Journal Pre-proof

What Does a Job Candidate's Age Signal to Employers?

Hannah Van Borm , Ian Burn , Stijn Baert

PII: S0927-5371(21)00038-5
DOI: https://doi.org/10.1016/j.labeco.2021.102003
Reference: LABECO 102003

To appear in: *Labour Economics*

Received date: 24 June 2020
Revised date: 20 February 2021
Accepted date: 14 May 2021

Please cite this article as: Hannah Van Borm , Ian Burn , Stijn Baert , What Does a Job Candidate's Age Signal to Employers?, *Labour Economics* (2021), doi: https://doi.org/10.1016/j.labeco.2021.102003

Highlights

- An applicant's older age signals to recruiters that the applicant has lower technological skills, flexibility, and trainability levels.
- An applicant's perceived technological knowledge and skills, flexibility, and trainability explain 41% of the total effect of age on a job applicant's interview chances.
- Negative association between age and invitation to interview probability is smaller when recruiters work for firms with a higher percentage of older employees.

What Does a Job Candidate's Age Signal to Employers?*

By Hannah Van Borm,[i] Ian Burn,[ii] and Stijn Baert[iii]

Abstract

Research has shown that hiring discrimination is a barrier for older job candidates in many OECD countries. However, little research has delved into why these job candidates face discrimination. Therefore, we have conducted an online scenario experiment involving recruiters to empirically investigate 15 potential stigmas related to older age drawn from a systematic review of the literature. We found that older age particularly signals to recruiters that the applicant has lower technological skills, flexibility, and trainability levels. Together, these perceptions explain about 41% of the effect of age on the probability of being invited to a job interview. Additionally, we found that the negative association between age and invitation probability is smaller when recruiters work for firms with a higher percentage of older employees.

Keywords: hiring, statistical discrimination, age, stereotypes.

JEL-classification: J71, J14, J24, J23.

* **Acknowledgements.** We thank Magnus Carlsson, Andrew Hayes, Olivier Marie, David Neumark, Matthew Notowidigdo, Panu Poutvaara and the participants of the 2018 Belgian Day of Labour Economists, the 2018 Summer School on Experimental Economics at the Paris School of Economics and the 2019 Workshop on Older Workers' Skills and Labour Market Behavior in Maastricht for their insightful comments and suggestions, which have helped to improve this study considerably. In addition, we thank our colleague Brecht Neyt for his advice.

[i] **Corresponding author.** Ghent University. Sint-Pietersplein 6, 9000 Ghent, Belgium. Hannah.VanBorm@UGent.be.

[ii] University of Liverpool and SOFI.

[iii] Ghent University, University of Antwerp, Université catholique de Louvain, IZA, GLO, and IMISCOE.

1. Introduction

The financing of Pay-As-You-Go pension systems, where labour income taxes paid by the working population are used to finance the pensions of the retired population, has become a major problem for many OECD countries (Barr, 2006; Attanasio, Kitao, & Violante, 2007; McGrattan & Prescott, 2017). That is, the increase in life expectancy (Attanasio et al., 2007; Kontis, Bennett, Mathers, Foreman, & Ezzati, 2017; OECD, 2019b), and decrease in fertility to below the replacement level (Attanasio et al., 2007; OECD, 2019b) has led to rising dependency ratios. The most commonly suggested solution for this financing problem is to make people work longer (Breyer & Kifmann, 2002; Munnell & Sass, 2009; Maestas & Zissimopoulos, 2010; Harkin, 2012; Kitao, 2014). When comparing the employment rate of individuals over 55 with those aged 25–54 in various developed countries, there is still a significant margin for improvement in this respect. In the United States, for example, 63.5% of the population aged 55–64 and 33.0% of the population aged 65 to 69 was employed in 2018, which is remarkably less than the 79.9% employment rate for the population aged 25–54 (OECD, 2019a).[1]

In practice, however, raising the employment rate of people aged 55 and over is not that straightforward. There are various explanations for the lower employment rates of people over 55, one of which is age discrimination in hiring (Boissonneault et al., 2020). Using correspondence field experiments, previous research has found considerable evidence of age discrimination in hiring in the United States (Johnson & Neumark, 1997; Lahey, 2008; Farber, Silverman, & von Wachter, 2016; Neumark, Burn, & Button, 2016, 2019; Neumark, Burn, Button, & Chehras, 2019; Neumark, 2018), the United Kingdom (Riach & Rich, 2010; Tinsley, 2012; Riach, 2015; Drydakis, MacDonald, Bozani, & Chiotis, 2017), and the European Union (Ahmed, Andersson, & Hammarstedt 2012; Riach, 2015;

[1] The remark has to be made that the employment rate of individuals aged 55 to 64 has been increasing in the last 25 years (Boissonneault, Mulders, Turek, & Carriere, 2020; OECD, 2017). However, despite this increasing trend, the gap between the employment rate of individuals between 55 and 64 years old and people between the ages of 25 and 54 remains large.

Baert, Norga, Thuy, & Van Hecke, 2016; Carlsson & Eriksson, 2019).[2, 3] Hiring discrimination pushes older individuals, especially those who are unemployed (Deelen, de Graaf-Zijl, & van den Berge, 2018) or unsatisfied with their current job (Scharn, Sewdas, Boot, Huisman, Lindeboom, & Van Der Beek, 2018) out of the labour market, forcing many to claim retirement benefits before full retirement age. To induce and enable people to work longer, it is, therefore, crucial to understand the mechanisms underlying age discrimination and craft policies to relax them.

In the economic literature, two main theoretical models provide explanations as to why employers may discriminate against older workers when hiring new employees: Arrow's (1973) model of statistical discrimination and Becker's (1957) model of taste-based discrimination. The model of statistical discrimination posits that age discrimination in the hiring process is driven by stereotypes concerning older workers' productivity (Arrow, 1973). When making hiring decisions, recruiters often have a limited amount of information about a job applicant, such as their age, gender, education level, and work experience. As a result, they might use this limited information as a signal for other, unobserved characteristics concerning the applicant's productivity (Arrow, 1973; Spence, 1973; Vishwanath, 1989; Blanchard & Diamond, 1994; Moscarini, 1997; Kroft, Lange, & Notowidigo, 2013; Eriksson & Rooth, 2014). Following this theory, older applicants might not be hired because older age signals, for example, lower levels of physical ability (Schmidt & Boland, 1986; Hummert, Garstka, Shaner, & Strahm, 1994; Finkelstein, Burke, & Raju, 1995; Kroon, van Selm, ter Hoeven, & Vliegenthart, 2016) or flexibility (Warr & Pennington, 1993; AARP, 1999; Büsch, Dahl, & Dittrich, 2009; McCann & Keaton, 2013).[4] On the other hand, the taste-based discrimination model indicates that employers discriminate against older job applicants because they, their employees, or their customers might experience a decrease in utility when interacting with older workers (Becker, 1957).[5, 6]

[2] A correspondence test is a type of field experiment often used to measure hiring discrimination. In these tests, fictitious résumés which vary only in terms of a specific characteristic of interest are sent to actual job openings. Subsequently, the callbacks involving these profiles are examined (Neumark, 2018). Correspondence tests are viewed by many as the golden standard for evidence for discrimination (Baert, 2018a; Neumark 2018).

[3] Previous research also found suggestive evidence for age discrimination in the labour market concerning dismissal (Johnson & Neumark, 1997; Roscigno, Mong, Byron, & Tester, 2007), promotions (Rosen & Jerdee, 1976,1977; Johnson & Neumark, 1997; Taylor & Walker, 1998; Adams, 2002), and training opportunities (Rosen & Jerdee, 1976, 1977; Johnson & Neumark, 1997; Taylor & Walker, 1998; Taylor & Urwin, 2001). For overviews of experimental research studying age discrimination in the labour market, see Baert et al. (2016), Baert (2018a), and Neumark (2018).

[4] For an overview of the empirical research on stereotypes (negative and positive) concerning older workers' productivity found in economics, industrial psychology, communication sciences, and related fields, see Burn et al. (2019). For a discussion on whether these stereotypes are, on average, correct or whether they are potentially erroneous, see Neumark, Burn, and Button (2019).

[5] The origin of negative attitudes towards collaborating with older workers may be linked to the theory of terror management developed by Greenberg, Pyszczyski, and Solomon (1986). This theory implies that these attitudes might be rooted in the fear of dying. This fear might

Although some of the correspondence studies mentioned above also tried to test these models empirically (e.g., Baert et al., 2016; Drydakis et al., 2017; Lahey, 2008; Neumark, Burn, & Button, 2016), the results of these studies concerning the explanations for age discrimination in hiring are often inconclusive. While correspondence field experiments are a great tool to measure hiring discrimination convincingly, the method is less suitable to gain insights into the causes of this discrimination. In recent years, however, researchers have used alternative ways to gain insights into the underlying mechanisms of age discrimination in hiring. In particular, Richardson, Webb, Webber, and Smith (2013) use a lab experiment to test which stereotypes predict discrimination. More concretely, they design a vignette experiment in which participants evaluate fictitious job applicants, of whom age is varied, for a hypothetical job vacancy. The authors examine whether a fictitious applicant's age affects perceptions about their reliability, sociability, trainability, and intellectual competence and to which extent these perceptions play a mediating role in the hiring decisions of the participants. They find that, although an applicant's age negatively affects evaluations of her or his trainability and sociability, the effect of the applicant's age on hiring evaluations was not mediated by these work-related competencies. Additionally, Burn, Button, Corella, and Neumark (2019) experimentally test whether the ageist language in job ads is correlated with hiring discrimination. These authors find that language related to ageist stereotypes is over-represented in the phrases selected by machine learning algorithms as predicting discrimination. Older workers are more likely to be discriminated against when job ads use sentences related to physical ability, technology, or communication skills.[7]

Evidence of stereotypes' effect on the hiring of older workers can also be found in survey research. Both Taylor and Walker (1998), Carlsson and Eriksson (2019), and Jensen, De Tavernier and Nielsen (2019) surveyed employers, respectively in the UK, Sweden, and Denmark, regarding their perceptions of older workers and different workplace practices, among which recruitment practices. Where both Taylor and Walker (1998) and Carlsson and Eriksson (2019) found an association between different perceptions of older workers (e.g., concerning their trainability) and recruitment practices,

lead younger individuals to distance themselves from older people to avoid reminders of their mortality (Greenberg, Pyszczynski, & Solomon, 1986; Martens, Goldenberg, & Greenberg, 2005; Nelson, 2005).

[6] For empirical literature that identifies overall negative attitudes and prejudices towards older individuals see, for example, Kite and Johnson (1988) and Nelson (2004).

[7] Another study that uses experimental data to explain hiring discrimination towards older (black) job applicants is Lahey and Oxley (2018). In their study, the researchers use eye-tracking to test various explanations for ethnic discrimination over the lifecycle. Their reseach, however, mainly focuses on race (and its interaction with age) and does not answer the question of which specific stereotypes and attitudes underlie hiring discrimination towards older job applicants, as done in the abovementioned studies.

Jensen et al. (2019) did not. Furthermore, Turek and Henkens (2019) used employer surveys from Poland to assess how likely employers are to recruit people over 50 years old and studied how the probability of inviting an older candidate to an interview varied as the skill requirements of the job post changed. These authors observed that older candidates were less likely to be hired in jobs requiring computer, physical, social, creative, and training skills.[8]

In the present article, we contribute to this literature by using a state-of-the-art vignette experiment in line with Richardson et al. (2013). In our experiment, we show participants with genuine hiring experience a series of fictitious résumés, varying over the applicants' age, gender, and some other attributes, which they had to evaluate for a hypothetical vacancy. The participants had to assess these fictitious profiles concerning various characteristics related to the productivity-related stigma identified in Burn et al.'s (2019) literature review and potential negative attitudes towards collaborating with older employees in line with the theory of taste-based discrimination. Consequentially, our design enables us to identify employer perceptions and attitudes towards older job candidates and to explore the degree to which these perceptions and attitudes act as drivers of potential age discrimination. We randomly assigned subjects to review applicants for one of eight job vacancies varying along four different dimensions, i.e., the required degree of skills, the level of customer contact, the required amount of physical effort, and the required level of technological knowledge associated with the job, allowing us to investigate the heterogeneity in unequal treatment of older job applicants by these four dimensions. Furthermore, we surveyed our participants concerning their background characteristics to identify the possible moderating effects of these characteristics on age discrimination in hiring (e.g., the participants' age and the percentage of older workers employed by their companies).

This study improves on the previous literature in four meaningful ways. First, we investigate a more systematic and extensive set of explanations for age discrimination in hiring compared to previous research (i.e., all factors identified by the existing theoretical and survey-based literature). While former contributions restricted their attention to a limited set of stereotypes (e.g., Richardson et al. (2013) and the studies based on employer surveys, mentioned before), our research takes into account (almost) all stereotypes identified by Burn et al. (2019) in their literature review. Additionally, while none of the former contributions investigates (nor controls for) taste-based discrimination (Becker, 1957), we consider employers', customers', and employees' potential negative attitudes

[8] For qualitative research on the relationship between attitudes and perceptions regarding older workers and discriminatory practices in the labour market, see Loretto and White (2006).

towards collaborating with older employees as an alternative explanation for age-based discrimination in hiring. Second, we also use a more systematic method (i.e., a vignette experiment and a thorough mediation analysis) to analyse the relative importance of these factors compared to most existing studies (except for Richardson et al. (2013)). Third, in our study, we account for many more of the pathways through which individual characteristics and job characteristics interact with age. The vignette experiment used by Richardson et al. (2013) and the surveys by Taylor and Walker (1998), Carlsson and Eriksson (2019), and Turek and Henkens (2019) only questioned employers about a limited number of characteristics involving the job (in the case of Taylor and Walker (1998) and Turek and Henkens (2019)) or the individual participants (in the case of Taylor and Walker (1998), Carlsson and Eriksson (2019), and Richardson et al. (2013)) as moderators of age discrimination. Lastly, our vignette study is an improvement over Richardson et al. (2013) in terms of scale and external validity. Our study features 2000 candidate evaluations (a much larger sample than that employed by Richardson et al. (2013)), studies both men and women (whereas Richardson et al. (2013) only used male applicants), considers several different types of job vacancies (Richardson et al. (2013) used only a position in the IT industry), and studies the behaviour of actual human resource managers (while Richardson et al. (2013) mainly worked with students).[9, 10] Taken together, these improvements provide scholars and policymakers with more general insights into the mechanisms underlying age discrimination (as well as its moderators).

2. Data

To gain insights into the potential drivers and moderators of age discrimination in hiring, we conducted a vignette experiment. A vignette experiment is an application of the factorial survey method (Rossi & Nock, 1982; Auspurg & Hinz, 2014) and is often used to study human judgements in the fields of psychology, sociology, and economics (Jasso, 2006; Derous, Nguyen, & Ryan, 2009; Derous, Ryan, & Nguyen, 2012; Eriksson & Kristensen, 2014; Rivera & Tilcsik, 2016; Ambuehl & Ockenfels, 2017; Auspurg, Hinz, & Sauer, 2017; Mathew, 2017; Cavalier, Hampton, Langford, Symes, & Young, 2018). Furthermore, previous research has used this type of experiment extensively to study

[9] Richardson et al. (2013) analysed 102 students and 52 experienced employees' evaluations of one younger versus one older job candidate.

[10] We asked recruiters to evaluate candidates in a wide range of fields, which may not have corresponded to the types of worker they typically hire. To improve the external validity in light of this concern, we provided the recruiters with the job descriptions from the O*NET classifications.

hiring discrimination and decisions in the labour market (Van Hoye & Lievens, 2003; Derous et al., 2009; Derous et al., 2012; Baert & De Pauw, 2014; Di Stasio, 2014; Baert, 2018b; Van Belle, Di Stasio, Caers, De Couck, & Baert, 2018; Van Borm & Baert, 2018; Damelang, Abraham, Ebensperger, & Stumpf, 2019; Van Belle, Caers, De Couck, Di Stasio, & Baert, 2019).

In these experiments, participants judge short, fictitious descriptions of individuals or situations depicted in the vignettes, for which the characteristics (the vignette factors) vary systematically or randomly over a predefined number of categories (the vignette levels) (Sauer, Auspurg, Hinz, & Liebig, 2011). One of the main advantages of a vignette experiment over non-experimental research is that the experimental manipulation of the vignette levels allows for a causal interpretation of the effect of each vignette factor on participants' evaluations (Wallander, 2009; Damelang & Abraham, 2016; Van Belle et al., 2018). Vignette experiments are more flexible than the correspondence field experiments often used to study hiring decisions. The latter experiments measure just the binary decision to offer a candidate an interview or not, while vignette experiments make it possible to investigate a broader array of decisions and the motivations behind these decisions. Hence, using a vignette experiment allowed us to survey the participants about their characteristics and beliefs regarding fictitious job applicants of varying ages, which we would not have been able to do had we conducted a correspondence experiment.

2.1. Vignette Design

In our experiment, we asked each participant to evaluate a deck of five unique vignettes in which we presented tabulated information about a fictitious job candidate (one per vignette).[11] More concretely, the fictive job candidates differed in five distinct characteristics, which varied over a predefined number of levels. We provide an overview of these different factors and their associated levels in Table 1 and discuss them below.

< Table 1 about here >

The main factor of interest in our experiment was age. Similar to Richardson et al. (2013), Lahey and

[11] As recommended by Auspurg and Hinz (2014), we work with tabular vignettes instead of text vignettes because these are better suited to decision tasks involving lists of decision criteria, such as evaluating fictitious résumés. Indeed, tabular vignettes' more straightforward presentations of vignette factors help participants to form more consistent judgements, which is especially useful in the context of our experiment in which participants have to evaluate multiple fictitious job applicants. Moreover, previous research has shown that tabular vignette designs produce similar evaluations to text vignettes (Auspurg & Hinz, 2014; Shamon, Dülmer, & Giza, 2019; Sauer, Auspurg, & Hinz, 2020).

Oxley (2018), and Carlsson and Eriksson (2019), we decided to use a continuous variable to reveal age on the profiles because it allows for a straightforward interpretation of our results.[12] More concretely, the ages of the applicants ranged from 32 to 63. We decided to use the age of 32 as a lower cut-off value because applicants at this age may already have enough experience in the labour market to compete with older job applicants (Lahey, 2008; Neumark et al., 2019). Additionally, we opted for an upper cut-off of 63 to avoid applicants too close to retirement age. To mimic real-life hiring decisions as closely as possible and cover up the main goal of the research to avoid social desirability bias, we also let the applicants differ regarding (i) gender (male or female), (ii) commuting distance (0–5 km, 5–10 km, 10–50 km, or more than 50 km), (iii) experience in the occupation (none, about 2 years, about 5 years, or about 10 years), and (iv) extracurricular activities (none, volunteer work, participating in sports, or engaging in cultural activities).[13] These additional factors and their levels are all elements that are typically revealed on résumés and were drawn from the previous literature (Olian, Schwab, & Haberfeld, 1988; Lahey, 2008; Nuijten, Poell, & Alfes, 2017; Carlsson, Reshid, & Rooth, 2018). To investigate whether these additional factors and their levels were perceived to be relevant for employers, we conducted a pilot test of our survey with 193 Belgian recruiters (see Section 2.3.). The results of this pilot indicated that the candidate characteristics mentioned in the vignettes were relevant and credible. Finally, we selected the factors and their levels in such a manner that no illogical or implausible combinations of vignette factors could occur (Auspurg & Hinz, 2014).

By combing all vignette levels for the five factors (i.e., 2 x 32 x 4 x 4 x 4), 4096 unique vignettes could be created (i.e., the vignette universe). Because we aimed to have each vignette evaluated at least five times, as advocated by Auspurg and Hinz (2014), it was not feasible to have all 4096 vignettes evaluated. Doing so would require an immense sample or having each participant evaluate a massive number of vignettes which could cause fatigue among the respondents and lead to less qualitative data (Auspurg & Hinz, 2014). To deal with this problem, we chose to draw a sample of vignettes using a D-efficient design. A D-efficient design selects combinations of vignette levels with the most statistical power, leading to a more efficient experimental design. That is, by using a D-efficient design, we need fewer vignette judgements (i.e., vignettes per participant, participants, or both) to

12 Other research has worked with a limited number of age levels or age ranges (e.g., ‘64 to 66 years old’) (Büsch et al., 2009; Farber et al., 2016; Neumark et al., 2016, 2019).

13 The choice to vary the candidate characteristics over five vignette factors was made based on the recommendation of Auspurg and Hinz (2014) to work with vignettes of midlevel complexity, i.e., vignettes in which approximately seven (plus or minus two) vignette dimensions are varied. By using a midlevel number of vignette dimensions, we avoid participants to become overburdened by a too complex vignette design and, at the same time, assure participants to be stimulated enough not to drop out because of boredom or fatigue, which might happen with an overly simple design in which participants have to rate several very similar vignettes.

attain the same amount of statistical power as a less efficient design. More concretely, following Auspurg and Hinz's (2014) algorithm, we selected 200 different vignettes, which resulted in a considerably high D-efficiency of 99.11.[14, 15] After sampling the 200 vignettes, we blocked them into 40 decks of 5 vignettes, again using Auspurg and Hinz's (2014) algorithm.[16] To avoid order effects, we randomised the sequence of display of the five different vignettes within each deck. These 40 decks were then randomly assigned to the participants. This method allowed us to use a large overall number of vignettes resulting in a higher level of statistical (D-)efficiency of the design (Auspurg & Hinz, 2014).

2.2. Online Survey and Data Collection

The vignette experiment was implemented via an online survey administered in English and offered to the participants using Amazon Mechanical Turk (hereafter 'MTurk').[17] MTurk is an online crowdsourcing platform on which individuals can hire 'workers' to perform particular tasks in return for financial compensation. Prior academic research, within and outside economics, has shown that MTurk is a legitimate source from which to collect high-quality and reliable data (Buhrmester, Kwang, & Gosling, 2011; Rand, 2012; Goodman, Cryder, & Cheema, 2013; Roulin, 2015) and is particularly useful for online experimental studies (Paolacci et al., 2010; Horton, Rand, & Zeckhauser, 2011; Amir, Rand, & Gal, 2012; Chandler & Kapelner, 2013; Crump, McDonnell, & Gureckis, 2013; Kuziemko, Norton, Saez, & Stantcheva, 2015; Halberstam & Knight, 2016; Berggren, Jordahl, & Poutvaara, 2017; DellaVigna & Pope, 2018; Neyt, Vandenbulcke, & Baert, 2018).

The participants in our experiment had to meet two criteria. First, we decided to restrict ourselves to participants from OECD countries. The maximum number of countries we could select in MTurk was 30. Therefore, we selected participants from the 30 largest OECD countries (in terms of

[14] To select the 200 vignettes, we used the freeware macro %Mktex developed by Kuhfeld (2010). Taking into account the number of factors and associated levels, the parameters one tends to identify, and the number of vignettes one wants to use in the experiment, this algorithm first builds a set of potential designs and subsequently searches for the design with the highest D-efficiency (Auspurg &Hinz, 2014). For more information, we refer to Auspurg and Hinz (2014).

[15] A D-efficient design enhances statistical precision by maximising both orthogonality and level balance (i.e. equal frequencies of all levels). An experimental design has a sufficiently high D-efficiency when the D-efficiency exceeds 0.90 (Auspurg & Hinz, 2014).

[16] Auspurg and Hinz (2014) recommend using no more than ten vignettes per participant. We decided to present only five vignettes to each participant because we wanted to avoid the survey time to get too excessive. Indeed, our participants had to evaluate each candidate regarding a high number of statements (i.e., 16). Evaluating multiple vignettes regarding many criteria could result in a high survey time which might cause fatigue among the participants leading to a lower quality of the data or high drop-out rates.

[17] We provide an example of the entire online survey in the Online Appendix.

inhabitants).[18, 19] Second, the participants had to have experience in evaluating job candidates. To ensure that only individuals who were highly experienced in recruitment participated in the experiment, the participants were required to fill in two screening questions regarding their experience in evaluating job applicants at the beginning of the survey.[20] Only when the participants passed these two screening questions were they redirected to the survey. To guarantee that the participants filled out the online survey entirely and accurately, we included an attention check. Only the participants who answered the attention check correctly were able to complete the task and receive financial compensation.[21] Between July and August 2018, 400 participants filled out the survey completely and accurately, resulting in a total of 2000 observations.

At the beginning of the online experiment, the participants were informed about their task, i.e., to evaluate job candidates for a job vacancy at their (hypothetical) firm.[22] More concretely, they were required to evaluate candidates for one of the following positions: (i) dental technician, (ii) door-to-door sales worker, (iii) packer, (iv) CNC machine operator, (v) lab technician (cytogenetic techniques), (vi) insurance sales agent, (vii) physiotherapist, and (viii) database administrator. We selected these occupations as they varied over four different job characteristics, i.e., the degree of (i) overall skill required, (ii) customer contact, (iii) physical effort, and (iv) technological knowledge needed to perform the job well. We selected the jobs based on data from the Occupational Information Network (O*NET).[23] For an overview of the selection criteria and the corresponding jobs, see Table A–1 in Appendix A. The job descriptions presented to the participants were based on the descriptions found

[18] More concretely, we aimed to reach participants from the following countries: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Japan, Mexico, the Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, South Korea, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States. Estonia, Iceland, Latvia, Lithuania, Luxembourg, and New Zealand were not selected.

[19] Although we allowed people from 30 OECD countries to participate in our study, most people who eventually participated in our experiment came from the United States (see Section 2.4. for a description of our final sample).

[20] The screening questions consisted of answering 'yes' to the question 'In the context of your current profession, do you have experience evaluating job applicants?'(yes or no) and at least '3 times' to the question 'How often were you actively involved in evaluating job applicants for an open job vacancy in the last year?' (1 time, 2 times, 3 times, 4 times, 5 times or more).

[21] The attention check consisted of answering 'completely agree' when asked to do so. If the participants failed to provide the correct answer, they could not complete the task and were presented with a message in which they were told that they had failed the attention check.

[22] We decided to let the participants evaluate job candidates for a hypothetical firm instead of their own firm to ensure the internal validity of our experiment.

[23] O*Net is an online databank developed by the U.S. Department of Labor/Employment and Training Administration summarising occupational information on thousands of jobs (National Center for O*NET Development, 2019).

on O*NET and formulated as uniformly as possible to avoid any potential effects of these descriptions. An overview of the job descriptions can be found in Table A–2 in Appendix A. We assigned the distinct job openings randomly to the participants in such a way that all eight vacancies were presented with equal probability (and did not correlate with the deck of fictitious profiles assigned).

After viewing their assigned job descriptions, the participants were asked to fill in a comprehension check. We included the comprehension check to test whether the participants' perceptions about their assigned job characteristics matched the objective job characteristics found on O*NET (which was indeed the case). Next, the participants were told that the candidates (for whom the profiles could be found on the following screens) had been pre-screened and summarised in a tabular way by an administrative secretary and that all candidates were eligible for the job. Additionally, they were informed that they should evaluate all the profiles accurately and that they could jump between the different candidates and adjust their ratings as desired.

Once the participants finished reading the aforementioned instructions, they were shown the tabulated summaries of the fictitious job applicants' characteristics. The applicants' characteristics appeared in the same order as they would occur in real résumés, i.e., in the order used in Table 1. The participants then evaluated the applicants in terms of the probability that they would invite the person to a job interview (i.e., the interview probability scale, following Van Belle et al. (2018))[24] and, more importantly, 15 different statements related to the theories of statistical discrimination (Arrow, 1973) and taste-based discrimination (Becker, 1957) (hereafter 'the employers' perceptions of candidates'). For the theory of statistical discrimination, 12 statements were developed based on the literature review of Burn et al. (2019), each questioning a certain perception regarding older job candidates' (drivers of) productivity put forward in the literature.[25] More concretely, we adopted items concerning the applicants' perceived: (i) mental abilities, (ii) social abilities, (iii) physical abilities, (iv) technological knowledge and skills, (v) flexibility, (vi) creativity, (vii) experience, (viii) motivation,

[24] We opted to use the interview probability scale and not the hiring probability scale since the invitation decision mimics the first decision to be made in practice. Prior research has shown that this first decision very much determines employment opportunities and related hiring discrimination (Baert et al., 2016).

[25] Burn et al. (2019) aimed at identifying all the age stereotypes concerning workers in their 50s and 60s put forward in economics, industrial psychology, communications, and related literature. That is, older workers are thought to be perceived of as: (i) having less ability to learn, (ii) being less flexible, (iii) being less attractive, (iv) having poorer communication skills, (v) being less physically capable, (vi) being less productive, (vii) being worse with technology, (viii) being less creative, (ix) having a poorer memory, (x) being hard of hearing, (xi) having a negative personality, (xii) being less productive, (xiii) being dependable, (xiv) being careful, (xv) being more experienced, (xvi) having better communication skills, and (xvii) having a warmer personality.

(ix) reliability, (x) accuracy, (xi) trainability, and (xii) reasonability of wage expectations.[26] With respect to the experience item, it is important to stress that recruiters were evaluating an applicant having the required experience conditional on the level of a fictitious candidate's experience in the occupation shown in the vignette. In addition, with regards to the theory of taste-based discrimination, we employed the same three statements used in Baert and De Pauw (2014) and Van Borm and Baert (2018) to measure employer, employees, and customers' attitudes towards collaborating with older workers (as perceived by the employers). All 15 statements were rated on 7-point Likert scales ranging from 1 (i.e., 'completely disagree') to 7 (i.e., 'completely agree'). An overview of all employers' perceptions of candidates and their corresponding statements are presented in Table 2.

< Table 2 about here >

After evaluating the five profiles, the participants were asked to fill in a post-experimental survey, in which they were questioned about: (i) their experiences and feelings of competency concerning evaluating job applicants for the presented vacancy, (ii) their tendency towards answering in a socially desirable manner, (iii) four personal characteristics, and (iv) four characteristics of their current job. As mentioned previously, these items were added in view of robustness analyses and analyses capturing moderators of age discrimination on the employer side.

First, the participants' experiences and feelings of competency concerning evaluating job applicants for the presented vacancy were captured using five statements, each of which were rated on a 7-point Likert scale ranging from 1 (i.e., 'completely disagree') to 7 (i.e., 'completely agree'). Examples of these statements include: 'I have experience in recruiting candidates for jobs that require a high level of education' and 'I felt, from my professional experience, competent enough to select job candidates for the vacancy described'.

Second, the recruiters' tendencies to answer in a socially desirable way were measured using the 13-

[26] Unlike Burn et al. (2019), we decided not to include perceptions about (i) attractiveness, (ii) hear impairment, (iii) negative personality, and (iv) personal warmth in our experiment since these elements would have been difficult to evaluate given the experimental design of our study. Additionally, we decided not to investigate perceptions of the overall productivity of older workers since this signal is contained in all other age signals. Furthermore, we decided to include motivation as a potential signal of age because motivation has been found to be an important signal for, among others, long-term unemployment and people applying for a job under a vacancy referral scheme, i.e., two groups to which older individuals often belong (Van Belle et al., 2018; Van Belle et al., 2019). Next, we also chose to take into account the signal regarding the perceived cost of labour of older workers based on the input of various participants of the 2018 Belgian Day of Labour Economists to which we presented the results of our pilot experiment with Belgian recruiters (see Section 2.3). We changed the wording of some of the age signals to make sure they were easy to evaluate given the experimental design (e.g., we changed 'adaptable' to 'flexible' and 'dependable' to 'reliable').

item version of the Marlowe-Crowne Social Desirability Scale (MC-SDS) developed by Reynolds (1982). The scale consists of 13 items describing behaviour that is culturally approved or sanctioned (e.g., 'There have been occasions when I took advantage of someone') and is one of the instruments used most to measure social desirability (Beretvas, Meyers, & Leite, 2002; Sârbescu, Costea, & Rusu, 2011; Baert, 2018b). The participants answered the 13 items with 'true' when the statement applied to them or 'false' when it did not. The answers were then recoded so that socially desirable answers received a score of 1, and non-social desirable answers received a score of 0. Summing the scores for all items yielded a total score for answering in a socially desirable manner of between 0 and 13. We divided this number by 13 to obtain a proportion between 0 and 1.

Third, the participants were asked to report their demographic characteristics. That is, they were asked for their gender (man or woman), age, nationality, and highest educational degree (university education, higher education outside the university, secondary education, or lower than secondary education).

Fourth and last, the participants answered four questions about their current job. More concretely, they were surveyed concerning: (i) how often they were involved in evaluating job candidates in their current job (daily, weekly, biweekly, monthly, once per semester, once a year, or less frequently), (ii) how long they had been involved in evaluating job candidates (less than one year, 1–5 years, or more than 5 years), (iii) their type of job (manager, specialist in personnel and career development, employment agency employee, management assistant, general administrative assistant, or other), and (iv) the percentage of the workforce in their company aged 50 or older.

2.3. Pilot Study

To assess whether both the experiment and the post-experimental survey were clear and well-constructed, we ran a pilot study with 193 genuine Belgian recruiters mentioned in job vacancies located in the Public Employment Agency of Flanders database (32,787 vacancies were screened, and 2697 unique email addresses were identified and contacted directly). The results of this pilot with 965 (i.e., 193 × 5) candidate evaluations, which are available upon request, were presented and discussed thoroughly at the 2018 Belgian Day of Labour Economists, resulting in the addition of the item on the perceived cost of labour for older workers (see above).

2.4. Data Description

In Table 3, we present summary statistics concerning the participant characteristics (Panel A), the randomised jobs (Panel B), and the interview probability scale (Panel C) for the sample as a whole, as well as for two subsamples (i.e., the participants who evaluated fictitious applicants younger than the sample mean of 47.5, and participants who evaluated job applicants older than 47.5).

< Table 3 about here >

As shown in Panel A of column (1), out of our total sample of 400 participants 44.7% were female and about half of the participants were younger than 35 (53.2%). Additionally, as aforementioned, a majority of our participants came from the United States (89.7%).[27] Next to this, most of our participants had a university degree (70.0%) and were involved in evaluating job candidates at least once a semester (95.2%). Furthermore, 35.2% of the participants had been involved in evaluating job candidates for more than five years, and more than half of the participants (59.2%) were employed by a firm in which at least 20% of the workforce was 50 years old and above.

Although it is clear that our participants matched the target population of people with experience in hiring, we cannot claim that our sample is representative of the population of real-world HR professionals as a whole. To get an idea of the representativeness of our sample, we compare some descriptive statistics of our sample with the sample of employers in the American Community Survey (ACS) (see Table A-3 in Appendix). As becomes clear from Table A-3, our sample of HR professionals is on average younger (i.e., 36 years old versus 45 years old), more male (i.e., 45% female versus 67% female), but equally educated as the sample of HR professionals from the ACS sample. To investigate how our results might change if our sample was more representative of real-world employers, we run our mediation analysis for the subsample of older participants (i.e., participants older or as equally old as 35) and the subsample of female participants. We discuss these results in section 4.1.

Looking at Panel B of column (1), we can also see that the different vacancies were evaluated with about the same frequency.

From columns (2), (3), and (4), we can conclude that the randomisation of the candidate's age over the different participants in the experiment (Panel A) was successful. Candidates younger than 47.5 were evaluated by participants who were similar in terms of gender, age, nationality, educational level, experience in evaluating job candidates, and estimated percentage of older workers in their

[27] The other participants in our sample came from the United Kingdom (2.5%), Italy (1.7%), Turkey (1.7%), Canada (0.8%), Japan (0.8%), The Netherlands (0.8%), Germany (0.8%), Ireland (0.8%), Spain (0.8%), and Poland (0.8%).

firm compared to older job candidates. The same is true for the randomisation of the candidates' ages over the different job vacancies (Panel B of columns 2 to 4). About the same number of older and younger candidates were evaluated for each of the eight job vacancies.

We return to the results presented in Panel C of Table 3 in Section 4, where we discuss the effect of someone's age on her/his chance of being invited to a job interview.

3. Statistical Framework

Before discussing our results, we describe the statistical framework we used to analyse the data discussed in the previous section.[28] We start with a bivariate analysis. First, we explore the total effect of a person's age on their chances of being interviewed. Based on the results of previous research, we expect age to have a negative effect on hiring chances (Johnson & Neumark, 1997; Lahey, 2008; Farber, Silverman, & von Wachter, 2016; Neumark, Burn, & Button, 2016, 2019; Neumark, 2018). Second, we test the relationship between the applicants' age and participants' stereotypical beliefs involving older workers or attitudes towards them. As mentioned previously, prior research has shown that employers have many stereotypes about older workers' productivity (Gordon & Arvey, 2004; Posthuma & Campion, 2009; Richardson et al., 2013; Burn et al., 2019). Furthermore, negative attitudes towards them exist, which could influence the taste to collaborate with them of employers, employees, and customers (Kite & Johnson, 1988; Nelson, 2004). Based on these previous studies, we expect age to have negative effects on applicants' perceived: (i) mental abilities, (ii) social abilities, (iii) physical abilities, (iv) technological knowledge and skills, (v) flexibility, (vi) creativity, (vii) motivation, (viii) trainability, and (ix) reasonability with respect to wage expectations. Expectations with respect to (x) reliability, (xi) accuracy, (xii) and experience (see Section 2.2) are less clear-cut. We also expect to find a negative effect of older age on the attitude towards collaborating with these workers on the part of employers, co-workers, and customers. In statistical terms, correlation coefficients between the age of the candidate and the candidate evaluations are presented. In addition, we regress the standardised versions of the evaluation items on the age of the candidate.

Next, we examine what proportion of the age gap in the interview probabilities can be ascribed to the 15 employers' perceptions of candidates. We decompose the total effect into different indirect

[28] All analyses mentioned in this section are run using Stata. The codes used for the different analyses are available upon request.

effects via the signals and attitudes and a remaining 'direct' effect. To do so, we run a multiple mediation model in which all signals and attitudes related to older workers are included jointly, following a system of linear regression equations (following Hayes (2013)):[29]

$$M_1 = \alpha_{M_1} + \beta_{M_1} CC + \gamma_{M_1} PC + \delta_{M_1} JC + \theta_1 Age + \varepsilon_{M_1}; \quad (1)$$

$$M_2 = \alpha_{M_2} + \beta_{M_2} CC + \gamma_{M_2} PC + \delta_{M_2} JC + \theta_2 Age + \varepsilon_{M_2}; \quad (2)$$

$$M_3 = \alpha_{M_3} + \beta_{M_3} CC + \gamma_{M_3} PC + \delta_{M_3} JC + \theta_3 Age + \varepsilon_{M_3}; \quad (3)$$

…

$$M_{15} = \alpha_{M_{15}} + \beta_{M_{15}} CC + \gamma_{M_{15}} PC + \delta_{M_{15}} JC + \theta_{15} Age + \varepsilon_{M_{15}}; \quad (15)$$

$$Y = \alpha_Y + \beta_Y CC + \gamma_Y PC + \delta_Y JC + \theta' Age + \epsilon_i M_1 + \epsilon_2 M_2 + \cdots + \epsilon_{15} M_{15} + \varepsilon_Y. \quad (16)$$

In equations (1) to (15), the M_i are the items related to the 12 potential age signals and 3 types of attitudes towards collaborating with an older worker mentioned in Table 2. *Age* stands for the job candidates' age, and *CC* is a vector of the other candidate characteristics (i.e., vignette factors). Moreover, *PC* and *JC* are vectors of, respectively, the participant and job characteristics mentioned in Table 3 and Table A–1. Furthermore, β_{M_i}, γ_{M_i}, δ_{M_i}, and θ_i are the (vectors of) parameters associated with *CC*, *PC*, *JC,* and *Age*, respectively. The α_{M_i} are the intercepts of the equations. In equation (16), *Y* is the interview probability. Furthermore, β_Y, γ_Y, δ_Y, and α_Y in equation (16) are equivalent to the parameters used in the equations (1) to (15). Moreover, in equation (16), the ϵ_i are the parameters related to the mediator scales. Lastly, θ' is the remaining direct effect of the candidate's age after controlling for the mediators. As mentioned above, our main interest lies in the indirect associations between the candidate's age and the interview probability via each of the mediators (i.e., the products $\theta_i \epsilon_i$). Following Hayes (2013), we estimate all 16 equations simultaneously and correct the standard errors ε_{M_i} and ε_Y for the clustering of the observations at the participant level. While the coefficients δ_{M_i} can be given a causal interpretation, such is not the case for the coefficients ϵ_i—we will return to this point in Section 5.

Last, we investigate whether certain participant and job characteristics might moderate the level of age discrimination in hiring. In this respect, we investigate interactions between the fictitious

[29] We also ran an explorative factor analysis to see whether the different perceptions of candidates held by employers could be clustered into different latent factors. No clear and unambiguous latent factors were found.

candidate's age and the aforementioned vectors *PC* and *CC*. Based on previous research, we expect that older participants might treat older job applicants more favourably compared to younger applicants because they might identify more with job applicants of a similar age (i.e., in-group bias; Finkelstein et al., 1995; Gordon & Arvey, 2004; Posthuma & Campion, 2009; van Dalen, Henkens, & Schippers, 2009; Jensen, De Tavernier, & Nielsen, 2019). Moreover, we expect that people with a higher percentage of older employees in their firm might also rate older job applicants more favourably because having contact with older workers might lead participants to have more positive attitudes towards this group or believe to a lesser extent in the stereotypes that exist about them (i.e., 'in-group contact hypothesis', Allport, 1979; Henkens, 2005; Jensen, De Tavernier, & Nielsen, 2019). Finally, previous academic research has shown that age discrimination in hiring can vary across different types of jobs due to the existence of job stereotypes. Therefore, we expect to find higher levels of age discrimination in jobs that require: (i) high overall skills, (ii) a high level of customer contact, (iii) considerable physical efforts, and (iv) high technological knowledge and skills (Macan et al., 1994; Finkelstein et al., 1995; Perry et al., 1996; Gordon & Arvey, 1986; Perry & Bourhis, 1998; Perry & Finkelstein, 1999; Goldberg et al., 2004; Posthuma & Campion, 2009; Jensen, De Tavernier, & Nielsen, 2019).

To investigate the abovementioned possible moderation effects, we run a multivariate regression analysis. First, we run a baseline model following this linear regression equation:

$$Y = \alpha_Y + \theta_y Age + \beta_Y CC + \gamma_Y PC + \delta_Y JC + \varepsilon_Y. \tag{17}$$

Next, we add different interaction terms to the regression analysis between *Age* and *PC* and *JC*. The interactions with respect to *PC* cannot be given a causal interpretation, as they may correlate with other, unobserved participant characteristics that may also influence the hiring probability for older job candidates. Again, the error term is corrected for the clustering of the observations at the participant level (so that that heteroscedasticity related to our ordinal outcome variable is corrected for automatically as well).

By way of various robustness checks, we also ran all the statistical analyses discussed above for various subsamples. More concretely, we ran the analyses for a subsample of: (i) participants with a residence in the United States, (ii) participants with a lot of experience in evaluating job candidates, (iii) participants who indicated that they felt competent to evaluate job applicants for the presented vacancy, (iv) participants with a low tendency towards socially desirable answering, (v) participants older than (or equally old as or younger than) 35 years, and (vi) female participants. Additionally, we ran the bivariate and moderation analysis using an ordered logit model. The results of some of these robustness checks are presented and/or mentioned below—the other results are available upon request.

4. Results

4.1. Drivers of Age Discrimination

Table 4 presents the results of the bivariate analysis described in Section 3. Panel B of this table corroborates the literature employing field experiments to measure age discrimination. That is, we find a highly significantly negative correlation between a candidate's age and their interview probability. This is also consistent with Panel C of Table 3, which indicates that the average rating on the interview probability scale is significantly higher for candidates younger than the sample mean than for older candidates. In addition, Figure 1, which depicts the average scores on the interview scale of the 2000 evaluated vignettes by the age of the fictitious candidate, is consistent with this evidence.

<Figure 1 about here >

< Table 4 about here >

More importantly, we find highly significantly negative correlations between the candidate's age and ten of the age signals (i.e., perceived social abilities, perceived physical abilities, perceived technological knowledge and skills, perceived flexibility, perceived creativity, perceived motivation, perceived reliability, perceived accuracy, perceived trainability, and perceived reasonability with respect to wage expectations). The highest correlations are found between the applicant's age and perceived physical abilities (i.e., −0.233), perceived trainability (−0.183), perceived flexibility (−0.145), and perceived technological knowledge and skills (−0.113). Correlations between the candidate's age and perceptions concerning their mental abilities and experience are weakly significant or not significant at all. Moreover, we also find a highly significant negative correlations between the candidate's age and the attitudes towards collaborating with this individual on the part of employers, employees, and customers (i.e., −0.099, −0.106, and −0.100, respectively). Bivariate regression analyses yield the same conclusions. For instance, we find that one additional year yields 2.6%, 2.0%, 1.6%, and 1.2% of a standard deviation lower scores on perceived physical abilities, trainability, flexibility, and perceived technological knowledge and skills, respectively.

Table 5 presents the results of the mediation analysis determining how much the individual signals contribute to the total age gap in the interview probability. Following Heckman, Pinto, and Savelyev (2013), we present the results as percentages of the total age effect explained by the 15 mediators.[30] Looking at the results for our total sample in column (1), we find there are three highly significant mediation effects.[31] First, we find a highly significant mediation effect of applicants' perceived technological knowledge and skills. That is, about 18% of the total age effect (with respect to the invitation probability) is explained by the perception of lower technological knowledge and skills. Additionally, we identify highly significant mediation effects of the applicants' perceived trainability and flexibility. Respectively, 12% and 11% of the total age effect is explained by these mediators. So, these three dominant stigma jointly explain about 41% of the total age effect. We return to the policy consequences of this finding in Section 5.

< Table 5 about here >

In addition, we find a significant, but less pronounced, mediating role for perceived mental abilities and perceived reasonability with respect to wage expectations (both explaining about 3% of the total age effect). Last, we find a highly significant mediation effect related to perceived experience. At first sight it might be surprising that this mediation effect has a positive sign, but, as mentioned in Section 2.2, it should be taken into account that this item received ratings that were conditional on the given candidate's experience in the occupation. Therefore, a greater age might reflect the negative signal of many years of irrelevant experience (and, as a consequence, a lower overall score with respect to experience relevant to performing well in the job).[32]

As a robustness check, we rerun this analysis for four substantial, homogeneous subsamples: participants with a residence in the United States (column 2), participants who evaluate job

[30] We, thus, divided the 15 indirect effects via the signals and attitudes on one's interview probability by the point estimate of the total (negative) effect of the candidates' age. As a result, the percentages with a positive sign found in Table 5 should be interpreted as negative mediation effects (e.g., the percentage related to perceived mental abilities (i.e., 3%) means that employers' perception that older job candidates have lower mental abilities explains 3% of the total negative effect of age on one's hiring chances) and the percentages with a negative sign as a positive mediation effect. Additionally, we constructed the 95% confidence intervals by dividing the lower and upper bounds of the bootstrapped 95% confidence intervals of the 15 mediation effects by the point estimate of the total (negative) effect of age on the candidates' interview probability.

[31] In this section, we speak of mediation 'effects' following the literature on mediation analysis. As mentioned previously, we are aware, however, that we cannot give these mediation effects a causal interpretation since the mediators are not exogenous. It is possible that our mediators still correlate with other unobserved employer perceptions and attitudes related to age. For this reason, the indirect 'effects' of the age signals and attitudes should be seen as associations rather than causal effects. We return to this point in Section 5.

[32] This significantly positive mediation effect with respect to perceived experience was also found in our pilot sample with 193 Belgian recruiters contacted via direct e-mail (see Section 2.2).

candidates at least once a semester (column 3), participants who felt competent to assess job candidates for the presented vacancy (i.e., participants who indicated at least '5' on the 7-point Likert Scale; column 4), and participants with a low tendency towards socially desirable answering (i.e., a score on the social desirability scale below the sample mean increased by one standard deviation; column 5). However, we found results comparable to those discussed above for all these subsamples. More concretely, for all three subsamples, the three dominant mediation effects remain those concerning (i) the perceived technological knowledge and skills, (ii) flexibility, and (iii) trainability of older job applicants. The only (slight) divergences occur among the sample of American participants, where the mediation effect related to perceived labour costs is only significant at the 10% significance level, and the sample of participants who felt competent to evaluate job candidates for the vacancy described, where the mediation effect regarding the experience of older job applicants became significant only at the 10% significance level and the mediation effect concerning the reliability of the job applicants became significant at the 5% significance level. It may not come as a surprise that the results regarding the American sample do not differ much from the total sample, however, since almost 90% of our sample comes from the United States.In Table A–4 in Appendix A, we replicate our mediation analysis after breaking down our sample by the gender of the fictitious candidates. Although the same dominant mediators are found for both genders, the mediation effects with respect to technological ability and flexibility are somewhat more prominent in the female subsample, while the mediation effect related to perceived trainability is more noticeable in the subsample of male candidates. Moreover, the mediation effect related to experience discussed above is driven by the male subsample.

Moreover, to have an idea of how our results would change if our sample would be more representative of real-life recruiters, we replicate our mediation analysis for the subgroup of participants who are older or as equally old as 35 (compared to the subgroup of participants who are younger than 35), as well as for the subgroup of female participants (compared to the subgroup of male participants) (see Table A–5 in Appendix A). As becomes clear from Table A-5., we find quite some similarities in the mediation results between the different subgroups, especially between the male and female subgroup. More concretely, we find that the mediation effect regarding the perceived flexibility of older job applicants is equally prominent in all subgroups. Moreover, the perceptions regarding the technological know-how and trainability of older job applicants seem to influence the hiring decisions of both male and female HR professionals in comparable ways. Despite these similarities, we also find some discrepancies in the mediation effects between the different subsamples. The mediation effect related to perceived trainability is driven by the older subsample, while the mediation effect regarding the perceived technological know-how is more pronounced in the younger subsample. Furthermore, the mediation effect regarding the perceived experience of older job applicants seems more noticeable in the male subgroup. These results, therefore, suggest that the perception regarding the trainability of older job applicants possibly influences HR professionals' hiring decisions more in the real-world than our initial results suggest and that the perceived experience of these job applicants might have less influence.

The mediation effects based on our data gathered via MTurk are, to a large extent, in line with the corresponding results obtained via the pilot sample of 193 Belgian recruiters (see Section 2.3). In particular, also within this sample, lower technological ability and flexibility were the two most dominant stigma mediating unfavourable interview decisions with respect to older job candidates, with a less prominent role for perceived trainability.

In conclusion, our results are in stark contrast to those of Richardson et al. (2013), who found no mediation effects.[33] At the full sample level, about 65% of the total age effect is explained by the mediators. However, a significant amount (i.e., about 35%) of the total effect is, therefore, not explained by our model, meaning that, although we attempted to capture the most relevant signals potentially explaining the lower hiring chances of older job applicants based on Burn et al.'s (2019) literature review, we were still not able to capture them all. A potential reason for this result is the fact that, given our experimental design, we were not able to investigate the signals regarding older employees' attractiveness, personality, and hearing impairments, all items mentioned by Burn et al. (2019)—see Section 2.2. Another explanation for the remaining significant direct effect might be the imprecise measurement of the different candidate evaluations. Measurement errors for the mediators could have resulted in downward-biased estimates for the mediation effects and an upward-biased estimate for the remaining direct effect (Judd & Kenny, 1981; VanderWeele, Valeri, & Ogburn, 2012). Looking at the 95% confidence intervals of the 15 mediation effects, it becomes clear that the average estimates of the mediation effects hide a decent amount of uncertainty.[34] Therefore, it seems credible to say that the other factors we are not accounting for and some additional uncertainty explains the missing share of discrimination.

[33] This divergence in results might be explained by the smaller scale and the different set-up of the experiment in Richardson et al. (2013), which was mentioned in Section 1.

[34] Adding up the upper limits of the 95% confidence intervals of the 15 individual mediation effect in Table 5 conveys that we can explain up to 143% of the variation in the most extreme case. This number exceeding 100% should not make us worry because it is very unlikely that the estimates of the mediation effects actually will be at the limit for every single factor.

4.2. Moderators of Age Discrimination

As mentioned in Section 3, we run a multivariate regression analysis to investigate whether certain participant and job characteristics might have an effect on the degree of age discrimination in hiring. Table 6 reports the results from this analysis.

< Table 6 about here >

First, in model (1), we estimate a baseline model in which we regress all candidate, participant, and job characteristics on the interview probability without including any interaction terms. We find (highly) significant effects of all candidate characteristics, with the exception of gender, on the probability of being invited to a job interview. In addition to the aforementioned age effect, we identify the positive effect of a limited commuting distance, two to ten years of experience, and the extracurricular activities mentioned on the résumé on the probability of being invited for a job interview. In addition, we find that, on average, younger participants give higher scores on the interview probability scale compared to participants 35 or over and participants who have a high percentage of older employees working at their firm. We find no differences in rating in terms of the four job dimensions.

Second, and more importantly, we estimate, in models (2), (3), and (4), the same baseline model, while including different sets of interaction terms. In model (2), we include interaction terms between the candidate's age and the different participant characteristics. We observe a significantly positive interaction effect[35] between the candidate's age and the percentage of older employees working at the participant's firm. This lower level of age discrimination among recruiters working for firms with a substantial number of older employees is in line with Allport's (1979) in-group contact hypothesis. In the regression models for which the results are presented in the next columns, we include interaction terms between the candidate's age and the four job characteristics (model (3))[36] as well as interactions with seven occupation indicators (model (4)). We find no significant interaction terms between the candidate's age and the various job characteristics and functions, with the exception of discovering that the unfavourable treatment towards older applicants is lower in jobs associated with high levels of required skills. Lastly, in column 5, we include the interaction terms concerning the

[35] As mentioned in Section 3, this interaction effect cannot be given a causal interpretation.

[36] In the context of a robustness check, we reran this analysis including interaction terms between the candidate's age and the comprehension check concerning the participant's perceptions on the characteristics of these jobs mentioned in Subsection 2.1. Results are available upon request.

participant and job characteristics jointly and find results similar to those found in columns 2 and 3.

5. Conclusion

To investigate the potential drivers and moderators of age discrimination in hiring, we conducted a vignette experiment in which genuine recruiters were asked to make fictitious hiring decisions regarding job applicants of different ages (ranging from 32 to 63 years old) for one out of eight job vacancies. Participants evaluated the applicants concerning 15 statements related to all dominant explanations for hiring discrimination towards older applicants found in the scientific literature. We found that older age signals lower social and physical abilities, motivation, and technological knowledge and skills. Additionally, the results showed that older age is associated with lower levels of flexibility, creativity, reliability, trainability, and higher costs of labour. We thus found clear evidence for the existence of most of the signals described in the literature (Burn et al., 2019). Moreover, our results suggest that statistical discrimination in hiring, as argued by Arrow (1973), is the main cause of the age gap in hiring probabilities. Indeed, we find that the applicant's perceived technological knowledge and skills, flexibility, and trainability explain 41% of the total effect of age on a job applicant's interview chances. There is little evidence that individual distaste on the part of employers, co-workers, and customers to collaborate with older workers contributes in a meaningful way to the gap. Finally, our analysis showed that the negative association between age and invitation probability is less prominent among recruiters working for firms with a higher percentage of older employees.

From a policy perspective, the solution to the statistical discrimination found in this study might be to provide employers with more candidate information. In particular, older workers might reduce their chances of being discriminated against by highlighting their flexibility and technological skills in their résumés. Additionally, policymakers seeking to help unemployed workers find a job may wish to offer these workers chances to gain the technological skills needed in the modern labour market (and to reduce the related stigma via awareness campaigns). These training programs may also signal to employers that these workers are willing and able to undergo training and adaptable to changing work requirements.

Our vignette experiment design does not come without limitations. First, although the estimated effect of an applicant's age on the tested employers' perceptions of candidates can be given a causal interpretation, the same does not hold for the estimated association of these perceptions of the candidates held by employers with the interview probability. Although we attempted to capture, based on the systematic literature review of Burn et al. (2019), the most relevant signals of age potentially explaining the lower hiring chances for older job applicants, it is still possible that they correlate with other unobserved prejudices. To measure causal mediation effects, we would have to experimentally manipulate the employers' perceptions of candidates separately, however, within our context, this was not feasible. Future research should, therefore, focus on experimentally manipulating the different age signals to detect any causal effect of the perceptions of candidates held by employers on the hiring chances as is done by Lahey (2008).

Second, our research is limited by its laboratory setting. In contrast to field experiments, lab experiments do not take place under real-life circumstances. In other words, participants are aware they are participating in an experiment and that their answers have no real-world consequences. Although a lab experiment is advantageous from a research-ethical point of view (Riach & Rich, 2004; Charness, Gneezy, & Kuhn, 2013) and essential for obtaining deeper insights into thought processes (Van Hoye & Lievens, 2003; Baert & De Pauw, 2014; Van Belle et al., 2018), it could induce a certain degree of hypothetical bias. That is, participants might behave differently in our experiment than in real life, for example, because they do not take the survey seriously or try to hide the fact that they are inclined to discriminate (i.e., social desirability bias).

In our study, we aimed to minimise this hypothetical bias in three ways. To make sure the participants in our experiment did not behave very differently than in real life, we designed our experiment to mimic real-life hiring decisions as closely as possible. By simultaneously manipulating different applicant characteristics, we aimed to imitate the complex nature of hiring decisions in the field, where HR managers and employers are also confronted with the evaluation of job applicants differing in several personal characteristics such as gender, educational level, and work experience (Shadish, Cook, & Campbell, 2002; Colquitt, 2008; Baert & De Pauw, 2014). Indeed, research has shown that the decisions made in vignette experiments are highly correlated with actual behaviour (Baert & De Pauw, 2014; Hainmueller, Hangartner, & Yamamoto, 2015; Van Belle et al., 2018). Therefore, this bias seems to play a lesser role in vignette experiments in general and in ours, in particular. Additionally, to assure the participants in our sample took the survey seriously, we screened all completed surveys on the quality of the data and retained only the participants who filled out the survey completely and accurately. For example, participants who filled out the survey in an extremely short amount of time, failed the attention check(s), or whom's answers were clearly of a low quality (i.e., all indications that the participants did not take the survey seriously) were excluded from our final sample. Lastly, we controlled for socially desirable answering by including a social desirability scale in our post-experimental survey and rerunning the analyses for the subset of participants with a low score on this scale. Because the results of the subgroup of participants with low scores on the social desirability scale are very much in line with the results we found for the total sample, we believe that social desirability bias plays a lesser role in our experiment. This result can be explained by the fact that the participants in our study were each presented with a limited number of vignettes varying over multiple factors, which made it impossible for the participants to know the experiment was about age and, therefore, to identify socially desirable answers (Mutz, 2011; Auspurg & Hinz, 2014).

Third, we asked the participants in our experiment to evaluate job candidates in a wide range of fields, which may not correspond to the types of worker they typically hire. Although this increases the external validity of our results and allows us to conduct a thorough heterogeneity analysis based on job characteristics, it might cause participants to answer differently from when they would make hiring decisions regarding job vacancies they are more familiar with. As a result, our results might be biased. Indeed, even though we aimed to limit this bias in two ways – that is, we provided the participants in our study with a detailed O*Net description of the occupations and conducted our experiment with real HR professionals who are all experienced in evaluating job candidates more generally – we find that our results for the total sample (slightly) differ compared to the results for the sample of participants who indicated to feel highly competent to evaluate job candidates for the vacancy they were presented with. We, thus, cannot claim that the way in which we distributed the fictitious vacancies over the participants does not influence our results. Future researchers should, therefore, keep this issue in mind when designing a study and should construct their experimental design in such a way as to allow them to test the robustness of their method to this potential bias. One way to do so is to let participants choose the vacancy for which they need to evaluate job applicants, adding an industry-specific treatment arm is another.

Fourth, to improve the scale and external validity of our research, we decided to work with tabular vignettes instead of text vignettes. Indeed, tabular vignettes' more straightforward presentations of dimensions can help participants form more consistent judgements in a short response time (Auspurg & Hinz, 2014), allowing us to collect more data in a relatively short amount of time. However, working with tabular vignettes might have some consequences. That is, by working with tabular vignettes, the five candidate characteristics are more visible to employers than these factors would be in a full job application (i.e., a cover letter and a CV). This raises the concern that the recruiters may respond differently than they would in a typical hiring situation, which could potentially affect the results. Because we did not vary the presentation form of our vignettes, we cannot assure that our choice for a tabular vignette does not influence our results. Although previous research has found that working with tabular vignettes produce similar evaluations to text vignettes (Auspurg & Hinz, 2014; Shamon, Dülmer, & Giza, 2019; Sauer, Auspurg, & Hinz, 2020), future research should experiment with the presentation form of vignettes to investigate whether tabular vignettes indeed form a good alternative for text vignettes in heterogeneous samples.

Fifth, although we diversified our sample considerably compared to Richardson et al. (2013), who mainly used students as participants, our sample of MTurk employers is not representative of real-life HR professionals. More concretely, our sample contains more men and younger people compared to the employers of the American Community Survey. Therefore, our results are not generalisable to the whole population of real-world HR professionals. To get some insight on how our results would change if our sample was more representative of real-life employers, we reran our mediation analysis for both the subsample of older and female participants. The results of this analysis suggest that the mediating effect of perceived trainability would possibly be more pronounced in real life, while the mediating effect of perceived experience potentially would have less influence. Although our efforts to deal with the issue of representativeness, future research should replicate the study with a more representative sample of HR professionals.

Sixth, although we attempted to increase the generalisability of our results by having participants from several OECD countries evaluate job candidates with different profiles for one out of eight vacancies from different sectors and varying over four different job characteristics, our results can still not be easily generalised to other contexts. It could be the case that the stigmas related to older age (as well as the extent to which older workers face discrimination) could be different in various types of jobs. Additionally, there may be variations in age stigmas (and the level of discrimination experienced by older job applicants) across countries because the prevalence of various age stereotypes might differ over different countries. Although we allowed individuals from 30 different OECD countries to participate in our study, most of our participants turned out to be from the United States. Consequentially, we were unable to conduct a thorough heterogeneity analysis by country, other than comparing our total sample with the American sample. Specific heterogeneities might, therefore, not be identified. However, we believe future research is needed to thoroughly identify the heterogeneities in the prevalence of different age stereotypes in various countries and settings.

Seventh, we focused on job applicants between the ages of 32 and 63. It might be interesting, however, to also investigate the stigmas and attitudes related to job applicants who exceed the age of 63. In a society where people often extend their work lives beyond the official retirement age, as is the case in the United States, it might be interesting to investigate how the different stigmas and attitudes related to age might vary between job applicants younger and as equally old as the official retirement age and job applicants older than the official retirement age.

Last, we only considered explicit age cues (i.e., an applicant's age) in our research. Potential implicit age cues, such as certain extracurricular activities or more old fashioned names, were thereby ignored (Derous & Decoster, 2017). To develop adequate policy actions, it is, however, also important to gain deeper insights into these implicit age cues. Derous and Decoster (2017), for example, found that implicit age cues could compromise the effectiveness of anonymous application procedures (i.e., procedures where non-job-related personal identifiers are not revealed on résumés to avoid discrimination in hiring based on these identifiers). Future research should, therefore, consider implicit age cues mentioned on résumés and investigate to what extent they are related to stigma other than those brought about by the explicit age cues on which we focused in this study.

References

AARP. (2000). *American business and older employees*. Washington, DC: AARP.

Adams, S. J. (2002). Passed over for promotion because of age: An empirical analysis of the consequences. *Journal of Labor Research*, *23*(3), 447–461.

Ahmed A., Andersson, L., & Hammarstedt, M. (2012). Does age matter for employability? A field experiment of ageism in the Swedish labour market. *Applied Economics Letters, 19*(4)*,* 403–406.

Allport, G. W. (1979). *The nature of prejudice.* Cambridge, MA: Perseus Books.

Ambuehl, S., & Ockenfels, A. (2017). The ethics of incentivizing the uninformed: A vignette study. *American Economic Review,* 107, 91–95.

Amir, O., Rand, D. G., & Gal, Y. K. (2012). Economic games on the Internet: The effect of $1 stakes. *Plos One*, *7*(2), e31461.

Arrow, K. J. (1973). The theory of discrimination. In O. Ashenfelter & A. Rees (Eds.), *Discrimination in labor markets.* Princeton, NJ: Princeton University Press.

Attanasio, O., Kitao, S., & Violante, G. L. (2007). Global demographic trends and social security reform. *Journal of Monetary Economics*, *54*(1), 144–198.

Auspurg, K., & Hinz, T. (2014). *Factorial survey experiments*. Thousand Oaks, CA: Sage.

Auspurg, K., Hinz, T., & Sauer, C. (2017). Why should women get less? Evidence on the gender pay gap from multifactorial survey experiments. *American Sociological Review*, *82*, 179–210.

Baert, S. (2018a). Hiring discrimination: An overview of (almost) all correspondence experiments since 2005. In S. Gaddis (Ed.), *Audit studies: Behind the scenes with theory, method, and nuance*. Methodos Series (Methodological Prospects in the Social Sciences), vol 14. Cham: Springer.

Baert, S. (2018b). Hiring a gay man, taking a risk?: A lab experiment on employment discrimination and risk aversion. *Journal of Homosexuality*, *65*(8), 1015–1031.

Baert, S., & De Pauw, A.-S. (2014), Is ethnic discrimination due to distaste or statistics?. *Economics Letters*, *125*(2), 270–273.

Baert, S., Norga, J., Thuy, Y., & Van Hecke, M. (2016). Getting grey hairs in the labour market. An alternative experiment on age discrimination. *Journal of Economic Psychology*, *57*, 86–101.

Barr, N. (2006). Pensions: Overview of the issues. *Oxford Review of Economic Policy*, *22*(1), 1–14.

Becker, G. S. (1957). *The economics of discrimination.* Chicago, IL: University of Chicago Press.

Beretvas, S .N., Meyers, J .L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne social desirability scale. *Educational and Psychological Measurements*, *62*, 570–589.

Berggren, N., Jordahl, H., & Poutvaara, P. (2017). The right look: Conservative politicians look better and voters reward it. *Journal of Public Economics*, *146*, 79–86.

Bhargava, S., Loewenstein, G., & Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated option. *Quarterly Journal of Economics, 132*(3), 1319–1372.

Blanchard, O. J., & Diamond, P. (1994). Ranking, unemployment duration, and wages. *The Review of Economic Studies*, *61*(3), 417–434.

Boissonneault, M., Mulders, J. O., Turek, K., & Carriere, Y. (2020). A systematic review of causes of recent increases in ages of labor market exit in OECD countries. *PloS one*, *15*(4), e0231897.

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science, 8*(4), 445–454.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. *Quarterly Journal of Economics, 131*(4), 1753–1794.

Breyer, F., & Kifmann, M. (2002). Incentives to retire later–a solution to the social security crisis?. *Journal of Pension Economics & Finance*, *1*(2), 111–130.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?. *Perspectives on Psychological Science, 6*(1), 3–5.

Burn, I., Button, P., Corella, L. M., & Neumark, D. (2019). Older workers need not to apply? Ageist language in job ads and age discrimination in hiring. *Unpublished manuscript.*

Büsch, V., Dahl, S. A., & Dittrich, D. A. (2009). An empirical study of age discrimination in Norway and

Germany. *Applied Economics*, *41*(5), 633–651.

Butler, R.N. (1969). Age-ism: Another form of bigotry. *Gerontologist*, *9*, 243–246.

Carlsson, M., & Eriksson, S. (2019). Age discrimination in hiring decisions: Evidence from a field experiment in the labor market. *Labour Economics*, *59*, 173–183.

Carlsson, M., Reshid, A. A., & Rooth, D. O. (2018). Neighborhood signaling effects, commuting time, and employment: Evidence from a field experiment. *International Journal of Manpower*, *39*(4), 534–549.

Cavalier Jr, J., Hampton, S. B., Langford, R., Symes, L., & Young, A. (2018). The influence of race and gender on nursing care decisions: a pain management intervention. *Pain Management Nursing*, *19*(3), 238–245.

Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing. *Journal of Economic Behavior & Organization*, *90*, 123–133.

Charness, G., Gneezy, U., & Kuhn, M. A. (2013). Experimental methods: Extra laboratory experiments-extending the reach of experimental economics. *Journal of Economic Behavior and Organization, 91*, 93–100.

Colquitt, J.A. (2008). From the editors publishing laboratory research in AMJ: A question of when, not if. *Academy of Management Journal, 51*(2008), 616–620.

Conley, T. D., Matsick, J. L., Moors, A. C., & Ziegler, A. (2017). Investigation of consensually nonmonogamous relationships: Theories, methods, and new directions. *Perspectives on Psychological Science, 12*(2), 205–232.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *Plos One*, *8*(3), e57410.

Damelang, A., & Abraham, M. (2016). You can take some of it with you!. *Zeitschrift für Soziologie*, 45, 91–106.

Damelang, A., Abraham, M., Ebensperger, S., & Stumpf, F. (2019). The hiring prospects of foreign-educated immigrants: A factorial survey among German employers. *Work, Employment and Society*. Retrieved from https://doi.org/10.1177/0950017018809897.

Deelen, A., de Graaf-Zijl, M., & van den Berge, W. (2018). Labour market effects of job displacement for prime-age and older workers. *IZA Journal of Labor Economics*, 7(1), 3.

DellaVigna, S., & Pope, D. (2018). Predicting experimental results: Who knows what? *Journal of Political Economy*, *126*(6), 2410–2456.

Derous, E., Nguyen, H. H., & Ryan, A. M. (2009). Hiring discrimination against Arab minorities: Interactions between prejudice and job characteristics. *Human Performance*, *22*(4), 297–320.

Derous, E., Ryan, A. M., & Nguyen, H. H. D. (2012). Multiple categorization in resume screening: Examining effects on hiring discrimination against Arab applicants in field and lab settings. *Journal of Organizational Behavior*, *33*(4), 544–570.

Di Stasio, V. (2014). Education as a signal of trainability: Results from a vignette study with Italian employers. *European Sociological Review*, *30*(6), 796–809.

Drydakis, N., MacDonald, P., Bozani, V., & Chiotis, V. (2017). Inclusive recruitment? Hiring discrimination against older workers. In A. Arenas, D. Di Marco, L. Munduate, & M. C. Euwema (Eds). *Shaping inclusive workplaces through social dialogue*. Cham: Springer.

Eriksson, S. & Rooth, D. O. (2014). Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment. *American Economic Review, 10*, 1014–1039.

Eriksson, T., & Kristensen, N. (2014). Wages or fringes? Some evidence on trade-offs and sorting. *Journal of Labor Economics*, *32*(4), 899–928.

Farber, H. S., Silverman, D., & Von Wachter, T. (2016). Determinants of callbacks to job applications: An audit study. *American Economic Review*, *106*(5), 314–18.

Finkelstein, L. M., Burke, M. J., & Raju, M. S. (1995). Age discrimination in simulated employment contexts: An integrative analysis. *Journal of Applied Psychology*, *80*(6), 652–663.

French, M. T., Mortensen, K., & Timming, A. R. (2019). Are tattoos associated with employment and wage discrimination? Analyzing the relationships between body art and labor market outcomes. *Human Relations*, *72*(5), 962–987.

Ginn, J., & Arber, S. (1996). Gender, age and attitudes to retirement in mid-life. *Ageing & Society*, *16*(1), 27–55.

Goldberg, C. B., Finkelstein, L. M., Perry, E. L., & Konrad, A. M. (2004). Job and industry fit: The effects of age and gender matches on career progress outcomes. *Journal of Organizational Behavior, 25*(7), 807–829.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making, 26*(3), 213–224.

Gordon, R. A., & Arvey, R. D. (1986). Perceived and actual ages of workers. *Journal of Vocational Behavior*, *28*(1), 21–28.

Gordon, R. A., & Arvey, R. D. (2004). Age bias in laboratory and field settings: A meta-analytic investigation. *Journal of Applied Social Psychology*, *34*(3), 468–492.

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behaviour. *Proceedings of the National Academy of Sciences*, *112*, 2395–2400.

Halberstam, Y., & Knight, B. (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, 143, 73–88.

Harkin, C. T. (2012). *The Retirement crisis and a plan to solve it*. Washington, DC: US Senate Committee on Health, Education, Labor and Pensions.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* New York, NY: Guilford Press.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*(6), 2052–2086.

Henkens, K. (2005). Stereotyping older workers and retirement: The managers' point of view. *Canadian Journal on Aging/La Revue Canadienne du Vieillissement*, *24*(4), 353–366.

Hicks, P. (2012). Later retirement: The win-win solution. *CD Howe Institute Commentary*, 345.

Horton, J. J., Rand, D. G., & Zeckhauser, R. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*(3), 399–425.

Hummert, M. L., Garstka, T. A., Shaner, J. L., & Strahm, S. (1994). Stereotypes of the elderly held by young, middle-aged, and elderly adults. *Journal of Gerontology*, *49*(5), 240–249.

Jasso, G. (2006). Factorial survey methods for studying beliefs and judgments. *Sociological Methods and Research, 34*(3), 334–423.

Jensen, P. H., De Tavernier, W., & Nielsen, P. (2019). To what extent are ageist attitudes among employers translated into discriminatory practices? *International Journal of Manpower.* doi: 10.1108/IJM-10-2018-0365.

Johnson, R. W., & Neumark, D. (1997). Age discrimination, job separations, and employment status of older workers. *Journal of Human Resources, 32*(4), 779–811.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, *5*, 602–619.

Kitao, S. (2014). Sustainable social security: Four options. *Review of Economic Dynamics*, *17*(4), 756–779.

Kite, M. E., & Johnson, B. T. (1988). Attitudes toward older and younger adults: A meta-analysis. *Psychology and Aging, 3(*3), 233–244.

Kontis, V., Bennett, J. E., Mathers, C. D., Li, G., Foreman, K., & Ezzati, M. (2017). Future life expectancy in 35 industrialised countries: Projections with a Bayesian model ensemble. *The Lancet*, *389*(10076), 1323–1335.

Kroft, K., Lange, F. & Notowidigdo, M. J. (2013). Duration dependence and labour market conditions: Evidence from a field experiment. *Quarterly Journal of Economics, 128*, 1123–1167.

Kroon, A. C., Van Selm, M., ter Hoeven, C. L., & Vliegenthart, R. (2018). Reliable and unproductive? Stereotypes of older employees in corporate and news media. *Ageing & Society*, *38*(1), 166–191.

Kuhfeld, W. F. (2010). *Statistical graphics in SAS: An introduction to the graph template language and the statistical graphics procedures*. SAS Publishing.

Kuziemko, I., Norton, M. I., Saez, E., & Stantcheva, S. (2015). How elastic are preferences for redistribution? Evidence from randomized survey experiments. *American Economic Review*, *105*(4), 1478–1508.

Lahey, J. N. (2008). Age, women, and hiring an experimental study. *Journal of Human*

Resources, *43*(1), 30–56.

Lahey, J. N., & Oxley, D. R. (2018). Discrimination at the intersection of age, race, and gender: Evidence from a lab-in-the-field experiment. *National Bureau of Economic Research, no. w25357*.

Loretto, W., & White, P. (2006). Employers' attitudes, practices and policies towards older workers. *Human Resource Management Journal*, *16*(3), 313–330.

Macan, T. H., Detjen, J. B., & Dickey, K. L. (1994). Measures of job perceptions: Gender and age of current incumbents, suitability, and job attributes. *Sex Roles*, *30*(1–2), 55–67.

Maestas, N., & Zissimopoulos, J. (2010). How longer work lives ease the crunch of population aging. *Journal of Economic Perspectives*, *24*(1), 139–60.

Maslyn, J. M., Farmer, S. M., & Bettenhausen, K. L. (2017). When organizational politics matters: The effects of the perceived frequency and distance of experienced politics. *Human Relations*, *70*(12), 1486–1513.

Mathew, S. (2017). How the second-order free rider problem is solved in a small-scale society. *American Economic Review*, 107, 578–581.

McCann, R. M., & Keaton, S. A. (2013). A cross cultural investigation of age stereotypes and communication perceptions of older and younger workers in the USA and Thailand. *Educational Gerontology*, *39*(5), 326–341.

McGrattan, E. R., & Prescott, E. C. (2017). On financing retirement with an aging population. *Quantitative Economics*, *8*(1), 75–115.

Montiel Olea, J. L., & Strzalecki, T. (2014). Axiomatization and measurement of quasi-hyperbolic discounting. *Quarterly Journal of Economics, 129*(3), 1449–1499.

Moscarini, G. (1997). Unobserved heterogeneity and unemployment Duration: A fallacy of composition. *Unpublished manuscript.*

Munnell, A. H., & Sass, S. A. (2009). *Working longer: The solution to the retirement income challenge*. Washington, DC: Brookings Institution Press.

Mutz, D. C. (2011). *Population-based survey experiments*. Princeton, NJ: Princeton University Press.

National Center for O*NET Development. *O*NET OnLine*. Retrieved from https://www.onetonline.org/.

Nelson, T. D. (Ed.). (2004). *Ageism: Stereotyping and prejudice against older persons*. Cambridge, MA: MIT Press.

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, *56*(3), 799–866.

Neumark, D., Burn, I., & Button, P. (2016). Experimental age discrimination evidence and the Heckman critique. *American Economic Review*, *106*(5), 303–308.

Neumark, D., Burn, I., & Button, P. (2019). Is it harder for older workers to find jobs? New and improved evidence from a field experiment. *Journal of Political Economy*, *127*(2), 922–970.

Neumark, D., Burn, I., Button, P., & Chehras, N. (2019). Do state laws protecting older workers from discrimination reduce age discrimination in hiring? Evidence from a field experiment. *Journal of Law and Economics, 62(2)*, 373–402.

Neyt, B., Vandenbulcke, S., & Baert, S. (2019). Are men intimidated by highly educated women? Undercover on Tinder. *Economics of Education Review*, *73*(12), 101914. doi: 10.1016/j.econedurev.2019.101914.

Nuijten, M. P., Poell, R. F., & Alfes, K. (2017). Extracurricular activities of Dutch university students and their effect on employment opportunities as perceived by both students and organizations. *International Journal of Selection and Assessment*, *25*(4), 360–370.

OECD. (2017). Pensions at a glance 2017: OECD and G20 indicators. Paris: OECD Publishing. Retrieved from http://dx.doi.org/10.1787/pension_glance-2017-en.

OECD. (2019a). Employment rate by age group (indicator). Paris: OECD Publishing. doi: 10.1787/084f32c7-en (Accessed on 16 June 2019).

OECD. (2019b). Society at a glance 2019: OECD social indicators. Paris: OECD Publishing. Retrieved from https://doi.org/10.1787/soc_glance-2019-en.

Olian, J. D., Schwab, D. P., & Haberfeld, Y. (1988). The impact of applicant gender compared to qualifications on hiring recommendations: A meta-analysis of experimental studies. *Organizational*

Behavior and Human Decision Processes, *41*(2), 180–195.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*(5), 411–419.

Perry, E. L., & Bourhis, A. C. (1998). A closer look at the role of applicant age in selection decisions. *Journal of Applied Social Psychology*, *28*(18), 1670–1697.

Perry, E. L., & Finkelstein, L. M. (1999). Toward a broader view of age discrimination in employment-related decisions: A joint consideration of organizational factors and cognitive processes. *Human Resource Management Review*, *9*(1), 21–49.

Perry, E. L., Kulik, C. T., & Bourhis, A. C. (1996). Moderating effects of personal and contextual factors in age discrimination. *Journal of Applied Psychology*, *81*(6), 628–647.

Posthuma, R. A., & Campion, M. A. (2009). Age stereotypes in the workplace: Common stereotypes, moderators, and future research directions. *Journal of Management*, *35*(1), 158–188.

Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology, 299*(2012), 172–179.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *38*(1), 119–125.

Riach, P. A. (2015). A field experiment investigating age discrimination in four European labour markets. *International Review of Applied Economics, 29(*5), 608–619.

Riach, P., & Rich, J. (2004). Fishing for discrimination. *Review of Social Economy, 62*, 465–486.

Riach, P., & Rich, J. (2010). An experimental investigation of age discrimination in the English labor market. *Annals of Economics and Statistics, 99,* 169–185.

Richardson, B., Webb, J., Webber, L., & Smith, K. (2013). Age discrimination in the evaluation of job applicants. *Journal of Applied Social Psychology*, *43*(1), 35–44.

Rivera, L. A., & Tilcsik, A. (2016). Class advantage, commitment penalty: The gendered effect of social class signals in an elite labor market. *American Sociological Review*, *81*, 1097– 1131.

Roscigno, V. J., Mong, S., Byron, R., & Tester, G. (2007). Age discrimination, social closure and employment. *Social Forces*, *86*(1), 313–334.

Rosen, B., & Jerdee, T. H. (1976). The influence of age stereotypes on managerial decisions. *Journal of Applied Psychology*, *61*(4), 428–432.

Rosen, B., & Jerdee, T. H. (1977). Too old or not too old. *Harvard Business Review*, *55*(6), 97–106.

Rossi, P. H., & Nock, S. L. (1982). *Measuring social judgements: The factorial survey approach*. Thousand Oaks, CA: Sage.

Roulin, N. (2015). Don't throw the baby out with the bathwater: Comparing data quality of crowdsourcing, online panels, and student samples. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 8*, 190–196.

Sârbescu, P., Costea, I., & Rusu, S. (2011). Psychometric properties of the Marlowe-Crowne Social Desirability Scale in a Romanian sample. *Procedia—Social and Behavioral Sciences*, *33*(1), 707–711.

Sauer, C., Auspurg, K., & Hinz, T. (2020). Designing multi-factorial survey experiments: Effects of presentation style (text or table), answering scales, and vignette order. *Methods, Data, Analyses, 14*(2), 195–214.

Sauer, C., Auspurg, K., Hinz, T., & Liebig, S. (2011). The application of factorial surveys in general population samples: The effects of respondent age and education on response times and response consistency. *Survey Research Methods, 5*, 89–102.

Scharn, M., Sewdas, R., Boot, C. R., Huisman, M., Lindeboom, M., & Van Der Beek, A. J. (2018). Domains and determinants of retirement timing: a systematic review of longitudinal studies. *BMC public health*, *18*(1), 1–14.

Schmidt, D. F., & Boland, S. M. (1986). Structure of perceptions of older adults: Evidence for multiple stereotypes. *Psychology and Aging, 1*, 255–260.

Schuh, S. C., Van Quaquebeke, N., Göritz, A. S., Xin, K. R., De Cremer, D., & Van Dick, R. (2016). Mixed feelings, mixed blessing? How ambivalence in organizational identification relates to employees' regulatory focus and citizenship behaviors. *Human Relations*, *69*(12), 2224–2249.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, NY: Houghton Mifflin Company.

Shamon, H., Dülmer, H., & Giza, A. (2019). The factorial survey: the impact of the presentation format of vignettes on answer behavior and processing time. *Sociological Methods & Research*, *1*(1), 1-43.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics, 87*(3), 355–374.

Taylor, P., & Urwin, P. (2001). Age and participation in vocational education and training. *Work, Employment and Society, 15*(4), 763–779.

Taylor, P., & Walker, A. (1998). Employers and older workers: Attitudes and employment practices. *Ageing & Society, 18*(6), 641–658.

Tinsley, M. (2012). *Too much to lose: Understanding and supporting Britain's older workers*. London: Policy Exchange.

Van Belle, E., Caers, R., De Couck, M., Di Stasio, V., & Baert, S. (2019). The signal of applying for a job under a vacancy referral scheme. *Industrial Relations*, *58*(2), 251–274.

Van Belle, E., Di Stasio, V., Caers, R., De Couck, M., & Baert, S. (2018). Why are employers put off by long spells of unemployment? *European Sociological Review*, *34*(6), 694–710.

Van Borm, H., & Baert, S. (2018). What drives hiring discrimination against transgenders? *International Journal of Manpower, 39*(4), 581–599.

Van Dalen, H. P., Henkens, K., & Schippers, J. (2009). Dealing with older workers in Europe: A comparative survey of employers' attitudes and actions. *Journal of European Social Policy*, *19*(1), 47–60.

Van Hoye, G., & Lievens, F. (2003). The effects of sexual orientation or hirability ratings: An experimental study. *Journal of Business and Psychology*, *18*(1), 15–30.

VanderWeele, T. J., Valeri, L. & Ogburn, E. L. (2012). The role of measurement error and misclassification in mediation analysis. *Epidemiology*, *23*, 561–564.

Vishwanath, T. (1989). Job search, stigma effect, and escape rate from unemployment. *Journal of Labor Economics, 7*, 487–502.

Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Science Research, 38*, 505–520.

Warr, P., & Pennington, J. (1993). Views about age discrimination and older workers. In P. Taylor, A. Walker, B. Casey & H. Metcalf (Eds). *Age and employment: Policies, attitudes, and practices*. London: Institute of Personnel Management.

Appendix A: Additional Tables

< Table A–1 about here >

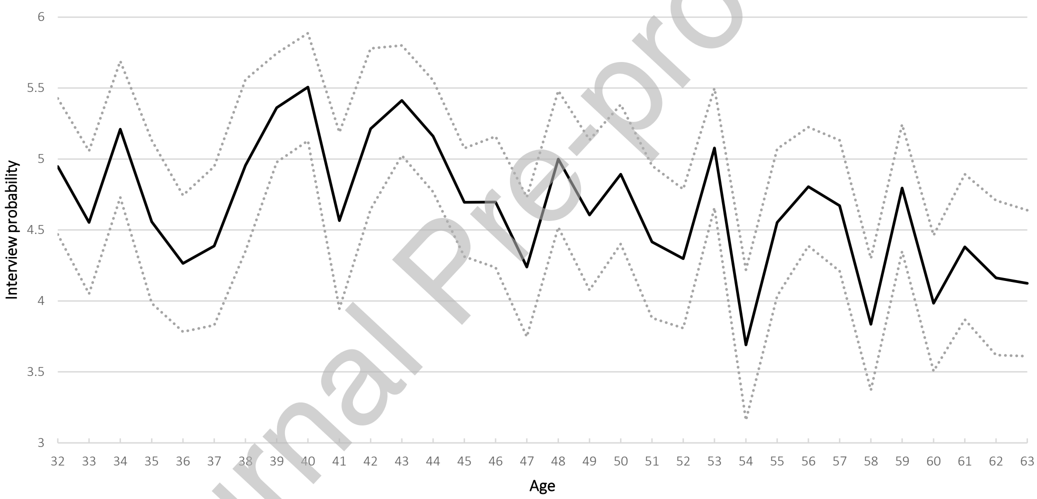
< Table A–2 about here >

< Table A–3 about here >

< Table A–4 about here >

< Table A–5 about here >

Figure 1. Interview Probability by Age



Notes: The thick line shows the (average) interview probability by candidate age. The dotted lines show the upper and lower bounds of the 95% confidence interval around these average values. The confidence bounds are corrected for the clustering of the observations at the participant level.

Table 1. Vignette Factors and Corresponding Levels Used in the Experimental Materials

Vignette factors	Vignette levels
Gender	{Male, Female}
Age	{32, 33, ..., 63}
Commuting distance	{0–5 km, 5–10 km, 10–50 km, More than 50 km}
Experience in the occupation	{None, About 2 years, About 5 years, About 10 years}
Extracurricular activities	{None, Volunteering, Sport activities, Cultural activities}

Notes: The factorial product of the vignette levels (i.e., 2 x 32 x 4 x 4 x 4) resulted in 4096 possible combinations. Forty sets of five vignettes were drawn from this vignette universe using a D-efficient design (D-efficiency: 99.11; Auspurg & Hinz, 2014) and distributed at random to the participants, as described in Subsection 2.1.

Table 2. Statements Used in the Experimental Materials

Signals and evaluation outcome	Statements
Perceived mental abilities	'I think this person has sufficient intellectual capacity to perform this job well.'
Perceived social abilities	'I think this person has sufficient social capacity to perform this job well.'
Perceived physical abilities	'I think this person has sufficient physical capacity to perform this job well.'
Perceived technological knowledge and skills	'I think this person has sufficient technological knowledge and skills to perform this job well.'
Perceived flexibility	'I think this person is sufficiently flexible to perform this job well.'
Perceived creativity	'I think this person is sufficiently creative to perform this job well.'
Perceived experience	'I think this person has sufficient experience to perform this job well.'
Perceived motivation	'I think this person is sufficiently motivated to perform this job well.'
Perceived reliability	'I think this person is sufficiently reliable to perform this job well.'
Perceived accuracy	'I think this person is sufficiently accurate to perform this job well.'
Perceived trainability	'I think this person is sufficiently trainable to perform this job well.'
Perceived reasonability with respect to wage expectations	'I think this candidate would have reasonable wage expectations.'
Attitude towards collaboration of employer	'I think I would enjoy collaborating with this person.'
Attitude towards collaboration of other employees	'I think other employees would enjoy collaborating with this person.'
Attitude towards collaboration of customers	'I think customers would enjoy collaborating with this person.'
Interview probability	'I will invite the candidate for a job interview for the described position.'

Note: In this table, we present the potential age signals, the evaluation outcome, and their corresponding statements as they were included in the online survey experiment. The participants evaluated each statement on a 7-point Likert scale ranging from 1 (i.e., 'completely disagree') to 7 (i.e., 'completely agree').

Table 3. Data Description by Fictitious Candidate's Age

	(1)	(2)	(3)	(4)
	Total sample [N = 2000]	Candidate's age below sample mean [N = 1018]	Candidate's age above sample mean [N = 982]	Difference (iii) – (ii)
A. PARTICIPANT CHARACTERISTICS				
Gender: female	0.447	0.434	0.461	0.027 [1.219]
Age: < 35 years old	0.532	0.530	0.535	0.004 [0.187]
Residence: United States	0.897	0.906	0.890	−0.017 [1.230]
Highest educational degree: university	0.700	0.708	0.691	−0.017 [0.820]
Frequency of hiring: ≥ once per semester	0.952	0.955	0.950	−0.005 [0.495]
Experience as HR professional: > 5 years	0.352	0.342	0.363	0.022 [1.015]
Percentage older employees in firm: ≥ 20%	0.592	0.587	0.598	0.10 [0.470]
B. JOB CHARACTERISTICS				
Dental technician	0.132	0.131	0.134	0.004 [0.249]
Door-to-door sales worker	0.125	0.123	0.127	0.004 [0.304]
Packer	0.120	0.116	0.124	0.008 [0.572]
CNC machine operator	0.115	0.113	0.117	0.004 [0.290]
Lab technician (cytogenetic techniques)	0.122	0.120	0.125	0.005 [0.369]
Insurance sales agent	0.132	0.141	0.123	−0.018 [1.202]
Physiotherapist	0.137	0.134	0.141	0.008 [0.516]
Database administrator	0.115	0.123	0.107	−0.016 [1.112]
C. EVALUATION OUTCOME				
Interview probability	4.656	4.851	4.454	−0.396*** [4.548]

Note: 47.5 is the sample mean candidate age. T-tests are performed to test whether the differences between the subsamples by candidate age are significantly different from 0. X^2-tests, which are more appropriate for binary outcomes, yield exactly the same conclusions. *** (**) ((*)) indicates significance at 1% (5%) ((10%)) significance level. T-statistics are in brackets.

Table 4. Bivariate Relation between Candidate's Age and Employers' Perceptions of Candidates

	Pearson correlation coefficients	Regression coefficients
A. SIGNALS		
Perceived mental abilities	−0.038 [0.089]	−0.004 [0.093]
Perceived social abilities	**−0.064** [0.004]	**−0.007** [0.005]
Perceived physical abilities	**−0.233** [0.000]	**−0.026** [0.000]
Perceived technological knowledge and skills	**−0.113** [0.000]	**−0.012** [0.000]
Perceived flexibility	**−0.145** [0.000]	**−0.016** [0.000]
Perceived creativity	**−0.091** [0.000]	**−0.010** [0.000]
Perceived experience	−0.012 [0.605]	−0.001 [0.609]
Perceived motivation	**−0.098** [0.000]	**−0.011** [0.000]
Perceived reliability	**−0.064** [0.004]	**−0.007** [0.005]
Perceived accuracy	**−0.062** [0.006]	**−0.007** [0.006]
Perceived trainability	**−0.183** [0.000]	**−0.020** [0.000]
Perceived reasonability with respect to wage expectations	**−0.061** [0.006]	**−0.007** [0.011]
Attitude towards collaboration of employer	**−0.099** [0.000]	**−0.011** [0.000]
Attitude towards collaboration of other employees	**−0.106** [0.000]	**−0.012** [0.000]
Attitude towards collaboration of customers	**−0.100** [0.000]	**−0.011** [0.000]
B. EVALUATION OUTCOME		
Interview probability	**−0.109** [0.000]	**−0.012** [0.000]

Notes: As discussed in Section 3, we present Pearson correlation coefficients between the candidate's age and the measured signals and attitudes (column 1). Spearman correlation coefficients were also calculated and led to exactly the same conclusions. In column 2, we present coefficient estimates for the simple linear regression model in which we regressed the standardised version of the signals and attitudes on the candidate's age. Regressions controlling for the other candidate characteristics and the participant characteristics included in Table 3 yield very similar coefficients. *P*-values are presented in brackets and corrected for the clustering of the observations at the participant level. Coefficients related to *p*-values below 5% are in bold. N = 2000.

Table 5. Multiple Mediation Analysis

	(1)		(2)		(3)		(4)		(5)	
	Total sample [N = 2000]		Subsample: American participants [N = 1795]		Subsample: Participants involved in hiring at least once a semester [N = 1905]		Subsample: Respondents who felt competent to evaluate job candidates for the described vacancy [N = 1770]		Subsample: Participants with tendency towards answering in a socially desirable manner below sample mean increased by 1 standard deviation [N = 1635]	
	% of total age effect explained by mediator	95% confidence interval	% of total age effect explained by mediator	95% confidence interval	% of total age effect explained by mediator	95% confidence interval	% of total age effect explained by mediator	95% confidence interval	% of total age effect explained by mediator	95% confidence interval
Perceived mental abilities	**3%****	[0%, 6%]	**4%****	[0%, 7%]	**3%****	[0%, 6%]	**3%****	[0%,6%]	**4%****	[0%, 8%]
Perceived social abilities	-1%	[-3%, 2%]	-1%	[-5%, 2%]	0%	[-3%, 3%]	-1%	[-4%, 2%]	0%	[-4%, 4%]
Perceived physical abilities	4%	[-4%, 12%]	0%	[-5%, 14%]	3%	[-5%, 12%]	2%	[-7%, 12%]	4%	[-6%, 15%]
Perceived technological knowledge and skills	**18%*****	[11%, 25%]	**18%*****	[10%, 26%]	**19%*****	[11%, 26%]	**19%*****	[12%, 27%]	**17%*****	[9%, 26%]
Perceived flexibility	**11%*****	[4%, 18%]	**13%*****	[6%, 21%]	**11%*****	[4%, 17%]	**13%*****	[6%, 21%]	**12%*****	[4%, 19%]
Perceived creativity	1%	[-3%, 5%]	0%	[-5%, 4%]	1%	[-3%, 5%]	2%	[-3%, 7%]	0%	[-4%, 5%]
Perceived experience	**12%*****	[4%, 21%]	**12%****	[3%, 21%]	**12%*****	[3%, 20%]	8%*	[-1%, 18%]	**13%****	[2%, 23%]
Perceived motivation	-2%	[-7%, 2%]	-2%	[-7%, 3%]	-3%	[-7%, 2%]	-2%	[-7%, 3%]	-5%*	[-10%, 1%]
Perceived reliability	-3%*	[-6%, 0%]	-4%*	[-8%, 0%]	-2%	[-6%, 1%]	**-4%****	[-8%, 0%]	-2%	[-6%, 2%]
Perceived accuracy	3%	[-1%, 7%]	2%	[-2%, 6%]	3%	[-1%, 7%]	4%*	[0%, 9%]	4%	[-1%, 9%]
Perceived trainability	**12%*****	[4%, 19%]	**11%*****	[3%, 18%]	**11%*****	[4%, 19%]	**11%*****	[3%, 19%]	**12%*****	[3%, 21%]
Perceived reasonability with respect to wage expectations	**3%****	[0%, 6%]	3%*	[0%, 6%]	**4%****	[0%, 7%]	**4%****	[0%, 8%]	**4%****	[0%, 8%]
Attitude towards collaboration of employer	1%	[-4%, 7%]	1%	[-6%, 7%]	1%	[-4%, 7%]	1%	[-5%, 6%]	1%	[-6%, 8%]
Attitude towards collaboration of other employees	2%	[-4%, 7%]	2%	[-5%, 9%]	2%	[-4%, 7%]	4%	[-4%, 11%]	2%	[-5%, 9%]
Attitude towards collaboration of customers	1%	[-3%, 6%]	3%	[-4%, 9%]	1%	[-4%, 6%]	0%	[-5%, 6%]	0%	[-6%, 7%]

Notes: The presented percentages are the results of the mediation model outlined in Section 3. We present the results as percentages of the total effect of age on the candidate's interview probability explained by the 15 mediators. *** (**) ((*)) indicates significance at 1% (5%) ((10%)) significance level. *p*-values are corrected for the clustering of the observations at the participant level. Percentages related to *p*-values below 5% are in bold. Bootstrapped 95% confidence intervals are presented between brackets. We present the lower bound and upper bound of the 95% confidence intervals as percentages calculated by dividing the lower and upper bounds of the mediation effect's 95% confidence intervals by the point estimate of the total effect of age on the candidate's interview probability.

Table 6. Multivariate Regression Analysis with Interview Probability as Outcome Variable

	(1)	(2)	(3)	(4)	(5)
A. CANDIDATE CHARACTERISTICS					
Female gender	−0.072 (0.062)	−0.061 (0.062)	−0.072 (0.063)	−0.069 (0.063)	−0.062 (0.062)
Commuting distance					
0–5 km	0.597*** (0.092)	0.600*** (0.093)	0.590*** (0.091)	0.586*** (0.091)	0.591*** (0.092)
5–10 km	0.461*** (0.091)	0.462*** (0.092)	0.454*** (0.090)	0.453*** (0.090)	0.455*** (0.092)
10–50 km	0.462*** (0.084)	0.447*** (0.085)	0.454*** (0.083)	0.451*** (0.083)	0.439*** (0.085)
More than 50 km (reference)					
Experience					
About 2 years	2.083*** (0.104)	2.077*** (0.104)	2.078*** (0.104)	2.079*** (0.104)	2.072*** (0.103)
About 5 years	2.666*** (0.112)	2.662*** (0.112)	2.666*** (0.112)	2.669*** (0.113)	2.662*** (0.112)
About 10 years	3.236*** (0.117)	3.230*** (0.117)	3.236*** (0.117)	3.236*** (0.117)	3.229*** (0.116)
None (reference)					
Extracurricular activities					
Volunteering	0.266*** (0.088)	0.265*** (0.088)	0.263*** (0.088)	0.267*** (0.088)	0.257*** (0.088)
Sport activities	0.220** (0.092)	0.219** (0.092)	0.224** (0.092)	0.228** (0.093)	0.223** (0.092)
Cultural activities	0.226** (0.094)	0.231** (0.094)	0.227** (0.094)	0.230** (0.095)	0.231** (0.094)
None (reference)					
Age	−0.030*** (0.004)	−0.044** (0.017)	−0.043*** (0.008)	−0.039*** (0.011)	−0.057*** (0.020)
B. PARTICIPANT CHARACTERISTICS					
Gender: female	0.164* (0.088)	−0.298 (0.359)	0.163* (0.088)	0.161* (0.088)	−0.224 (0.358)
Age: < 35 years old	0.298*** (0.102)	0.544 (0.369)	0.294*** (0.101)	0.291*** (0.102)	0.565 (0.366)
Residence: United States	−0.092 (0.124)	−0.568 (0.534)	−0.086 (0.124)	−0.081 (0.126)	−0.727 (0.527)
Highest educational degree: university	−0.161* (0.096)	−0.099 (0.383)	−0.161* (0.096)	−0.160* (0.096)	−0.112 (0.386)
Frequency of hiring: ≥ once per semester	0.280* (0.167)	0.467 (0.479)	0.286* (0.166)	0.293* (0.167)	0.540 (0.481)
Experience as HR professional: > 5 years	0.011 (0.107)	0.496 (0.387)	0.013 (0.107)	0.010 (0.110)	0.616 (0.384)
Percentage older employees in firm: ≥ 20%	0.248*** (0.087)	−0.628* (0.359)	0.248*** (0.087)	0.249*** (0.088)	−0.654** (0.358)
Candidate's age x Gender: female		0.010 (0.008)			0.008 (0.008)
Candidate's age x Age: < 35 years old		−0.005 (0.008)			−0.006 (0.008)
Candidate's age x Residence: United States		0.010 (0.012)			0.013 (0.011)
Candidate's age x Highest educational degree: university		−0.001 (0.008)			−0.001 (0.008)
Candidate's age x Frequency of hiring: ≥ once per semester		−0.004 (0.011)			−0.005 (0.011)

	(1)	(2)	(3)	(4)	(5)
Candidate's age x Experience as HR professional: > 5 years		−0.010 (0.008)			−0.013 (0.008)
Candidate's age x Percentage of older employees in firm: ≥ 20%		0.019** (0.008)			0.019** (0.008)
C. Job characteristics					
Level of required skills in occupation: high	0.062 (0.087)	0.065 (0.088)	−0.748** (0.348)		−0.806** (0.355)
Level of required customer contact in occupation: high	0.221* (0.116)	0.214* (0.117)	−0.163 (0.456)		−0.049 (0.459)
Level of required physical effort in occupation: high	0.196 (0.120)	0.190 (0.121)	0.099 (0.492)		0.215 (0.496)
Level of required technological skills in occupation: high	−0.008 (0.126)	−0.019 (0.126)	−0.358 (0.502)		−0.465 (0.494)
Occupation					
Door-to-door sales worker				0.322 (0.686)	
Packer				0.309 (0.727)	
CNC machine operator				−0.230 (0.711)	
Lab technician (cytogenetic techniques)				−0.336 (0.664)	
Insurance sales agent				−0.944 (0.656)	
Physiotherapist				−0.444 (0.721)	
Database administrator				−0.829 (0.760)	
Dental technician (reference)					
Candidate's age x Level of required skills in occupation: high			0.017** (0.007)		0.018** (0.007)
Candidate's age x Level of required customer contact in occupation: high			0.008 (0.010)		0.006 (0.010)
Candidate's age x Level of required physical effort in occupation: high			0.002 (0.010)		−0.000 (0.010)
Candidate's age x Level of required technological skills in occupation: high			0.007 (0.011)		0.009 (0.011)
Candidate's age x Door-to-door sales worker				−0.002 (0.015)	
Candidate's age x Packer				−0.003 (0.015)	
Candidate's age x CNC machine operator				0.005 (0.016)	
Candidate's age x Lab technician (cytogenetic techniques)				0.008 (0.015)	
Candidate's age x Insurance sales agent				0.026* (0.014)	
Candidate's age x Physiotherapist				0.015 (0.015)	
Candidate's age x Database administrator				0.018 (0.017)	
Observations			2000		

Notes: The presented statistics are coefficient estimates and standard errors in parentheses for the regression model outlined in Section 3. More concretely, in column (1), we present the coefficient estimates of our baseline model. In column (2), (3), (4), and (5), we present the coefficient estimates of four models in which we, subsequently, add different interaction terms to the baseline model. Standard errors are corrected for the clustering of the observations at the participant level. *** (**) ((*)) indicates significance at the 1% (5%) ((10%)) significance level.

Table A–1. Jobs and Corresponding Job Characteristics Used in the Experimental Materials

Job	Required skills	Level of customer contact	Level of physical effort	Required technological skills
Dental technician	Low	Low	Low	Low
Door-to-door sales worker	Low	High	Low	Low
Packer	Low	Low	High	Low
CNC machine operator	Low	Low	Low	High
Lab technician (cytogenetic techniques)	High	Low	Low	Low
Insurance sales agent	High	High	Low	Low
Physiotherapist	High	Low	High	Low
Database administrator	High	Low	Low	High

Note: Jobs were selected and categorised based on data provided by O*NET, as described in Subsection 2.2.

Table A–2. Job Descriptions Used in the Experimental Materials

Job function	Job description
Dental technician	'This employee will be responsible for the construction or repair of partial or full dentures and other dental constructions.'
Door-to-door sales worker	'This employee will be responsible for selling goods or services door-to-door or on the street.'
Packer	'This employee will be responsible for packaging a wide variety of products and materials (in an industrial environment).'
CNC machine operator	'This employee will be responsible for setting up machines that mill, shape and/or engrave plastic or metal work pieces.'
Lab technician (cytogenetic techniques)	'This employee will be responsible for analysing chromosomes (in biological material such as amniotic fluid, bone marrow, and blood) in view of studying, diagnosing, or treating genetic diseases.'
Insurance sales agent	'This employee will be responsible for selling insurance, including life, property, accident, and health insurance.'
Physiotherapist	'This employee will be responsible for physically (physiotherapeutically) guiding individuals with exceptional physical needs due to gross motor development disorders or other disorders.'
Database administrator	'This employee will be responsible for the implementation, testing, management, security, and reworking of computer databases using data management systems.'

Note: Job functions and descriptions were provided by O*NET, as described in Subsection 2.2.

Table A–3. Comparison between Participant Characteristics and Characteristics of HR Professionals in ACS

	(1)	(2)
Participant characteristics	Mean among participants in experiment	Mean among HR professionals in ACS
Female gender	0.447	0.670
Age	36.089	45.363
Highest educational degree: secondary education or lower	0.150	0.140
Highest educational degree: tertiary education	0.850	0.860

Notes: We combined ACS data conducted in the years 2010- 2019 and selected all respondents with occ2010 occupation codes 0130 (Human Resources Managers), 0620 (Human Resources, Training, and Labour Relations Specialists), and 5360 (Human Resources Assistants, Except Payroll and Timekeeping).

Table A–4. Multiple Mediation Analysis by Fictitious Candidate Gender

	(1)		(2)	
	Subsample: Female candidates [N = 997]		Subsample: Male candidates [N = 1003]	
	% of total age effect explained by mediator	95% confidence interval	% of total age effect explained by mediator	95% confidence interval
Perceived mental abilities	2%	[-2%, 7%]	4%*	[0%, 8%]
Perceived social abilities	1%	[-3%, 6%]	-2%	[-6%, 2%]
Perceived physical abilities	2%	[-8%, 13%]	5%	[-6%, 16%]
Perceived technological knowledge and skills	**20%*** **	[8%, 32%]	**15%*** **	[6%, 24%]
Perceived flexibility	**15%*** **	[4%, 26%]	**9%** **	[0%, 17%]
Perceived creativity	-2%	[-8%, 4%]	4%	[-2%, 9%]
Perceived experience	5%	[-10%, 20%]	**15%*** **	[5%, 26%]
Perceived motivation	-8%	[-18%, 2%]	0%	[-4%, 4%]
Perceived reliability	-1%	[-7%, 5%]	-5%*	[-9%, 0%]
Perceived accuracy	2%	[-5%, 9%]	4%*	[0%, 8%]
Perceived trainability	**10%** **	[1%, 19%]	**16%*** **	[5%, 26%]
Perceived reasonability concerning wage expectations	4%*	[0%, 9%]	2%	[-1%, 5%]
Attitude towards collaboration of employer	4%	[-4%, 12%]	-1%	[-8%, 5%]
Attitude towards collaboration of other employees	4%	[-8%, 15%]	0%	[-6%, 5%]
Attitude towards collaboration of customers	1%	[-7%, 9%]	2%	[-3%, 8%]

Notes: We present the results as percentages of the total effect of age on the candidate's interview probability explained by the 15 mediators. *** (**) ((*)) indicates significance at 1% (5%) ((10%)) significance level. *p*-values are corrected for the clustering of the observations at the participant level. Percentages related to *p*-values below 5% are in bold. Bootstrapped 95% confidence intervals are presented between brackets. We present the lower bound and upper bound of the 95% confidence intervals as percentages calculated by dividing the lower and upper bounds of the mediation effect's 95% confidence intervals by the point estimate of the total effect of age on the candidate's interview probability.

Table A–5. Multiple Mediation Analysis by Participants' Age and Gender

	(1)		(2)		(3)		(4)	
	Subsample: Participants < 35 years old [N = 1065]		Subsample: Participants ≥ 35 years old [N = 935]		Subsample: Female participants [N=895]		Subsample: Male participants [N=1105]	
	% of total age effect explained by mediator	95% confidence interval	% of total age effect explained by mediator	95% confidence interval	% of total age effect explained by mediator	95% confidence interval	% of total age effect explained by mediator	95% confidence interval
Perceived mental abilities	3%	[-3%, 9%]	2%	[-3%, 7%]	4%	[-1%, 9%]	2%	[-1%, 5%]
Perceived social abilities	-1%	[-5%, 2%]	0%	[-5%, 5%]	-1%	[-8%, 6%]	0%	[-3%, 3%]
Perceived physical abilities	6%	[-7%, 20%]	4%	[-12%, 21%]	8%	[-6%, 22%]	0%	[-10%, 11%]
Perceived technological knowledge and skills	**25%*****	[12%, 39%]	10%	[-4%, 23%]	**19%*****	[7%, 31%]	**18%*****	[10%, 27%]
Perceived flexibility	**14%****	[3%, 25%]	**15%****	[2%, 28%]	**12%****	[2%, 22%]	**11%****	[2%, 20%]
Perceived creativity	3%	[-3%, 8%]	-6%	[-15%, 3%]	1%	[-5%, 7%]	-1%	[-6%, 5%]
Perceived experience	9%	[-9%, 28%]	-6%	[-42%, 29%]	13%	[-4%, 30%]	**12%****	[3%, 21%]
Perceived motivation	-3%	[-8%, 3%]	-2%	[-13%, 8%]	-1%	[-7%, 4%]	-2%	[-9%, 4%]
Perceived reliability	0%	[-4%, 4%]	-5%	[-12%, 3%]	-6%	[-13%, 1%]	-1%	[-5%, 3%]
Perceived accuracy	4%	[-2%, 10%]	1%	[-6%, 7%]	3%	[-4%, 9%]	3%	[-2%, 9%]
Perceived trainability	10%*	[0%, 21%]	**18%****	[4%, 32%]	**8%****	[0%, 17%]	**15%****	[2%, 27%]
Perceived reasonability concerning wage expectations	4%*	[0%, 9%]	3%	[-2%, 7%]	2%	[-3%, 8%]	4%*	[0%, 8%]
Attitude towards collaboration of employer	-2%	[-9%, 6%]	10%	[-2%, 22%]	-2%	[-11%, 6%]	5%	[-3%, 12%]
Attitude towards collaboration of other employees	4%	[-3%, 12%]	-2%	[-15%, 10%]	2%	[-7%, 11%]	2%	[-5%, 9%]
Attitude towards collaboration of customers	-1%	[-7%, 4%]	1%	[-10%, 12%]	3%	[-4%, 10%]	0%	[-7%, 6%]

Notes: We present the results as percentages of the total effect of age on the candidate's interview probability explained by the 15 mediators. *** (**) ((*)) indicates significance at 1% (5%) ((10%)) significance level. p-values are corrected for the clustering of the observations at the participant level. Percentages related to p-values below 5% are in bold. Bootstrapped 95% confidence intervals are presented between brackets. We present the lower bound and upper bound of the 95% confidence intervals as percentages calculated by dividing the lower and upper bounds of the mediation effect's 95% confidence intervals by the point estimate of the total effect of age on the candidate's interview probability.