpha + \beta S + \delta X + \varepsilon. \quad [1]$$

Our outcome variable ***A*** will first be the average age of applicants to the job posting, then the median and 75th percentile of the age of applicants, the share of applicants over 40 years old, and finally, the share of applicants providing no age identifying information. The latter analysis can be viewed as studying selection into reporting age. If we find no such effect, then the interpretation of the results for average age and the share aged 40 and over. If we do find an effect, then we have to think about the effects of the treatment on both margins – the extensive margin effect on reporting age, and the intensive margin effect on age reported conditional on reporting age. Moreover, if we do find a response, we explore whether there is other evidence from the resumes indicating that older workers were less likely to report age, such as the length of the job history and the vintage of the name based on Social Security data.

Assuming we conclude that there is reason to believe that non-reporting of age in response to ageist language is associated with older ages, we can also collapse the outcome for reporting an age of 40 or over and not reporting an age to create a single out for whether an applicant reported age of 40 or over (where those who did not report age also did not report age over 40); if we find a reporting response, we will estimate models for this outcome as well.

[12] As a robustness exercise, we will also examine how the shares of older workers and younger workers vary when we use the sample of all applicants.

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.[13] 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. For all our analyses, we consider how the significance tests are affected by accounting for multiple comparisons.

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).

Next, we test for heterogeneous effects by treatment type. We first estimate whether there is a difference in responses when using stereotyped language determined by the machine learning similarity scores from Burn et al. (2020) and when using the stereotyped language provided by the AARP ($AARP$). To do this, we modify Equation 1 to include an interaction between the dummy variable for stereotyped language in an ad and a dummy variable for using the AARP language.

$$A = \alpha + \beta_1 S + \beta_2 (S \times AARP) + \delta X + \varepsilon \quad [2]$$

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 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 the individual stereotypes individually before entering all three simultaneously, as shown below. To better understand the mechanisms behind our results, we compare the effect of getting all three treatments to the summed effect of the three treatments, as well as to the effect of the AARP treatment.

[13] 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.

$$A = \alpha + \beta_1 S + \beta_2 (S \times P) + \beta_3 (S \times T) + \beta_4 (S \times AARP) + \delta X + \varepsilon. \quad [3]$$

Heterogeneous Effects

The previous section focused on the average effect of stereotyped language on application patterns of older workers. Next, we turn our attention to understanding how these effects differ across occupation, gender, demographics, and cities. We begin by examining how the effect of stereotyped language differs across occupations. To do this, we expand Equation 1 by including an interaction between the dummy variable for stereotyped ads with dummy variables for each occupation.[14]

$$A = \alpha + \beta_1 (S \times O_1) + \beta_2 (S \times O_2) + \beta_3 (S \times O_3) + \delta X + \varepsilon. \quad [4]$$

Second, we test for heterogeneity in how men and women respond to the stereotyped language. For this analysis, we use the first names collected from the resume. We use U.S. Social Security records to classify a name as either male, female, or unknown. Then, we calculate the average age (using our three measures) for those we can identify as men based on their first name, the average age for those we can identify as women based on their first name, and the average age of those that we cannot assign gender based on their first name. We then include an interaction between a dummy variable for being female (***F***) or having an unknown gender (***U***) and the dummy variable for a stereotyped ad.

$$A = \alpha + \beta_1 S + \beta_2 (S \times F) + \beta_3 (S \times U) + \delta X + \varepsilon. \quad [5]$$

Our next set of heterogeneity analyses explore how worker characteristics mediate the effect of ageist language. We test how the effect of ageist language differs by educational attainment, previous experience, previous experience in the occupation, current employment status, and length of the current unemployment spell. We then explore what effect local labor market characteristics have on our results. We test how our results vary when we interact our treatment indicator with the number of other ads available, the speed at which our ad is no longer on the first page, and the local unemployment rate. We may add other related analyses depending on what information we can code consistently from the

[14] We explain these analyses in terms of Equation [1], but we will also carry them out for Equation [2].

resumes, and we may omit some of the analyses mentioned above if the coding cannot be done consistently. We will note deviations from the pre-analysis plan.

For all the above heterogeneous effects analyses, we take Equation [1] as our base and interact the mediating factor (M) we are interested in with our treatment indicator (S).

$$A = \alpha + \beta_1 S + \beta_2 (S \times M) + \delta X + \varepsilon. \quad [6]$$

Robustness Checks

Our first set of additional analyses will 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 ($\boldsymbol{S}$) with the cosine similarity scores and the ageist sentiments measure derived from our MTURK survey.

Further robustness checks will be conducted which were not specified in the pre-analysis plan. We will be explicit in the text and table and figure notes which analyses were pre-specified and which were exploratory.

Current Status

The experiment has been running since January 2021. We have currently posted 154 jobs (as of November 7) and collected 1,598 resumes, from which we can recover age for 1,377. As of November 1, 2021, we have 121 remaining jobs to post. Due to job ads failing to remain up for the full three days or being flagged before posting by the job board, we will run longer than our target of 54 weeks of data collection. Based on our current posting rate, we anticipate final data collection will be completed on April 8. However, we will be carrying an analysis of all of the data collected in the first 54 weeks, per our registered prespecified analysis plan, and will be able to have a complete draft based on these data.

References

Adams, Scott. 2004. "Age Discrimination Legislation and the Employment of Older Workers." *Labour Economics*, Vol. 11, pp. 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*, Vol. 57, pp. 86-101.

Bem, S.L. and Bem, D.J., 1973. Does Sex-biased Job Advertising "Aid and Abet" Sex Discrimination? *Journal of Applied Social Psychology*, 3(1), pp.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*, Vol. 10, pp. 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*, Vol. 8, pp. 25-46.

Burn, Ian, Patrick Button, Luis Felipe Munguia Corella, David Neumark. 2020. "Older Workers Need Not Apply? Ageist Language in Job Ads and Age Discrimination in Hiring." NBER Working Paper No. 26552.

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*, Vol. 46, pp. 514-23.

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.

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*, Vol. 101, pp. 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*, Vol. 34, pp. 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(4): 276-84.

Hanson, Andrew, Zachary Hawley, Hal Martin, and Bo Liu. 2016. "Discrimination in Mortgage Lending: Evidence from a Correspondence Experiment." *Journal of Urban Economics* 92: 48-65.

He, Haoran, David Neumark, and Qian Weng. "Do Workers Value Flexible Jobs? A Field Experiment. Forthcoming in *Journal of Labor Economics.*

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*, Vol. 17, pp. 409-46.

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

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*, Vol. 117, pp. 441-4.

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

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

Lahey, Joanna. 2008M. "Age, Women, and Hiring: An Experimental Study." *Journal of Human Resources*, Vol. 43, pp. 30-56.

Maestas, Nicole. 2010. "Back to Work: Expectations and Realizations of Work after Retirement." *Journal*

of Human Resources, Vol. (45), pp. 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 (2), pp. 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*, Vol. 106, pp. 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(2), pp. 922-970.
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(2), pp. 373-402.
Ryan, Ellen Bouchard, Sheree Kwong See, W. Bryan Meneer, and Diane Trovato. 1992. "Age-Based Perceptions of Language Performance Among Younger and Older Adults." *Communication Research* 19(4): 423-43.
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*, Vol. 117, 586-626.

Table 1: 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 2: Cities Posted on First Rotation

Week	New York	Houston	Salt Lake City	Sarasota
1	11-Sec-Tech-T	10-Adm-Phys-M	10-Sec-Comm-T	12-Adm-Tech-M
4	05-Adm-Phys-M	07-Ret-Tech-M	08-Ret-All3-T	02-Sec-Cont-T
7	12-Ret-AARP-M	09-Sec-AARP-T	04-Adm-AARP-M	02-Ret-Phys-T
10	11-Sec-AARP-M	10-Adm-Tech-M	10-Sec-Tech-T	12-Adm-Phys-T
13	05-Adm-Cont-T	07-Ret-Phys-M	08-Ret-Comm-T	02-Sec-AARP-M
16	12-Ret-Phys-T	09-Sec-All3-T	04-Adm-All3-M	02-Ret-Comm-M
19	11-Sec-Cont-M	10-Adm-Cont-M	10-Sec-Cont-T	12-Adm-Comm-T
22	05-Adm-Comm-M	07-Ret-Comm-T	08-Ret-AARP-T	02-Sec-All3-M
25	12-Ret-All3-M	09-Sec-Comm-T	04-Adm-Cont-T	02-Ret-All3-M
28	11-Sec-Phys-M	10-Adm-Comm-M	10-Sec-AARP-T	12-Adm-AARP-T
31	05-Adm-AARP-T	07-Ret-All3-T	08-Ret-Phys-M	02-Sec-Phys-M
34	12-Ret-Tech-M	09-Sec-Cont-T	04-Adm-Tech-T	02-Ret-Tech-M
37	11-Sec-Comm-M	10-Adm-AARP-T	10-Sec-Phys-T	12-Adm-All3-M
40	05-Adm-All3-T	07-Ret-AARP-T	08-Ret-Tech-M	02-Sec-Comm-M
43	12-Ret-Cont-M	09-Sec-Phys-M	04-Adm-Comm-T	02-Ret-AARP-T
46	11-Sec-All3-M	10-Adm-All3-M	10-Sec-All3-T	12-Adm-Cont-T
49	05-Adm-Tech-T	07-Ret-Cont-M	08-Ret-Cont-T	02-Sec-Tech-M
52	12-Ret-Comm-M	09-Sec-Tech-M	04-Adm-Phys-T	02-Ret-Cont-T

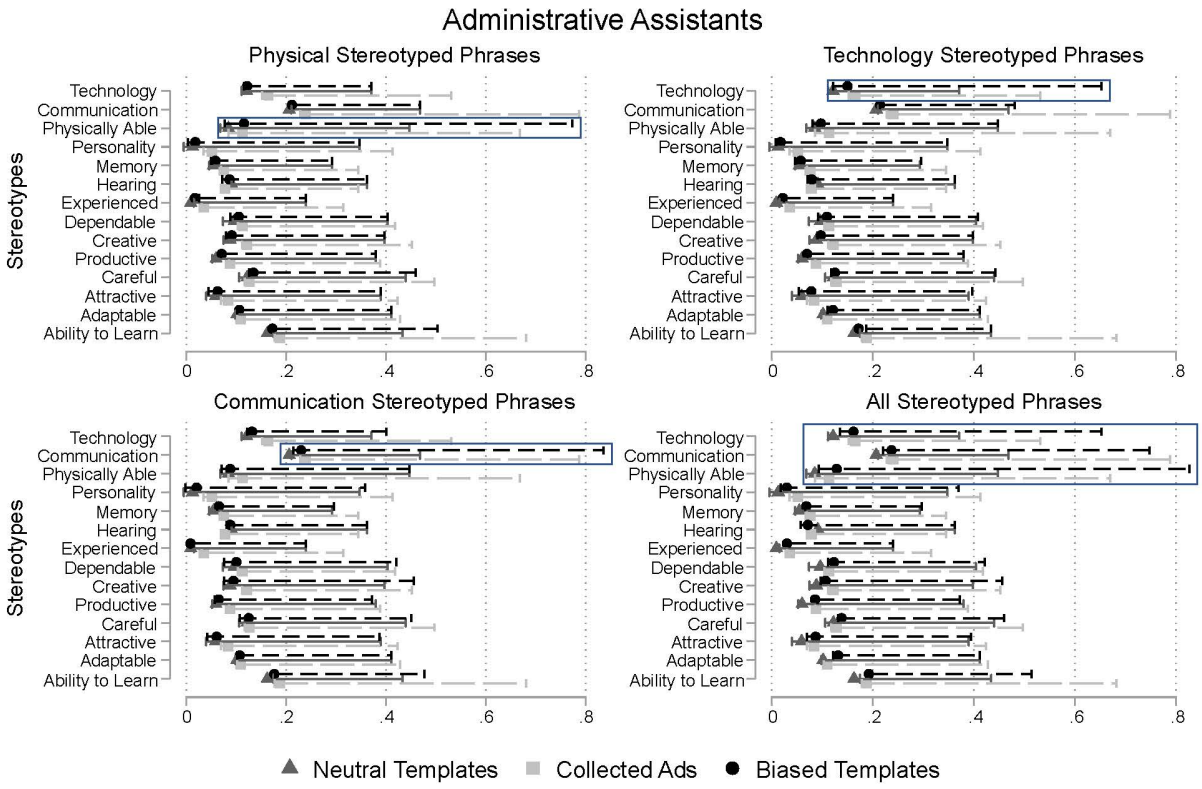
Note: This table reports the posting schedule for the first rotation cities. In each cell, we report the template number (see Appendix A for the text of each template). We report the occupation: Administrative Assistant (Adm), Retail Sales (Ret), and Security Guard (Sec). We report the treatment phrases used: control phrases only (Cont), communication skills treatment (Comm), physical ability treatment (Phys), technology skills treatment (Tech), all three treatment phrases (All3), and AARP treatment phrases (AARP). Finally, we indicate the day of the week the ad will be posted: Monday (M) or Tuesday (T). The format of each cell is "Template #-Job Category-Treatment-Day".

Table 3: Manipulation Checks for Applicants

Template #	Administrative Assistant	Retail	Security Guard
1	put “full-time” or “part-time” in the subject line	send us your resume and your earliest possible starting date and we will be in touch.	Email your resume and potential work hours to apply.
2	please send resume and some interview hours to the provided email	specify your ideal work hours.	Send resumes and propose interview time/dates via email.
3	Please email your resume and availability for more information.	(put full or part-time in the subject line)	If you’re interested in this posting send us your resume and list a few hours for a sit-down interview.
4	Email resumes with a start date in the header!	list your schedule availability.	Reply to this ad with your resume plus your earliest start date and a member of our team will contact you.
5	Put “Re: Administrative Assistant Position” in the subject line.	Email us a resume and indicate part-time or full-time if you are interested.	EMAIL YOUR RESUME AND TYPICAL WEEKLY SCHEDULE FOR CONSIDERATION
6	Be sure to let us know when your earliest possible start date is!	(List your hours.)	If you are qualified to this position contact us soon with your workday preferences.
7	If you fit this description, then please submit your resume by email and preferred work hours!	Please send in your preferred schedule and resume to apply.	SPECIFY FULL OR PART-TIME
8	Please send resume via email with the heading “Administrative Opening”.	Let us know if you are looking for part-time or full-time work in the header of email.	Tell us how many hours you’d like to sign up for!
9	Simply reply to this posting by email with “part time” or “full time” and attach your resume.	list of available work hours if interested.	Send us an email with your resume and a few possible dates for an interview.
10	If you’re interested in this position, email us your resume and let us know the earliest you begin working.	Please send a resume and potential in-person interview hours.	Email resumes and say part-time/full-time to express interest.
11	Apply with your resume and your preferred work hours.	Please indicate your M-F availability in the email.	Please email your resume or summary of experience and suggested work hours for review.
12	Put “Office Assistant Position” in the first line of your email.	let us know many hours per week you prefer.	Contact us by email with your resume and start date.

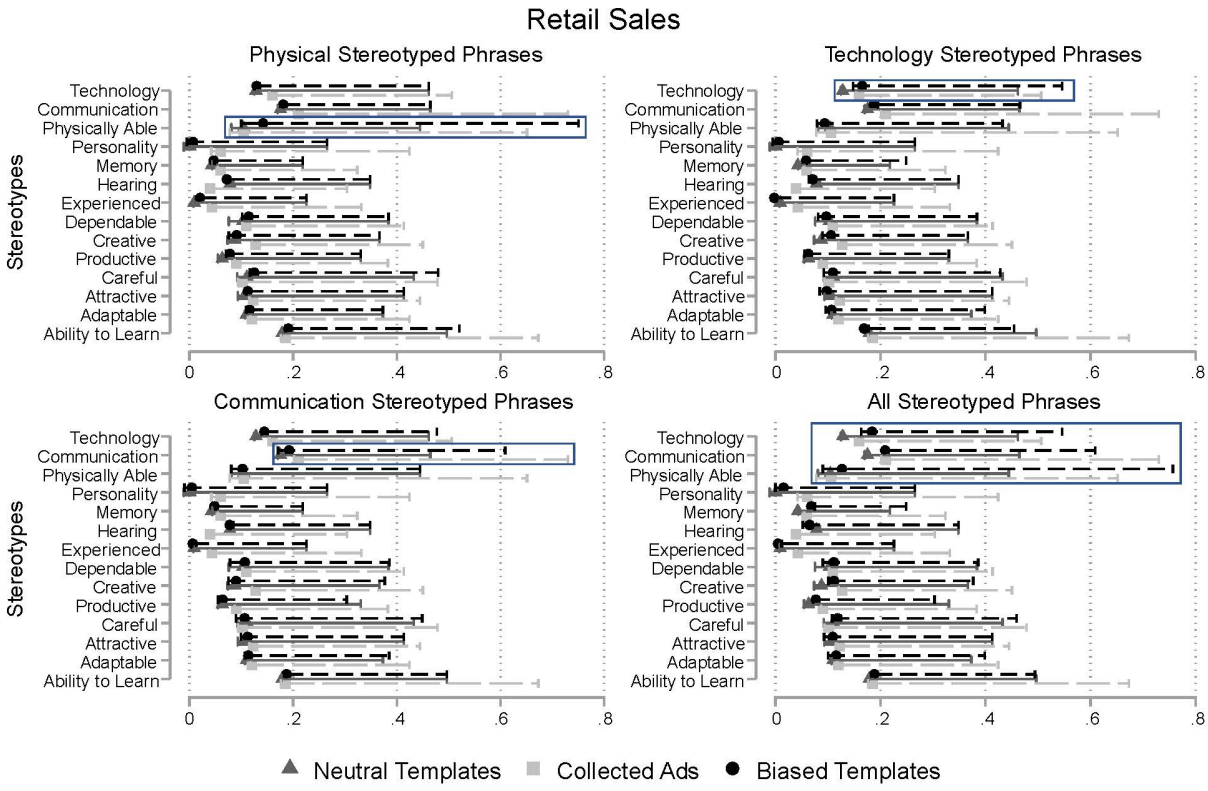
Note: This table lists our manipulation checks for each template.

Figure 1a: Cosine Similarity Score of Administrative Assistant Templates



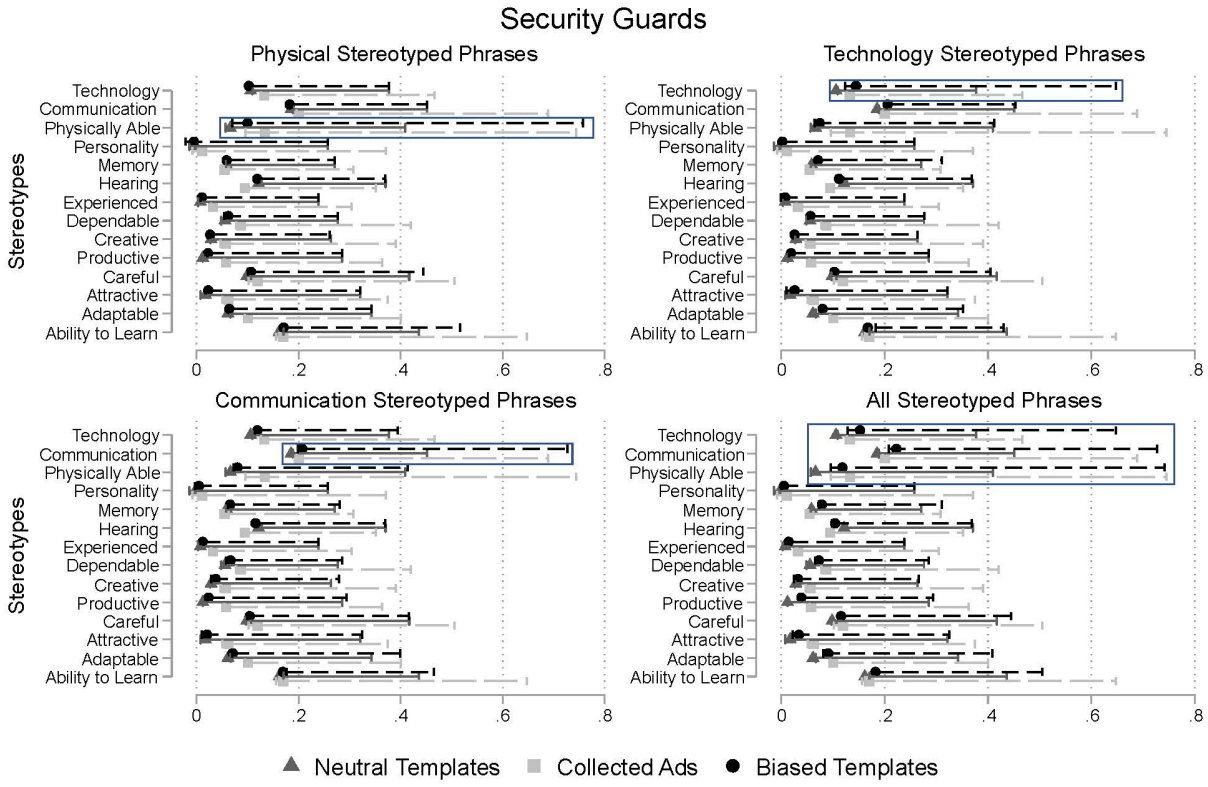
Note: Graphs display median to 99th percentile range of trigram semantic similarity scores for each stereotype for Administrative Assistant ads. The average trigram semantic similarity score for each stereotype is represented by the respective shape for each template. Control templates contain trigrams from the created ad templates with only non-stereotyped phrases included. Collected ads comprise trigrams from all Administrative Assistant job ads from Burn et al. 2020. Treatment templates contain trigrams from the created ad templates with the respective stereotyped phrase or phrases included.

Figure 1b: 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. Control templates contain trigrams from the created ad templates with only non-stereotyped phrases included. Collected ads comprise trigrams from all Retail Sales job ads from Burn et al. 2020. Treatment templates contain trigrams from the created ad templates with the respective stereotyped phrase or phrases included.

Figure 1c: 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. Control templates contain trigrams from the created ad templates with only non-stereotyped phrases included. Collected ads comprise trigrams from all Security Guard job ads from Burn et al. 2020. Treatment templates contain trigrams from the created ad templates with the respective stereotyped phrase or phrases included.

Appendix A: Validating Stereotyped Phrases with MTURK

To ensure that our treatment and control phrases reflect perceived ageist sentiments, we ran an incentivized survey. We recruited participants through an online platform, Amazon MTURK. We restricted the sample to those residing in the United States and, to guarantee that the median age of the sample was roughly 50, used age-based quotas with a third of the sample in each of the following age bins: 25 to 35, 45 to 55, and 55+. Because the age bins are pre-set by Amazon MTURK and MTURK's age data may not be up-to-date, we ask participants to self-report their age. The bins we use to collect self-reported age (25 to 35, 35 to 50, and 50+) broaden the MTURK bins to cover a gap in the age distribution and adjust the highest cutoff to 50, in line with our benchmark age for older workers in the survey. We did not balance their recruitment sample on race or gender.[15]

Our experiment consisted of two parts, the baseline survey and the incentivized survey, which were conducted using Qualtrics. In the baseline survey, we recruited 50 respondents who met our criteria and passed the manipulation check (Figure B1). The responses from the baseline survey were used solely to provide "correct" answers for the next step when we incentivize our second group of respondents to predict the responses of this first group. For the incentivized survey, we recruited 151 respondents who met our criteria and passed the manipulation checks (Figure B2).[16] In Table B1, we report the self-reported demographic composition of our sampled MTURK respondents. The sample is relatively more

[15] Respondents who met the sample restrictions were excluded if they failed a manipulation check as the first step of the survey. This manipulation check acted as both a Turing test (ensuring our respondents were human) and a check for American English language fluency (to reduce the chance someone has masked their location). All questions were free response to prevent individuals or computer programs from clicking through and getting the right answers by accident. The first three questions test for English understanding by asking respondents to complete the analogy (e.g., Canine is to Dog as Feline is to ___, for which cat, kitten, Cat, or cats would all be correct). One of these questions (Pen is to Whiteout as Pencil is to ___) was designed to elicit different responses between American and U.K. English speakers. If participants listed "eraser," the correct answer in American English, they were allowed to proceed, while participants who responded with "rubber," the answer in U.K. English, were excluded. (This turned out not to filter out anyone.) The last question was a free response that asked respondents to write two sentences of at least 140 characters telling us who their favorite band or artist is and why. These were checked after the fact to ensure the participant was paying attention and was capable of writing in coherent English sentences.

[16] We recruited 150 respondents initially – 50 in each age bin – but one person completed the survey and then refused to accept payment. Thus, we ended up with 151 subjects because the quota filling on MTURK only counts people who accepted payment.

educated, white, and female than the U.S. population as a whole. Consistent with the age-based quotas we set for recruiting participants, the sample is also relatively older, which may explain some of the differences in the other demographic characteristics. In line with our target, we generated a sample with a median age near 50, finding roughly 55% of participants over that threshold and 45% below.

Baseline Survey

The baseline survey had three separate blocks of questions. In the first block of questions, subjects were asked to give their informed consent to participate in the survey (Figure B3). In the second block, participants were shown a series of job requirements and told they were from job ads posted online (Figure B4). For each requirement, respondents were asked whether they personally agreed or disagreed with the statement that "[Treatment or control requirement] is biased against workers over the age of 50." Within each of the blocks that asked about whether or not phrases were perceived as biased, there were three pages for each respective stereotype: communication, physical, and technology. All the treatment and control phrases were tested on their respective pages, but the pages were not labeled by stereotypes that respondents viewed. Respondents had to proceed sequentially through the pages and could not go back or skip between them. The ordering of the questions was randomized within each page. We adopted this approach to force respondents to compare treatment and control phrases in a specific stereotype category without prompting them about the specific age-related stereotype in question. The order of questions on a page was randomized for each respondent. Responses to the questions were given on a Likert scale with the options: Strongly agree, Somewhat agree, Neither agree nor disagree, Somewhat disagree, Strongly disagree.

In the third block, we concluded by collecting the demographic characteristics of our sample. They were asked to report their age, gender, education, and race/ethnicity.

Survey

The second survey was split into five blocks of questions. The first asked the participants to give their informed consent. The second block of questions repeated questions from the baseline survey and

asked respondents about their own opinions on job requirements. The purpose of this was to allow us to test how the two groups differed and account for them in our analysis.

The next two parts formed the crux of the incentivized survey. In each part, the respondents were asked to guess how the participants in the baseline survey rated the job requirements, and they were rewarded for how correctly they guessed. Before starting each of these blocks of questions, respondents were sent to a landing page that emphasized the new prompt and cash incentive.

The third set of questions asked respondents to predict what the previous survey respondents answered when they were shown job requirements from the job ads (Figure B5). For each requirement, they were asked whether they thought previous respondents agreed or disagreed with the statement that "This job requirement is biased against workers over the age of 50". They were shown the same Likert scale as was shown to the baseline respondents. Respondents were informed that either the third set of questions in the survey or the fourth (described below) would be randomly selected for a cash incentive based on their answers to the questions in that part. They were told they would earn bonus pay, which was to be calculated based on how close they were to the correct answer. If they correctly predicted what the average participant said, they would earn \$0.32 per question. Incorrect answers received less money, with the penalty increasing the further they were from the correct answer. Payouts were calculated using the quadratic scoring rule

$$P_{iq} = 0.32 - 0.02 \times (\overline{A_q} - A_{iq})^2 \,, \qquad [B.1]$$

where P_{iq} was the payout to individual i for question q based on the average response to question q by previous respondents ($\overline{A_q}$) and their own answer to question q (A_{iq}).[17] For example, they would earn \$0.24 if the correct answer was "somewhat agree," and they guessed "somewhat disagree." All the treatment and control phrases were tested on their respective pages, but the pages were not labeled by

[17] Respondents were told that their earnings, "would be calculated according to the formula: M=\$0.32 − \$0.02 *(Average Previous Answer −-Your Prediction)^2." To illustrate their payoff, they were presented the same example as shown in the text.

stereotypes that respondents viewed. Respondents had to proceed sequentially through pages and could not go back or skip between them. The question order was randomized within each page.

The fourth block of questions differed from the third in that we asked respondents to guess how the older participants in the baseline survey (those over 50 years old) responded (Figure B6). Before starting this block of questions, respondents were sent to a landing page that emphasized both the scoring rule and the age of respondents for whom they were guessing. The instructions read, "For each requirement, please state whether you think previous respondents over the age of 50 agree or disagree with the statement that "This job requirement is biased against workers over the age of 50." The structure of this section was identical to the third block of questions.

Finally, we ended with a demographic survey where we collected data on age, gender, education, and race/ethnicity (Figure B7).

Results

The answers from the MTURK survey participants are provided in Figure B8. 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 fairly 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. They were more mixed with respect to our machine learning-derived treatment phrases. 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.

Table A1: Demographics of MTURK Sample

Demographic Characteristic	Number of Respondents	Percent of Sample
A. Level of Education		
Postgraduate Degree	25	16.6%
Bachelor's Degree	60	39.7%
Some College or 2 Year Degree	49	32.4%
High School Graduate or Less	17	11.3%
B. Age Group		
21 to 35 years old	45	29.8%
35 to 50 years old	23	15.2%
Over 50 years old	83	55.0%
C. Sex		
Female	83	55%
Male	68	45%
D. Race and Ethnicity		
White	125	82.8%
Black or African American	8	5.3%
Asian	6	4.0%
Hispanic or Latino	4	2.7%
Other	3	2.0%
Two or More	5	3.3%

Note: MTURK participants self-reported their demographic characteristics. Respondents who selected two or more of the race and ethnicity categories were grouped into the "Two or More" group.

Figure A1: MTURK Advertisement for Baseline Survey

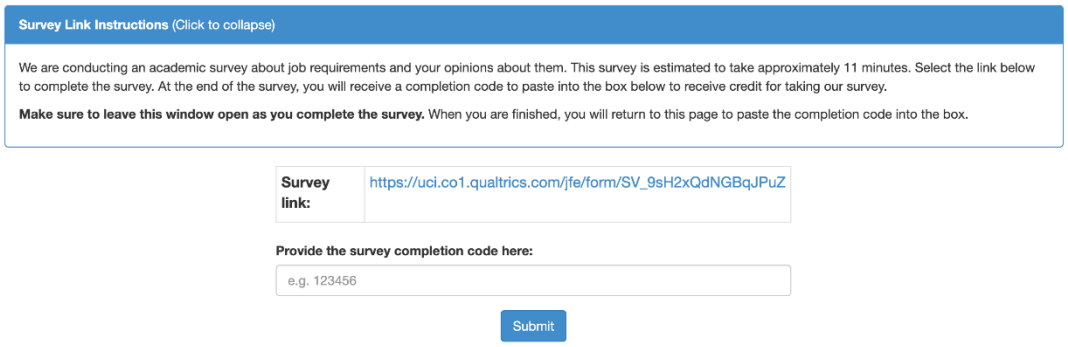


Figure A2: MTURK Advertisement for Incentivized Survey

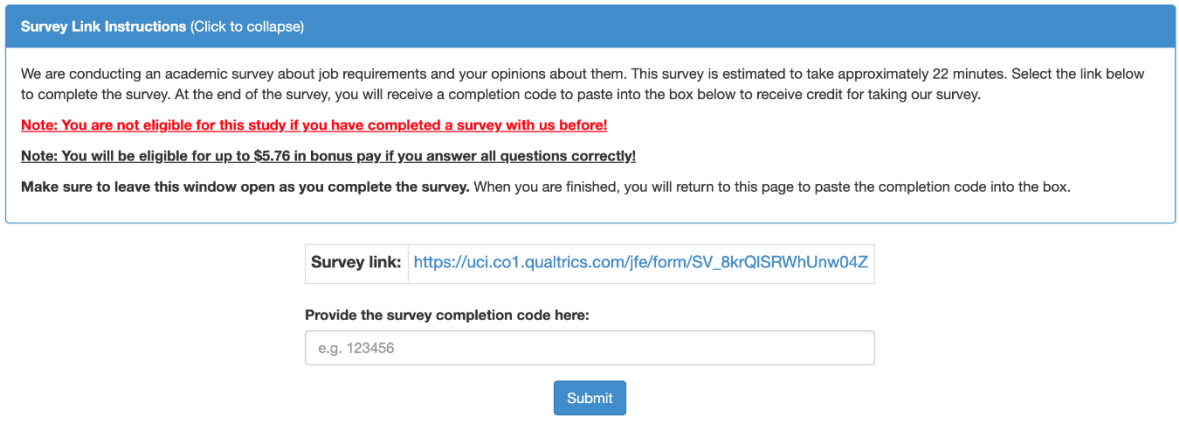


Figure A3: Informed Consent



Thank you for participating in our survey. Before you continue, you must provide your informed consent that you agree to participate in the survey.

Lead Researcher
David Neumark, Distinguished Professor of Economics
Department of Economics
949-824-8496
dneumark@uci.edu

Please read the information below and ask questions about anything that you do not understand. A researcher listed above will be available to answer your questions.

You are being asked to participate in a research study. Participation in this study is voluntary. You may choose to skip a question or a study procedure. You may refuse to participate or discontinue your involvement at any time without penalty or loss of benefits. You are free to withdraw from this study at any time. **If you decide to withdraw from this study you should notify the research team immediately**.

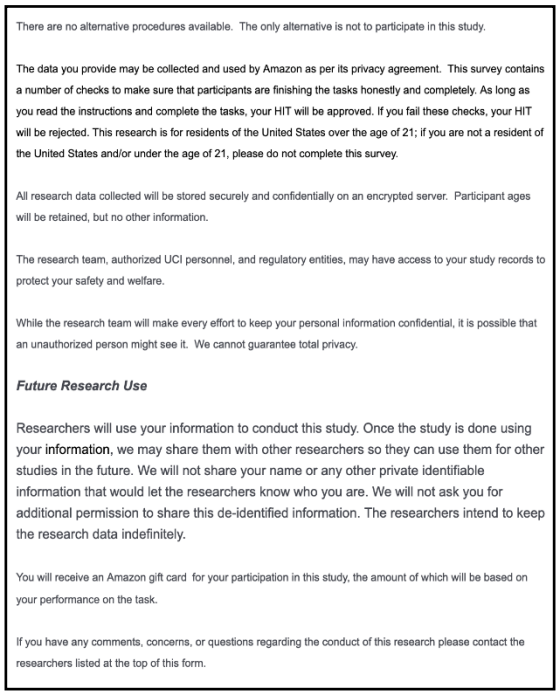
You are being asked to participate in a research study that explores the effects of ageist language or age stereotypes in job ads, to understand how this affects who applies for jobs.

You are eligible to participate in this study if you are age 21 or older and live in the United States.

The research procedures involve reading a list of job requirements, rating the phrasing of the job requirements as ageist (on a 1-5 scale), providing a similar rating for how ageist the average person would view the phrasing of the job requirements, and providing a similar rating for how ageist a worker aged 50-65 would view the phrasing of job requirements. The research procedures will last about 15 minutes.

Possible risks/discomforts associated with the study are non-existent.

There are no direct benefits from participation in the study. However, this study may explain how employers influence the ages of who applies to their jobs.



There are no alternative procedures available. The only alternative is not to participate in this study.

The data you provide may be collected and used by Amazon as per its privacy agreement. This survey contains a number of checks to make sure that participants are finishing the tasks honestly and completely. As long as you read the instructions and complete the tasks, your HIT will be approved. If you fail these checks, your HIT will be rejected. This research is for residents of the United States over the age of 21; if you are not a resident of the United States and/or under the age of 21, please do not complete this survey.

All research data collected will be stored securely and confidentially on an encrypted server. Participant ages will be retained, but no other information.

The research team, authorized UCI personnel, and regulatory entities, may have access to your study records to protect your safety and welfare.

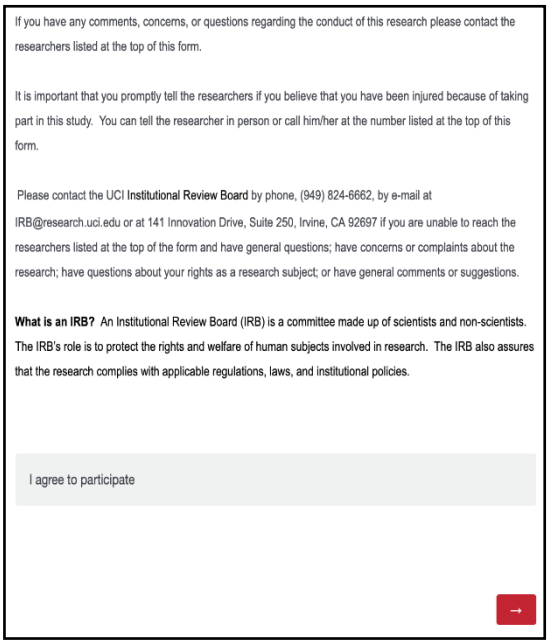
While the research team will make every effort to keep your personal information confidential, it is possible that an unauthorized person might see it. We cannot guarantee total privacy.

Future Research Use

Researchers will use your information to conduct this study. Once the study is done using your information, we may share them with other researchers so they can use them for other studies in the future. We will not share your name or any other private identifiable information that would let the researchers know who you are. We will not ask you for additional permission to share this de-identified information. The researchers intend to keep the research data indefinitely.

You will receive an Amazon gift card for your participation in this study, the amount of which will be based on your performance on the task.

If you have any comments, concerns, or questions regarding the conduct of this research please contact the researchers listed at the top of this form.



If you have any comments, concerns, or questions regarding the conduct of this research please contact the researchers listed at the top of this form.

It is important that you promptly tell the researchers if you believe that you have been injured because of taking part in this study. You can tell the researcher in person or call him/her at the number listed at the top of this form.

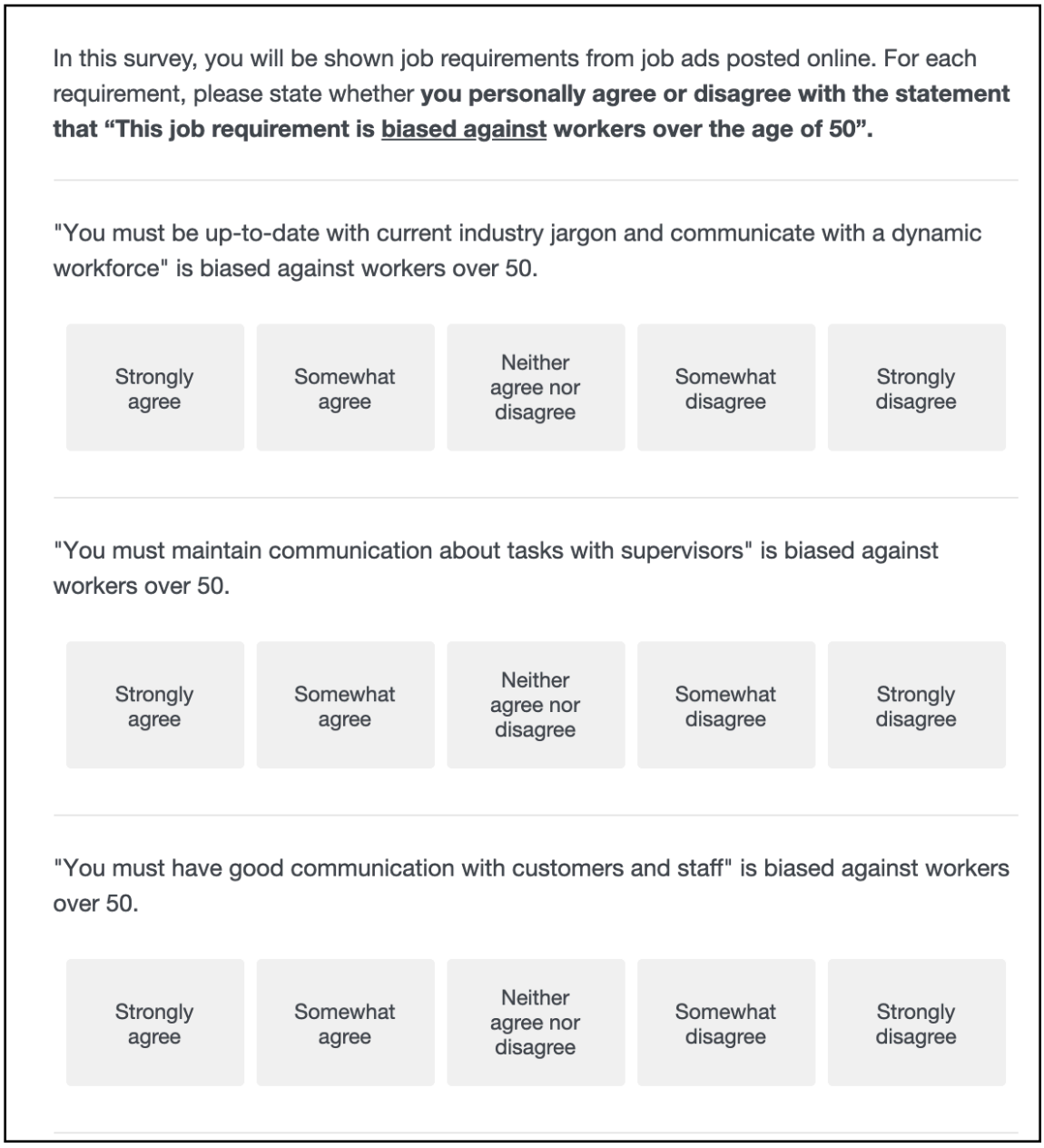
Please contact the UCI **Institutional Review Board** by phone, (949) 824-6662, by e-mail at IRB@research.uci.edu or at 141 Innovation Drive, Suite 250, Irvine, CA 92697 if you are unable to reach the researchers listed at the top of the form and have general questions; have concerns or complaints about the research; have questions about your rights as a research subject; or have general comments or suggestions.

What is an IRB? An Institutional Review Board (IRB) is a committee made up of scientists and non-scientists. The IRB's role is to protect the rights and welfare of human subjects involved in research. The IRB also assures that the research complies with applicable regulations, laws, and institutional policies.

I agree to participate

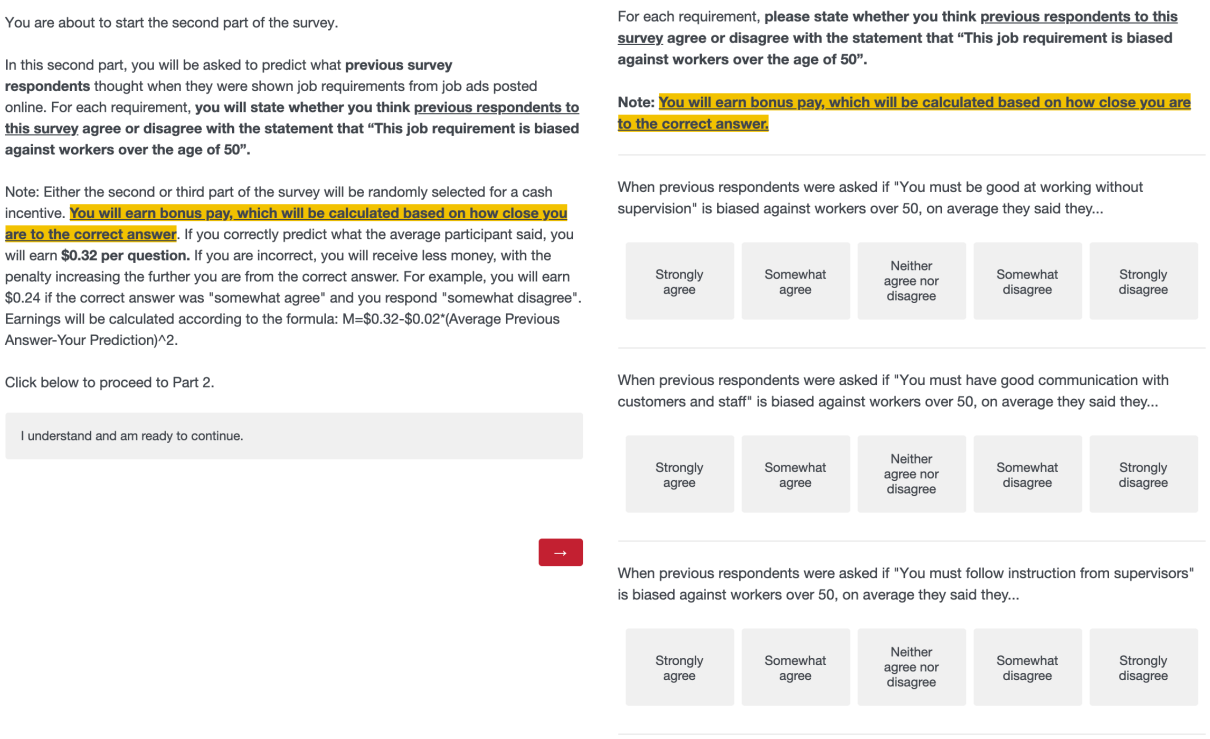
→

Figure A4: Block 1 Example (Exists in Survey 1 and Survey 2)



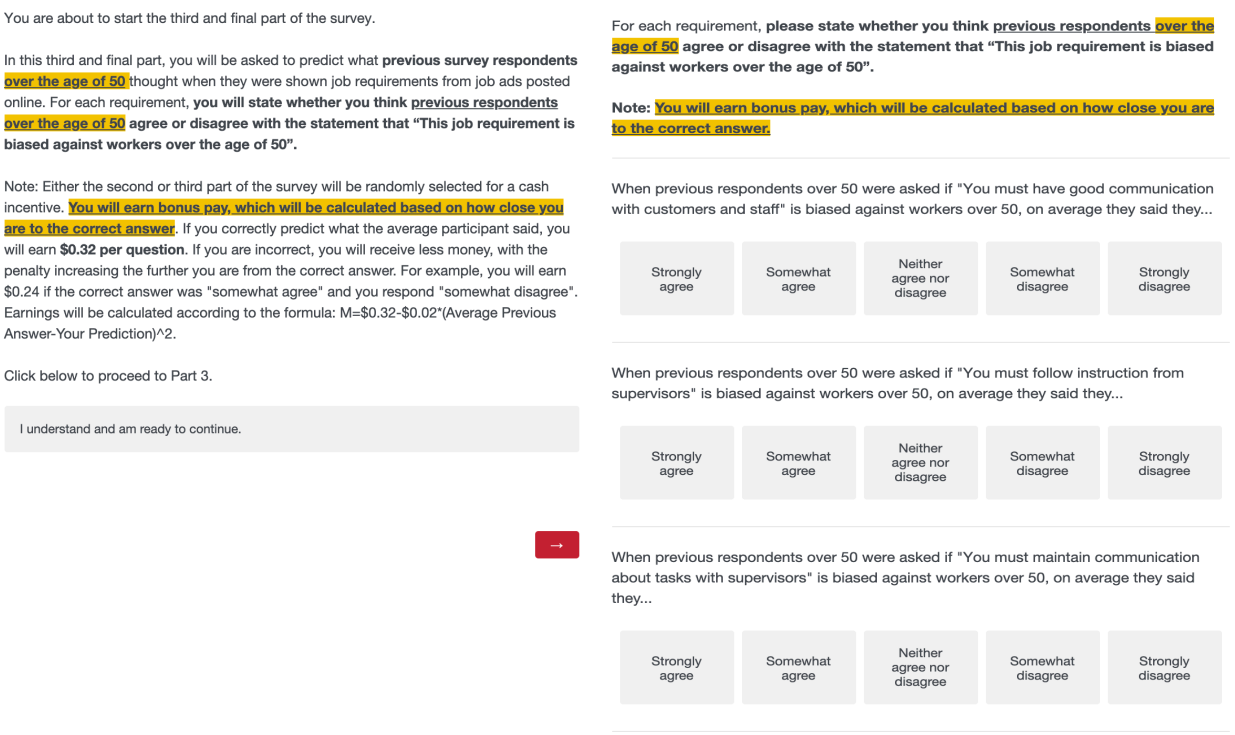
Note: Respondents were asked about their opinions for all of the treatment and control phrases within a given age-related stereotype. The treatment and control phrases within the communication, physical, and technology categories were shown on three separate pages, and the Qualtrics software disallowed skipping or going back between sections. The order of phrases was randomized for each participant, and they were required to provide an answer in order to proceed to the next section of the survey. In this example, the first three phrases for the communication category are shown.

Figure A5: Block 2 Landing Page and Example



Note: After survey participants completed the block soliciting their personal opinions on whether certain phrases were age-biased, they were directed to the landing page pictured on the left-hand side of the figure. The landing page stressed the incentivized nature of the subsequent section and that bonus pay would be based on how accurately they could identify what the average respondent to the baseline survey said. The questions in the section reflected the new prompt and a highlighted portion at the top of each page reminded MTURK participants that their answers were incentivized based on accuracy.

Figure A6: Block 3 Landing Page and Example (Exists only Survey 2)



You are about to start the third and final part of the survey.

In this third and final part, you will be asked to predict what **previous survey respondents over the age of 50** thought when they were shown job requirements from job ads posted online. For each requirement, **you will state whether you think previous respondents over the age of 50 agree or disagree with the statement that "This job requirement is biased against workers over the age of 50".**

Note: Either the second or third part of the survey will be randomly selected for a cash incentive. **You will earn bonus pay, which will be calculated based on how close you are to the correct answer**. If you correctly predict what the average participant said, you will earn **$0.32 per question**. If you are incorrect, you will receive less money, with the penalty increasing the further you are from the correct answer. For example, you will earn $0.24 if the correct answer was "somewhat agree" and you respond "somewhat disagree". Earnings will be calculated according to the formula: M=$0.32-$0.02*(Average Previous Answer-Your Prediction)^2.

Click below to proceed to Part 3.

I understand and am ready to continue.

→

For each requirement, **please state whether you think previous respondents over the age of 50 agree or disagree with the statement that "This job requirement is biased against workers over the age of 50".**

Note: You will earn bonus pay, which will be calculated based on how close you are to the correct answer.

When previous respondents over 50 were asked if "You must have good communication with customers and staff" is biased against workers over 50, on average they said they...

Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree

When previous respondents over 50 were asked if "You must follow instruction from supervisors" is biased against workers over 50, on average they said they...

Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree

When previous respondents over 50 were asked if "You must maintain communication about tasks with supervisors" is biased against workers over 50, on average they said they...

Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree

Figure A7: Demographic Questions

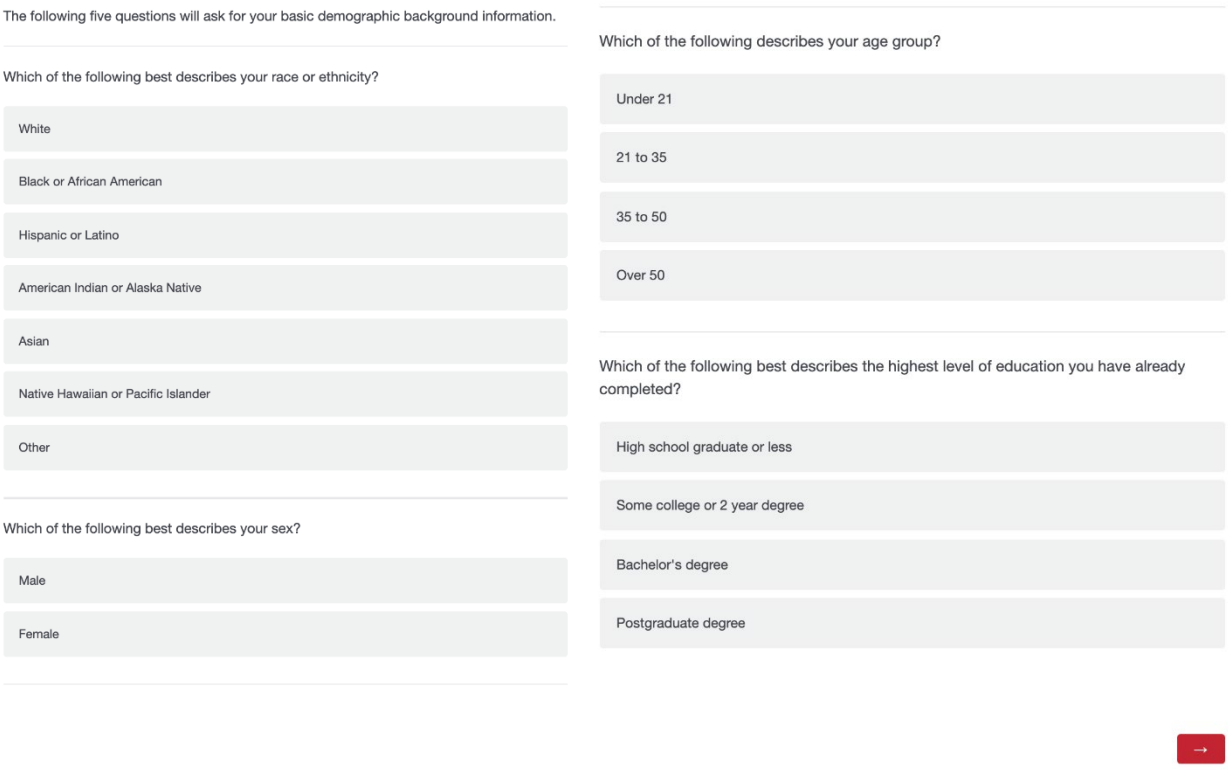
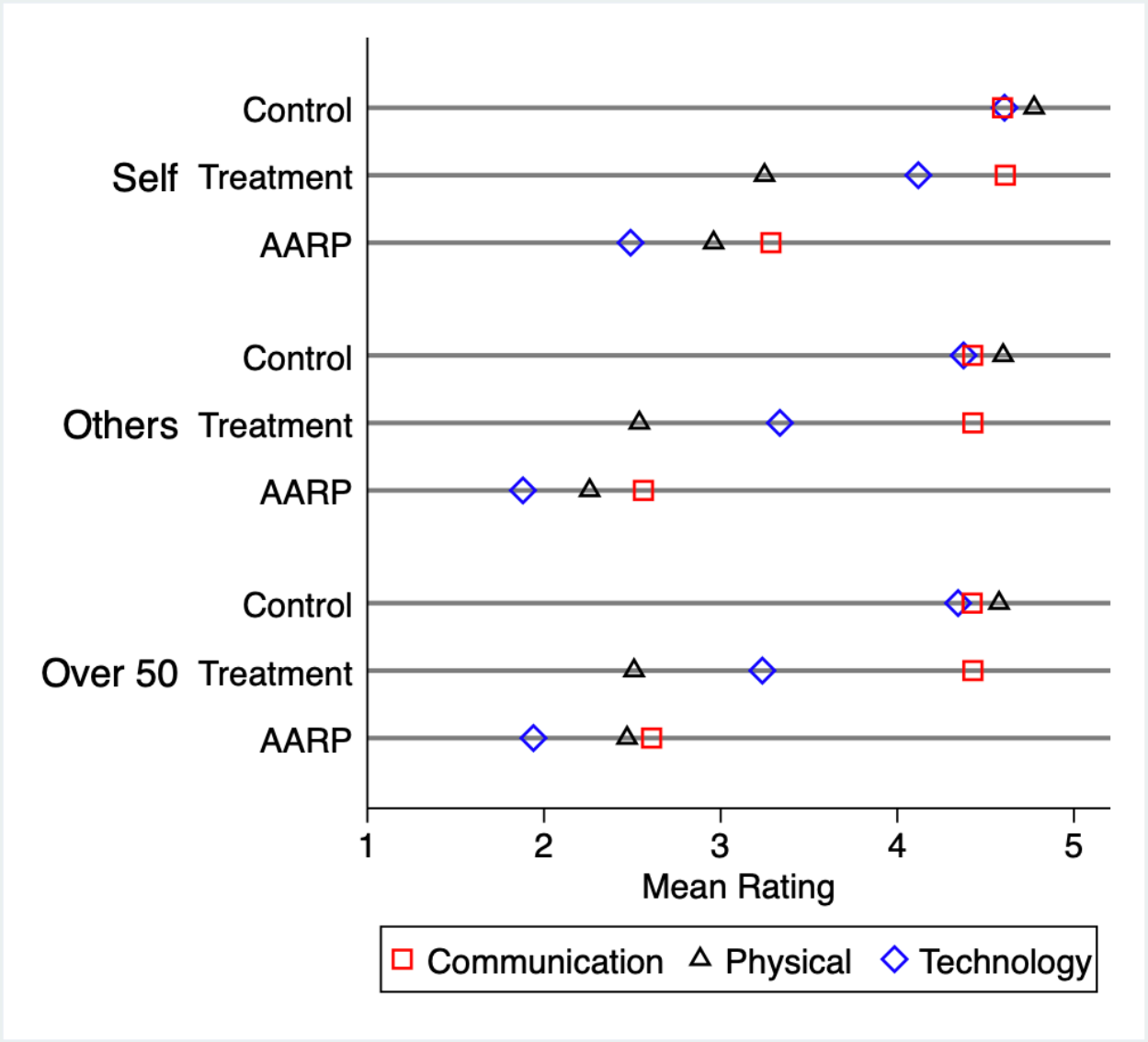


Figure A8: Survey Results



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.