

# Age Discrimination in Hiring Decisions: A Factorial Survey among Managers in Nine European Countries

Jelle Lössbroek <sup>1,\*</sup>, Bram Lancee<sup>2</sup>, Tanja van der Lippe<sup>1</sup> and  
Joop Schippers<sup>3</sup>

<sup>1</sup>Department of Sociology, Utrecht University, 3584 CH Utrecht, The Netherlands, <sup>2</sup>Department of Sociology, University of Amsterdam, 1018 WV Amsterdam, The Netherlands and <sup>3</sup>Department of Economics, Utrecht University, 3584 EC Utrecht, The Netherlands

\*Corresponding author. Email: [j.j.h.lossbroek@uu.nl](mailto:j.j.h.lossbroek@uu.nl)

Submitted April 2019; revised May 2020; accepted May 2020

## Abstract

This article analyses old-age discrimination in managers' hypothetical hiring decisions. We expect that older job candidates are less likely to be hired than equally qualified younger candidates. Statistical discrimination theory argues that when recruiters have more information about the candidate's skills, age is less important for hiring decisions. Given inconclusive results of previous studies, we elaborate on the theory by focusing on the content rather than the amount of information. We argue that information is primarily influential if it debunks, rather than confirms, ageist stereotypes. To test this argument, a factorial survey was conducted among 482 managers in nine European countries. The findings show that older candidates indeed receive lower hireability scores, and this finding is robust across countries and sectors. However, we do not find that stereotype-rejecting information moderates age discrimination: it does not matter whether recruiters have information that debunks or confirms ageist stereotypes; age is equally important in both situations. Our findings suggest that for hiring decisions, the valuation of applicants' skills and their age are largely independent.

## Introduction

While many European governments have increased retirement ages and restricted early exit options (Van Dalen, Henkens and Schippers, 2007), job opportunities for older workers are bleak. Labour laws protect older workers to a large extent from being fired due to their age (Van Dalen *et al.*, 2007; Martin *et al.*, 2014), but age discrimination is consistently shown to prevent older people from acquiring new jobs (Bendick, Brown and

Wall, 1999; Büsch, Dahl and Dittrich, 2009; Karpinska, Henkens and Schippers, 2013a; Oude Mulders *et al.*, 2014). Workplace age discrimination can decrease opportunities and job satisfaction of older employees, performance of organizations, and sustainability of social security provisions (Snape and Redman, 2003; Weiss and Maurer, 2004; Kluge and Krings, 2008). Indeed, long-term unemployment is highest among the older working population (Büsch *et al.*, 2009), primarily

due to a lack of labour demand from employers (Vickerstaff, Cox and Keen, 2003).

In stark contrast to the virtual consensus on the existence of age discrimination in the hiring process is the scarce examination of its reasons (Martin *et al.*, 2014). This article contributes to the literature by studying why employers discriminate. A central assumption in the theory of statistical discrimination is that employers strive to select the most productive candidate but have limited information about job candidates (Phelps, 1972; Aigner and Cain, 1977; Fang and Moro, 2011; Ewens, Tomlin and Wang, 2014; Oude Mulders *et al.*, 2018). To cope with this lack of information, employers use the age of a job candidate as a proxy for expected productivity, assuming that older employees are less productive.

For example, employers often assume that an older candidate is less healthy than a younger candidate, believing older workers are less healthy in general (Steinberg *et al.*, 1996; Oude Mulders *et al.*, 2014). As a result, employers are less inclined to hire older workers. However, when employers have information about the health of an individual job candidate, there is no need to use age as a proxy for expected productivity, to the extent that this is health-related.

So far, studies did not find empirical support for statistical age discrimination (Lahey, 2008; Neumark, Burn and Button, 2016; Carlsson and Eriksson, 2017). This may be because existing work predominantly focused on the *availability* of information, without recognizing that the *content* of information may matter even more. Studies on ethnic discrimination find that stereotype-rejecting information sends a stronger signal than stereotype-confirming information (Ewens *et al.*, 2014; Schaeffer, Höhne and Teney, 2015). For example, negative information about a candidate's educational attainment has a stronger influence on the likelihood to be hired for native Germans than for immigrants, since the stereotype is that native Germans are higher educated than immigrants (Schaeffer *et al.*, 2015). Following this line of argumentation, the information that an older job seeker is particularly healthy rejects an ageist stereotype and is likely to reduce the negative effect of the candidate's age. By contrast, a less healthy candidate likely confirms employer's expectations. In this study, instead of analysing the presence or absence of information, we assess how hiring decisions are influenced by information that confirms stereotypes or rejects stereotypes.

To test this argument, we use a factorial survey, also known as a vignette experiment (Wallander, 2009; Di Stasio and Gërkhani, 2015). Like field experiments, vignettes are a suitable instrument to deal with social desirability norms by studying age-based preferences

indirectly. Vignettes allow for a greater control over numerous (hypothetical) candidates than field experiments, which enables studying relevant properties of candidates simultaneously. As our data were collected among actual managers, the external validity is higher than in commonly used student data (Rosen and Jerdee, 1976; Perry and Bourhis, 1998; Weiss and Maurer, 2004; Rosen and Rupp, Vodanovich and Credé, 2006; Karpinska, Henkens and Schippers, 2011). We use vignettes included in the European Sustainable Workforce Survey (Van der Lippe *et al.*, 2016), collected in 2015–2016 among 482 managers in nine European countries.

## Age Discrimination on the Labour Market

### The role of age

In spite of laws against discrimination, age discrimination is an important explanation for the position of older job applicants. Differences in the extent of discrimination are predominantly the result of personnel decisions made by middle-level managers (Karpinska *et al.*, 2013a; Karpinska, Henkens and Schippers, 2013b; Martin *et al.*, 2014; Oude Mulders *et al.*, 2014) also when adjusted for competing factors, such as equal opportunity policies (Bendick *et al.*, 1999) or formal rules (Taylor and Walker, 1998).

There are two sources of (age) discrimination: taste-based and statistical discrimination. The former refers to a general attitude of disliking and distaste towards a group, in this case older people. For this type of discrimination, people could be willing to make less economically efficient choices in order to satisfy their preference to avoid working with certain people (Becker 1971; Weiss and Maurer, 2004; List, 2013; Ewens *et al.*, 2014). According to statistical discrimination theory, employers aim to maximize individual applicants' net contribution to their organization; group membership is used as a proxy for absent information about an individual's productivity (Arrow 1973; Fang and Moro, 2011; Ewens *et al.*, 2014). Consequently, employers base individual hiring decisions on the group's (perceived) productivity. Put differently, statistical discrimination implies that a candidate's age is used as a proxy for invisible (negative) qualities that, on average, are (believed to be) related to a higher age (Ewens *et al.*, 2014). Thus, employers believe that older age comes with unfavourable characteristics such as reduced health, worn out human capital, or lower flexibility. While older people may be well appreciated in other roles, such as

neighbour or (grand) parent, they are disadvantaged for (supposedly) being less productive in the workplace.

Research indeed shows that employers assume that older workers score worse on productivity-related characteristics such as health and motivation (Finkelstein, Higgins and Clancy, 2000; Gray and McGregor, 2003; Loretto and White, 2006; Ng and Feldman, 2012; Principi, Fabbietti and Lamura, 2015). Empirical evidence provides a mixed answer as to whether these assumptions are true (Büsch *et al.*, 2009; Ng and Feldman, 2012), and even for assumptions which are rooted in empirical reality employers often overestimate the correlation between the proxy and true productivity (Pager and Karafin, 2009). Regardless of these assumptions being rooted in reality, employers will keep using the proxy as long as they feel that it works (Birkelund, 2016).

Previous studies have not yet been able to assess the importance of either source of age discrimination, but they generally indicate that age discrimination in labour market hiring exists. Therefore, our first hypothesis is:

*H1: Employers give older job candidates lower hireability scores compared to equally qualified younger candidates.*

### The role of information

Next, statistical discrimination theory predicts that the more information on a job candidate's productivity is available, the less employers rely on age as a selection criterion. A study on the American housing market indeed indicates that landlords discriminate less on ethnicity if they have more information about potential renters than if they only know their ethnicity (Ewens *et al.*, 2014). However, in age discrimination studies, the role of information is only scarcely studied. Bendick *et al.* (1999) find that age discrimination is particularly prominent in the early hiring phase, when employers know little about the applicants. This is in line with the argument that discrimination is more pronounced when employers know less about the individual candidate, although other differences between early and later hiring phases obscure conclusions about the role of information. Three labour market studies tested whether information about candidates reduces the age penalty: an American study analysed information such as absence, computer training, vocational training (Lahey, 2008), another American study analysed information such as language and computer skills (Neumark *et al.*, 2016), and a Swedish study analysed willingness to train and

unemployment status (Carlsson and Eriksson, 2017). None of these studies found that the availability of information affected the discrimination against older workers.

However, besides the availability of information on the productivity of a job candidate, the content may matter too. Social psychologists show that information that rejects existing stereotypes is more easily remembered than stereotype-reinforcing information (Stangor and McMillan, 1992). According to Ewens *et al.* (2014), stereotype-rejecting information acts as a 'surprise signal' and is for that reason more influential than stereotype-confirming information. Their study on ethnic discrimination in the American housing market supports this mechanism: landlords more often invited applicants with 'typically White' names than applicants with 'typically Black' names. In comparison to this base difference, applications containing negative information (e.g. a poor credit rating) showed a strong decline in invitation rates for White applicants, and a smaller decline for Black applicants (Ewens *et al.*, 2014). By contrast, Thijssen *et al.* (2019) do not find that adding personal information reduces hiring discrimination against Turks in Germany and the Netherlands. A German labour market study provides indirect support for the 'surprise signal' hypothesis: both native Germans and immigrants are disadvantaged by low educational attainment, but the disadvantage was far smaller for immigrants, i.e. for workers who are stereotyped to be lower educated (Schaeffer *et al.*, 2015).

Following statistical discrimination theory, stereotype-confirming information barely counts as relevant information, whereas stereotype-rejecting information does count as relevant information. Along this line of reasoning, information that matches ageist stereotypes (e.g. an unhealthy older applicant) merely confirms what employers already assumed, and may not substantially reduce the influence of age in hiring decisions. By contrast, stereotype-rejecting information substantially alters the expected productivity of the job candidate. This may be even more relevant in the case of older workers, as the variance in their productivity is arguably larger than among younger workers (Neumark *et al.*, 2016; Carlsson and Eriksson, 2017). As a consequence, hiring older job candidates is perceived to be 'riskier' than hiring younger candidates (Daniel and Siebert, 2005). Thus, a 'surprise signal' not only provides information on the *actual* productivity of a candidate but it also reduces the *insecurity* that employers may have as a result of a candidate's age (Ewens *et al.*, 2014; Neumark, 2016).

We thus expect that age is less influential in hiring decisions when employers have information about the job candidate that debunks ageist stereotypes, compared to having information that confirms stereotypes:

*H2: The influence of a job candidate's age on hiring decisions is more strongly reduced by stereotype-rejecting information than by stereotype-confirming information.*

## Data and Methods

### Data

We collected data as part of the European Sustainable Workforce Survey in 2015–2016, a cross-national organization survey that studies managers' and employees' behaviour in nine European countries: Bulgaria, Finland, Germany, Hungary, Netherlands, Portugal, Spain, Sweden, and the United Kingdom (Van der Lippe *et al.*, 2016). Organizations were approached using stratified random sampling, based on organization size (40–99 employees; 100–249 employees; 250+ employees) and organization sector (financial services; health care; higher education; manufacturing; telecommunications; transportation). The diversity of these organizations ensures that our findings are not dependent on possible country-specific or sector-specific issues. Organizations that refused to participate were replaced using a matched sampling strategy with organizations in the same stratum of sector and size. Among organizations agreeing to participate, the within-organization response rate for department managers is about 80 per cent ( $N=922$ ). They are treated as employers in this study, as they are involved in selecting personnel and

thus have actual experience with this task. After completing the survey, respondents were asked to participate in the vignette study (without specifying its content). Five hundred and four respondents participated in the vignette study, of whom 499 rated all profiles. Following best practice (Wallander, 2009), respondents were asked three post-experiment questions about the difficulty, realism, and their experience with rating such candidates. Removing the 17 respondents who skipped these questions results in a total of 482 respondents.<sup>1</sup> Each respondent received a vignette set of six profiles, resulting in 482 vignette sets containing 2,892 profiles. Because the age-related survey questions were followed by a broad variety of questions and topics, there is no reason to assume that 'carry-over' effects influenced the vignette findings (Oude Mulders *et al.*, 2014). Table 1 shows descriptive respondent characteristics.

### Research Design

We conducted a vignette experiment (factorial survey design). In this design, respondents are presented with a short scenario (see Appendix, Figure 1) and the instruction to rate six hypothetical job candidates who vary on several items of theoretical interest. Respondents were instructed to imagine they (in their current job) had to fill vacancies for typical jobs in their departments with matching salaries. The hypothetical job candidates were presented simultaneously as the respondents in our pilot study indicated that this reflects real-world hiring decisions.

Vignettes are frequently used to study inclined behaviour in employer hiring decisions (Karpinska *et al.*, 2011; Oude Mulders *et al.*, 2014) due to four key strengths. First, it gives great control over the characteristics of candidates. Second, respondents use a holistic approach in judging candidates, rather than rating separate aspects of a profile; this is more in line with actual hiring decisions. Although decisions are hypothetical and not 'real', previous studies suggest that people behave as if the decisions are real (Hainmueller, Hangartner and Yamamoto, 2015).<sup>2</sup> Third, we selected actual managers, rather than students, which is frequently done in age discrimination studies (Rosen and Jerdee, 1976; Weiss and Maurer, 2004; Rupp *et al.*, 2006; Karpinska *et al.*, 2011). This ensures that the external validity of the design is high. Fourth, vignettes can circumvent social desirability norms that could skew answers related to sensitive topics such as discrimination (Wallander, 2009).<sup>3</sup>

**Table 1.** Descriptive statistics of the respondents

	Mean	Std. Dev.	Min	Max
Manager characteristics				
Age	45.61	8.72	25	71
Gender (1 = female)	0.40	—	0	1
Seniority (years as manager)	6.17	5.68	0	30
Country				
Bulgaria	0.12	—	0	1
Finland	0.09	—	0	1
Germany	0.08	—	0	1
Hungary	0.09	—	0	1
Netherlands	0.25	—	0	1
Portugal	0.14	—	0	1
Spain	0.10	—	0	1
Sweden	0.07	—	0	1
United Kingdom	0.06	—	0	1

### Vignette Characteristics

The candidate profiles included eight characteristics: two demographic characteristics, age and gender, and six productivity indicators: experience, performance, training participation, motivation, health, and retirement intentions. Jointly, these indicators capture the three core dimensions of human capital: the amount of human capital (experience and training participation), the usage of human capital (performance and motivation), and the availability of human capital (health and retirement intentions). These indicators were selected as factors that are both important to employers' hiring decisions and commonly believed to be correlated to age.<sup>4</sup>

The most important demographic, *AGE*, has six values, with steps of four years: 43, 47, 51, 55, 59, and 63. This operationalization strikes a balance: on the one hand, it contains more age categories than commonly used in vignette studies on older workers, and, to be able to compare older workers with younger workers, they are spread over a larger range than in comparable vignette studies (Karpinska *et al.*, 2013a,b; Oude Mulders *et al.*, 2018). *GENDER* has two values: male and female.

To study the extent to which employers use age as a cue for productivity, we use six productivity indicators. For each indicator, we briefly discuss age stereotypes. The six indicators signal relatively low, or relatively high productivity. Dichotomies keep vignettes relatively simple, allowing respondents to weigh all available information rather than ignoring some of the information to make the decision easier. Also, this accustoms to common respondent behaviour to treat the middle value of a trichotomy as either the high or low value rather than a distinct value (Buskens, 1999). Table 2 shows all possible values.

*EXPERIENCE* is measured as having *little* or *considerable* experience in comparable positions.

Experience can make employees more effective and efficient in completing their tasks. Older workers have had more time to build up experience in related tasks than younger workers; indeed, experience is mentioned as a positive age stereotype about older workers (Henkens, 2005). *PERFORMANCE* is measured as *average* or *above-average* performance in a recent assessment.<sup>5</sup> Recent performance refers to the quality and quantity of an employee's work and can indicate future performance. Although older workers are stereotyped to be more precise, and, hence, deliver a higher quality, they are also stereotyped to deliver a lower quantity. Older workers are often believed to perform worse than younger counterparts (Martin *et al.*, 2014). *TRAINING* is measured as *whether or not a candidate had received training relevant for the job*. Participation in training can help employees to increase or maintain their human capital and shows a willingness to learn. Older workers, on average, show lower training participation than younger workers. Employers know this (Gringart, Helmes and Speelman, 2005; Loretto and White, 2006; Kluge and Krings, 2008; Ng and Feldman, 2012; Martin *et al.*, 2014), and the stereotype of lower willingness to learn has been found to be an important obstacle for older job candidates (Carlsson and Eriksson, 2017). For *MOTIVATION*, respondents are either *reasonably motivated* or *highly motivated*.<sup>6</sup> Motivation shapes the extent to which employees use their potential. More motivated workers contribute more, yet older workers are often believed to be less motivated than younger workers (Ng and Feldman, 2012). *HEALTH* is captured by the number of sick days, which reflects both physical and mental health. Having used *15 sick days* signals bad health; having used *2 sick days* signals good health. Health covers both physical health (e.g. greater vulnerability to sickness, reduced stamina and reduced

**Table 2.** Vignette characteristics

		Variables which are included in each profile					
Age	43	47	51	55	59	63	
Gender	Male			Female			
		Variables of which four out of six are included in each profile					
		Negative value			Positive value		
Experience	Has little experience in comparable positions			Has considerable experience in comparable positions			
Performance	Average performance in recent assessments			Above-average performance in recent assessments			
Training	Had no relevant training last year			Had various forms of training relevant to this job			
Motivation	Appears to be reasonably motivated			Appears to be highly motivated			
Health	Took 15 sick days in the past year			Took 2 sick days in the past year			
Retirement	Intends to retire early			Intends to retire at the statutory retirement age			

physical strength) and mental health (e.g. lower cognitive performance, loss of memory and stress resistance). Evidence on the relation between age and health is mixed (Steinberg *et al.*, 1996; Kluge and Krings, 2008; Oude Mulders *et al.*, 2014). Still, employers reported a fear of older applicants being less healthy and, through long-term absence, more expensive (Finkelstein *et al.*, 2000). *RETIREMENT* is signalled by the *intention to retire early* or to *retire at the statutory retirement age*. From the employer's perspective, retirement entails a loss of human capital and the need to replace the worker, which can lead to temporary understaffing, recruitment costs, training costs, and lower productivity during this transition period. Older workers are often believed to prefer early retirement, and this stereotype has been associated with ageism (Duncan and Loretto, 2004).

### Vignette Design

Respondents rated six candidates on scales ranging from 1 (extremely unlikely) to 10 (extremely likely) based on the question *How likely is it that you would hire these candidates?* This approach is based on previous studies (Oude Mulders *et al.*, 2014). A rating design provides not only an ordering but also the distance between the preferences, which is more informative about the strength of preferences than a ranking design. We used an orthogonal design to construct the vignettes, i.e. a random value was assigned to each variable. The values were drawn within the boundaries of the restrictions described below, which are used to prevent illogical cases and to guarantee that each vignette set can be used to test the interaction hypothesis. Each candidate has a different value for AGE; three candidates are women, three are men; four of the six productivity indicators are present on each candidate profile, the other two are absent. Having different indicators on the profiles stimulates respondents to spend more time comparing the candidates, and resembles real-world hiring decisions in which employers also have different information about different candidates. For candidates aged 43, 47, and 51, *RETIREMENT* was always absent: mentioning early retirement intentions would be out of place for younger candidates, and unnatural information may lead respondents to conclude there is something 'wrong' with this candidate (Wallander, 2009; Rich, 2014). For candidates aged 55, 59 and 63, *RETIREMENT* was once absent, once 'intends to retire at statutory retirement age' and once 'intends to retire early'. For the other five indicators, each was absent at least once, Positive Information at least once, and Negative Information at least once. Within the profiles, information was always

presented in the same order, which roughly corresponds to the order in which managers learn about this information in real-life situations. Profiles were randomly assigned a name ranging from 'Candidate A' to 'Candidate F' and presented in alphabetical order; response scales were also presented in alphabetical order. Appendix Figure A1 shows an example of a vignette set.

Table 3 shows the descriptive statistics of the vignette characteristics and their bivariate correlations. By design, age and gender are included on all 2,892 vignettes, and Retirement is included on a third of the vignettes (964). The other five vignette characteristics are all very close to the average of 2,121. The mean scores of the dichotomous variables are all close to 0.50, and their correlations are very weak and in all but two instances insignificant. This indicates that our orthogonal factorial design was successful.

### Estimation Strategy

To analyse the hireability scores employers attached to each vignette, we use linear regression analysis. We use dummy variables to compare positive information versus negative information for the six productivity indicators. Because each candidate profile contains four out of six productivity indicators and two absent indicators, it is impossible to estimate a model including all twelve dummy variables of our interest simultaneously (positive and negative information for the six indicators with absent information as reference category). All variance would already be captured by the inclusion of 11 out of 12 dummy variables.<sup>7</sup> Therefore, we use saturated models, i.e. models that capture all variance in the dependent variable that could be captured by the vignette variables. These saturated models of dummies allow us to control for all variation in the other indicators and to focus on the age effects. To test whether positive and negative information about experience differentially affect the importance of age, we select the vignettes that include experience. We then estimate the interaction term between age and having positive information on experience (negative information being the reference category), while this model is again saturated regarding the other productivity indicators. The same procedure is used for the other productivity indicators.

Respondents rating multiple vignettes might lead to interrater correlation, although there is no consensus on this (Wallander, 2009). Using fixed-effects or random-effects takes respondent characteristics into account in estimating the influence of candidates' characteristics on their rating (Di Stasio and Gërkhani, 2015). We used Hausmann tests to compare fixed-effects and

**Table 3.** Vignette characteristics descriptive statistics and bivariate correlations

		N	Mean	1	2	3	4	5	6	7
1	Age	2,892	53.00							
2	Gender	2,892	0.50	-0.03*						
3	Experience	2,117	0.50	-0.03	-0.00					
4	Performance	2,123	0.51	0.03	0.05*	-0.01				
5	Training	2,093	0.49	0.00	0.01	0.00	0.02			
6	Motivation	2,119	0.47	0.01	0.04	-0.04	0.00	0.01		
7	Health	2,152	0.49	-0.01	-0.03	-0.03	0.00	-0.04	-0.04	
8	Retirement	964	0.50	0.01	-0.00	0.04	-0.04	0.03	-0.06	-0.07

Notes: For gender, one represents a female profile; for the other dichotomous characteristics, one represents positive information. Respondent N = 482.  
\*P < 0.05.

**Table 4.** Random-effects linear regression of the influence of age on hireability score

	M4a	M4b
Female	0.06 (0.09)	0.05 (0.09)
Age (linear, in 4-year steps)	-0.40*** (0.03)	
Reference category: age = 43		
Age = 47		-0.25 (0.16)
Age = 51		-0.49** (0.16)
Age = 55		-1.09*** (0.16)
Age = 59		-1.64*** (0.16)
Age = 63		-1.84*** (0.16)
Constant	6.57*** (0.19)	6.07*** (0.19)
Vignette N	2,892	2,892
Respondent N	482	482
Respondent random effects	Yes	Yes
Saturated model	No	No

Note: Standard errors in parentheses.

\*P < 0.05,

\*\*P < 0.01,

\*\*\*P < 0.001,

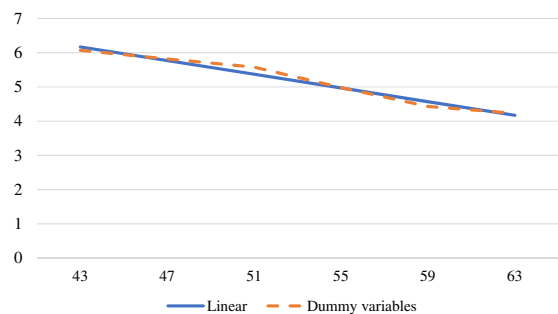
two-tailed test. Dependent variable: hireability score (1–10). Unstandardized coefficients.

maximum-likelihood random-effects models, which indicated that coefficients were consistent for nearly all models, supporting our choice for random-effects (Di Stasio and Van de Werfhorst, 2016).<sup>8</sup>

## Results

### Main Results

Table 4 shows that age affects the candidate’s hireability score. Job candidates with a higher age have lower hireability scores, which we interpret as evidence for age discrimination. The magnitude of the age effect is large: model M4a shows that a 4-year increase in age reduces the hireability score by 0.4. Descriptive statistics show



**Figure 1.** Hireability score (1–10) per age group, age modelled linear and non-linear

that compared to the average score of 5.26, an average 43-year-old candidate scores 6.15 on the 1-10 scale, whereas an average 63-year-old only scores 4.31. Model M4b and Figure 1 show that the age effect approximates a linear relationship reasonably, although the difference between 51 and 55 and between 55 and 59 is particularly large.<sup>9</sup> For the purpose of parsimony, we treat the age effect as linear.

To test the robustness of the age effect, we compared various subsamples based on several managerial characteristics (see Appendix Table A1). Candidate age is somewhat less important for older managers, male managers, and managers who correctly identified the purpose of the vignette experiment in the post-experiment questions; however, for all groups, the age effect is statistically significant, and between-group differences are regularly not statistically significant. Managerial time stress, experience in hiring decisions, difficulty with making the rating decisions, and opinions about vignette realism did not shape the effect size of candidate age. We also compared subsamples based on organization characteristics: candidate age was most important in Eastern Europe, the Transportation sector, departments

with few high skilled workers, and larger departments; the proportion of older workers did not matter. Country-differences seem to some extent to be related to average effective retirement ages (AERA) (see Appendix Table A2): discrimination is highest in countries with the lowest AERA (Bulgaria and Hungary) and weakest in countries with the highest AERA (Sweden and United Kingdom). Although the level of age discrimination varies, age discrimination strongly affects candidates' hireability score in all subgroups, supporting Hypothesis 1.

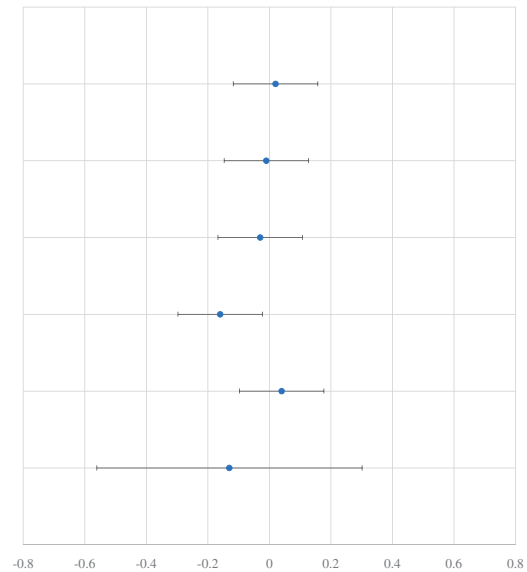
In Table 5, we test whether stereotype-confirming and stereotype-rejecting information differently moderates the impact of age (Hypothesis 2). For each of the six stereotypes, the left-hand column presents the effect of positive information compared to negative information on the hireability score. For all six indicators, signalling higher productivity increases the hireability score. The effect sizes vary strongly between the six indicators; on a ten-point scale, positive information about candidate motivation increases the hireability score with 0.36 (compared to negative information), positive information about retirement increases this with 0.67. Possibly, managers attach greater value to positive scores for some characteristics than for others; alternatively, the values chosen for some indicators may be more extreme than others.

In the right-hand column, we estimate the interaction between age and the positive information for each productivity indicator, which is also plotted in Figure 2. For training, experience, motivation, health, and retirement intentions, the interaction is not statistically significant. The direction of these interactions is not consistent either. For performance, it is significant ( $P = 0.02$ ), but the coefficient is negative rather than positive. The effect size is rather small for most interactions, it is only meaningful for retirement and performance, with over one-third of the age main effect. Finally, sensitivity analyses (presented below) indicate that the performance interaction is not very robust. Considering the inconsistent directions, mostly absent statistical significance and the sensitivity analyses, we reject Hypothesis 2: we do not find evidence that age discrimination is lower when ageist stereotypes regarding productivity are debunked.

### Sensitivity Analyses

To assess whether modelling choices influenced the results, we performed various sensitivity analyses. They are described below; tables are available in the Supplementary data.

First, Table 4 indicated that the age effect was not perfectly linear [as also indicated by Carlsson and Eriksson (2017), although at different ages]. Possibly,



**Figure 2.** Interaction effects age and positive information with 95% confidence intervals

the interaction only exists for certain ages, and it is suppressed by the other ages. Hence, Table 5 was replicated using two non-linear operationalization of age. One operationalization was by interacting dummy variables for each age category (reference category: age = 43) with the positive information dummies, like in Table 5. No interaction term was statistically significant. Alternatively, we used a threshold-operationalization: Table 5 was replicated, replacing the linear age variable by an age dummy in which 43 was considered young, and 47+ was considered old. It was also replicated for the four other possible cut-off points. The interactions were nearly always insignificant, except for the negative interaction between age and performance for cut-off points 59 and 63, comparable to the results in Table 5.

Second, as mentioned earlier, the level of age discrimination was lower for managers who correctly suspected that we studied the importance of age, compared to managers who suspected a different purpose or skipped the expected purpose question. In the same sense, social desirability might induce them to attach greater value to the information. However, replicating Table 5 in subgroups based on purpose did not change our conclusions.

Third, vignette studies are arguably best at capturing inclined behaviour if the choices are comparable to choices that the respondents are familiar with. Therefore, Table 5 was replicated for subgroups based on three post-experiment questions on the extent to which respondents believed the vignettes were easy or hard to rate, were realistic, matched their experience.



**Table 5.** Maximum-likelihood random-effects regression of the influence of age and positive information on hireability score

	X = Training	X = Experience	X = Motivation	X = Performance	X = Health	X = Retirement
Female	0.06 (0.11)	0.05 (0.11)	0.09 (0.11)	0.10 (0.11)	0.05 (0.11)	-0.10 (0.16)
Age	-0.35*** (0.05)	-0.36*** (0.04)	-0.35*** (0.05)	-0.37*** (0.04)	-0.33*** (0.04)	-0.35*** (0.10)
Positive information on X	0.51*** (0.11)	0.51*** (0.18)	0.36* (0.17)	0.52*** (0.11)	0.57*** (0.11)	0.67*** (0.15)
Age <sup>a</sup> positive information	0.02 (0.07)	-0.01 (0.07)	-0.03 (0.07)	-0.16* (0.07)	0.04 (0.07)	-0.13 (0.22)
Constant	7.77*** (0.40)	5.24*** (0.40)	6.00*** (0.39)	5.80*** (0.39)	5.39*** (0.39)	5.73*** (0.76)
Vignette N	2,093	2,117	2,119	2,123	2,152	964
Respondent N	482	482	482	482	482	482
Respondent random effects	Yes	Yes	Yes	Yes	Yes	Yes
Saturated model	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses.

\* $P < 0.05$ ,

\*\* $P < 0.01$ ,

\*\*\* $P < .001$ ,

two-tailed test. Dependent variable: hireability score (1–10). Unstandardized coefficients.

The results were comparable to those in Table 5 in all subgroups.

Fourth, statistical discrimination is plausibly most likely to be a rational strategy among recruiters who lack the time to assess individual candidates (Birkelund, 2016). Following this reasoning, the interaction between age and productivity indicators would only be expected for managers with little time stress. Subsample analysis showed that for managers with little time stress, only the interaction between age and performance was significant (again in the unexpected direction); for managers with a lot of time stress, no interaction was significant.

Fifth, since all candidate profiles were presented on the same (web) page, it could be argued that managers have ranked the candidates rather than given individual ratings. Following this reasoning, the vignettes should not be analysed as clustered observations, but as a single rank-order observation. Therefore, Tables 4 and 5 were replicated using conditional logit models (Allison and Christakis 1994). The results strongly resembled those of Tables 4 and 5.

Sixth, we combined the six productivity indicators into a single variable, 'SURPRISE SIGNAL', based on whether the information confirmed or rejected ageist stereotypes. For experience, the negative information of having little experience is stereotype-rejecting; for the other variables, the positive information is stereotype-rejecting. Each job candidate received a score for the proportion of vignette indicators containing stereotype-rejecting information ranging from 0 (all information confirmed stereotypes) to 1 (all information rejected stereotypes). This variable performed comparably to the individual items.

Seventh, a jackknife procedure was applied to ascertain that the results were not driven by an outlier country or sector. Regardless of which country or sector was omitted from the analysis, age was a comparably strong factor in employers' evaluations; coefficients for the six productivity indicators fluctuated moderately but remained similar to those in Table 5. The interaction between performance and age was found in about half of the jackknife replications.

Eighth, based on Ewens and colleagues' argument on neighbourhood composition (2014), it could be argued that the argument on noisy signals of older workers' age is particularly relevant in organizations with many older workers. However, subsample analysis of departments with few (<20%) and with many (>20%) older workers shows comparable results.

Ninth, we replicated our models using tobit regression analyses, yielding comparable results.

## Discussion and Conclusion

Although the 50+ working population steadily increases and is evermore incentivized to keep working longer, organizations are reluctant to hire older applicants. This may in part be due to statistical age discrimination, as many studies (McGregor and Gray, 2002; Loretto and White, 2006; Ng and Feldman, 2012; Principi *et al.*, 2015) show that managers frequently believe older workers are less motivated, healthy and productive. Previous statistical discrimination studies, that compared the level of discrimination between situations of a lot versus very little information, yielded inconclusive results. Therefore, we tested whether the extent of discrimination was influenced by the content, rather than the amount, of information about the productivity of older job candidates. To test our hypothesis, we used data from the European Sustainable Workforce Survey, collected at 259 organizations in nine European countries (Van der Lippe *et al.*, 2016). This provided us with vignette data from 482 department managers.

Our findings confirm the widespread presence of age discrimination. In each subpopulation in our sample, older candidates received substantially lower hireability scores. Nevertheless, the level of age discrimination varied strongly: the age penalty was around twice as high in some countries (Hungary, Bulgaria) or sectors (Transport, Manufacturing) as compared to less discriminatory strata. This may indicate that more traditional sectors and countries are more prone to age discrimination, although this finding should be tempered by the fact that the samples are not completely representative of their populations. Future research is invited to delve deeper in these country and sector differences.

Additionally, we tested whether the extent to which employers discriminated was influenced by the availability of different type of information about candidates' productivity. If managers use old age as a heuristic for assumed lower productivity, age discrimination should be lower when managers have more useful information to assess an individual's productivity. Following the study on the housing market by Ewens *et al.* (2014), we expected that stereotype-rejecting information was more influential than stereotype-confirming information, as the latter does not really supply new information. We found no support for this mechanism. Candidate age remained the central criterion for managers, and its effect was not moderated by any of the six productivity indicators (training, experience, motivation, performance, health, and retirement intentions), even though managers consistently gave better ratings to more productive candidates. The absence of support for the

hypothesis related to statistical discrimination is in line with previous studies on age discrimination (Lahey, 2008; Carlsson and Eriksson, 2017) and hints at the presence of animus-based (taste-based) discrimination: managers are less willing to hire older workers because they simply like them less than younger workers. If age discrimination is indeed primarily based on non-statistical grounds, this supports the LinkedIn policy change per 2016 of no longer showing dates of birth: no matter how impressive the other information is, older age still reduces job chances.

Alternatively, statistical discrimination effects may have eluded us. Although our findings on age, productivity indicators, and subsample analyses were generally consistent with theoretical predictions, we recommend future scholars to develop different research designs to capture the relation between information and age discrimination. First, a design in which managers were given fewer candidates and productivity indicators per decision may be more conducive to studying the role of information. It is possible that respondents experienced information overload, because eight different pieces of information were used. An overview shows that most vignette studies include six pieces at most (Clark *et al.*, 2014). Presenting six candidates on a single page may also have contributed to the overload, even though respondents in our pilot study explicitly preferred judging all candidates on a single page. This may explain why all managers rely strongly on age as heuristic: a lack of variation in level of statistical discrimination. An alternative could be to show respondents either candidates with few productivity indicators, or candidates with many productivity indicators, and compare the extent to which managers rely on age between these groups. This approximates the design of a field experiment that—perhaps in part due to this—offers a uniquely clear support for statistical discrimination theory (Ewens *et al.*, 2014). This raises the risk of reduced realism: compared to landlords selecting candidates for an apartment, employers tend to have more relevant information in actual hiring decisions, and rating almost empty profiles may not accurately reflect real-world behaviour (Neumark, 2016). Still, we invite future scholars to develop research designs that do vary in relevant information that can also be absent without damaging the realism, such as experimentally varying the presence of recommendation letters.

Second, since our vignettes concern ‘typical jobs’ at heterogeneous organizations, we selected six frequently mentioned age-stereotypes that are important to all jobs. A more homogeneous sample would have lower generalizability, but would allow including stereotype-based indicators that are particularly relevant for specific jobs,

such as adaptability, ambition, creativity and interest in new technology (Gringart *et al.*, 2005; Carlsson and Eriksson, 2017). A more homogenous sample would also attend to the complication that jobs differ in terms of ‘ideal age’ (Perry and Bourhis, 1998). Alternatively, the difference between stereotype-rejecting and stereotype-confirming of the current indicators could be made more pronounced. The differences were chosen to generate plausible profiles but could be enlarged to increase variance. Since pension ages vary across countries and sectors, the retirement intention could be operationalized to mirror age steps: what would be the difference between two otherwise identical candidates, one 59 years old and one 63 years old, both intending to retire in 3 years?

Third, although our age range was larger and more gradual than in most vignette studies, it may be desirable to study a larger range of ages also including ages below 43. The belief of what is ‘old’ may vary between countries, sectors and types of jobs, and people in their forties may already experience age discrimination in hiring decisions (Büsch *et al.*, 2009; Ahmed, Andersson and Hammarstedt, 2012; Riach, 2015; Carlsson and Eriksson, 2017).

Fourth, we used a random-with-boundaries, orthogonal method for constructing the candidate vignettes. While these are commonly used to study the hiring of older workers (Di Stasio, 2014), a D-efficient design such as a Bayesian efficient design may be preferable (De Bekker-Grob, Ryan and Gerard, 2012; Clark *et al.*, 2014; Dülmer, 2016); we invite future researchers, where feasible, to use these more efficient approaches as they plausibly would yield smaller standard errors.

In this study, it was evident that although the level of age discrimination varies based on country, sector, organizational, and managerial characteristics, the variation is between bad and worse: candidates substantially older than fifty need exceptionally positive characteristics to receive a high hireability score. Even if managers know a lot about the productivity of a candidate, the candidate likely still faces age discrimination.

## Notes

- 1 We compared vignette respondents with those who only completed the survey. At  $P < 0.05$ , there were no significant differences between respondents and non-respondents; at  $P < 0.10$ , non-respondents were on average 1.1 year younger than respondents. As sensitivity analyses indicate, lower respondent age is associated with a higher level of age discrimination

but not with the relation between age and other characteristics. If anything, we slightly underestimate the age discrimination effect.

- 2 The relation between inclined behaviour and actual behaviour is not beyond controversy (Pager and Quillian, 2005). An important counterargument is that of social desirability, which is discussed below.
- 3 Abreu (1999), for example, argues that different findings in studies on (racial) discrimination are primarily caused by whether respondents were aware of the topic of the study. Despite the vignette design and the anonymity, social desirability could still be present albeit to a lesser extent. Sensitivity analyses showed that respondents who correctly identified the purpose of the study discriminated less than respondents who did not, yet the relation between age and other characteristics is comparable.
- 4 Employers evidently also consider other characteristics, yet these are often less applicable to older workers, or too function specific to be used in a study in six sectors in nine countries. For example, during decades of work experience, the value of educational attainment is gradually replaced by more recent experience, performance and training.
- 5 Since it is unlikely that employers seriously consider candidates who perform below average, the contrast is between average and above-average.
- 6 As with PERFORMANCE, a candidate scoring ‘not motivated’ would not be considered regardless of the other variables.
- 7 This also renders it impossible to analyse the influence of the presence of information: the absence of information for a specific characteristic cannot be studied independently from the presence of information for other characteristics. To compare presence with absence of information, one would have to vary the amount of information given about each candidate. However, this may come at the price of reduced realism, as hiring decisions normally involve a substantial amount of information about each candidate (Neumark, 2016).
- 8 Compared to fixed-effects and OLS models, coefficients are highly similar, and our conclusions are the same.
- 9 Alternating the reference category shows that the difference between adjacent age categories is significant between 51 and 55 and between 55 and 59 (at  $P < 0.001$ ), and not significant for the other adjacent categories (at  $P < 0.05$ ). Each age category is significantly different from each non-adjacent category (at  $P < 0.01$ , often at  $P < 0.001$ ).

## Supplementary Data

Supplementary data are available at *ESR* online.

## Funding

This work was supported by the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 340045; the European Commission (H2020 649255), and The Netherlands Organization for Scientific Research (NWO), (016.Vidi.185.041). The funding bodies were not involved in the study design; in the collection, analysis, and interpretation of data; in the writing of the article; and in the decision to submit the article for publication.

## References

- Abreu, J. M. (1999). Conscious and nonconscious African American stereotypes: impact on first impression and diagnostic ratings by therapists. *Journal of Consulting and Clinical Psychology*, *67*, 387–393.
- Ahmed, A. M., Andersson, L. and Hammarstedt, M. (2012). Does age matter for employability? A field experiment on ageism in the Swedish labour market. *Applied Economics Letters*, *19*, 403–406.
- Aigner, D. J. and Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *ILR Review*, *30*, 175–187.
- Allison, P. D. and Christakis, N. A. (1994). Logit models for sets of ranked items. *Sociological Methodology*, *24*, 199–228.
- Arrow, K. (1973). The theory of discrimination. *Discrimination in Labor Markets*, *3*, 3–33.
- Becker, G. S. (1971). *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Bendick, M., Brown, L. E. and Wall, K. (1999). No foot in the door: an experimental study of employment discrimination against older workers. *Journal of Aging and Social Policy*, *10*, 5–23.
- Birkelund, G. E. (2016). Rational laziness—when time is limited, supply abundant, and decisions have to be made. *Analyse & Kritik*, *01*, 203–225.
- Büsch, V., Dahl, S.-Å. and Dittrich, D. A. V. (2009). An empirical study of age discrimination in Norway and Germany. *Applied Economics*, *41*, 633–651.
- Buskens, V. (1999). *Social Networks and Trust*. Amsterdam: Thela Thesis.
- Carlsson, M. and Eriksson, S. (2017). The effect of age and gender on labor demand—evidence from a field experiment. Linnaeus University Working Papers, 1–55
- Clark, M. D. et al. (2014). Discrete choice experiments in health economics: a review of the literature. *PharmacoEconomics*, *32*, 883–902.
- Daniel, K. and Siebert, W. S. (2005). Does employment protection reduce the demand for unskilled labour?. *International Economic Journal*, *19*, 197–222.

- De Bekker-Grob, E. W., Ryan, M. and Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health Economics*, 21, 145–172.
- Di Stasio, V. (2014). Education as a signal of trainability: results from a vignette study with Italian employers. *European Sociological Review*, 30, 796–809.
- Di Stasio, V. and Gërkhani, K. (2015). Employers' social contacts and their hiring behavior in a factorial survey. *Social Science Research*, 51, 93–107.
- Di Stasio, V. and Van de Werfhorst, H. G. (2016). Why does education matter to employers in different institutional contexts? A vignette study in England and the Netherlands. *Social Forces*, 95, 77–106.
- Dülmer, H. (2016). The factorial survey: design selection and its impact on reliability and internal validity. *Sociological Methods & Research*, 45, 304–347.
- Duncan, C. and Loretto, W. (2004). Never the right age? Gender and age-based discrimination in employment. *Gender, Work and Organisation*, 11, 95–115.
- Ewens, M., Tomlin, B. and Wang, L. C. (2014). Statistical discrimination or prejudice? A large sample field experiment. *Review of Economics and Statistics*, 96, 119–134.
- Fang, H. and Moro, A. (2011). Theories of statistical discrimination and affirmative action. A survey. In Benhabib J., Bisin A. and Jackson M. O. (Eds.), *Handbook of Social Economics*. Amsterdam: Elsevier, pp. 133–200.
- Finkelstein, L. M., Higgins, K. D. and Clancy, M. (2000). Justifications for ratings of older and young job applicants: an exploratory content analysis. *Experimental Aging Research*, 26, 3, 263–83.
- Gray, L. and McGregor, J. (2003). Human resource development and older workers: stereotypes in New Zealand. *Asia Pacific Journal of Human Resources*, 41, 338–353.
- Gringart, E., Helmes, E. and Speelman, C. P. (2005). Exploring attitudes towards older workers among Australian employers. *Journal of Aging and Social Policy*, 17, 85–103.
- Hainmueller, J., Hangartner, D. and Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 2395–2400.
- Henkens, K. (2005). Stereotyping older workers and retirement: the managers' point of view. *Canadian Journal on Aging/La Revue canadienne du vieillissement*, 24, 353–366.
- Karpinska, K., Henkens, K. and Schippers, J. (2011). The recruitment of early retirees: a vignette study of the factors that affect managers' decisions. *Ageing and Society*, 31, 570–589.
- Karpinska, K., Henkens, K. and Schippers, J. (2013a). Hiring retirees: impact of age norms and stereotypes. *Journal of Managerial Psychology*, 28, 886–906.
- Karpinska, K., Henkens, K. and Schippers, J. (2013b). Retention of older workers: impact of managers' age norms and stereotypes. *European Sociological Review*, 29, 1323–1335.
- Kluge, A. and Krings, F. (2008). Attitudes towards older workers and Human Resource practices. *Swiss Journal of Psychology*, 67, 61–64.
- Lahey, J. N. (2008). Age, women, and hiring: an experimental study. *Journal of Human Resources*, 43, 30–56.
- List, J. A. (2013). The nature and extent of discrimination in the marketplace: evidence from the field. *The Quarterly Journal of Economics*, 119, 49–89.
- Loretto, W. and White, P. (2006). Employers' attitudes, practices and policies towards older workers. *Human Resource Management Journal*, 16, 313–330.
- Martin, G. et al. (2014). In the name of meritocracy: managers' perceptions of policies and practices for training older workers. *Ageing and Society*, 34, 992–1018.
- McGregor, J. and Gray, L. (2002). Stereotypes and older workers: the New Zealand experience. *Social Policy Journal of New Zealand*, 18, 163–177.
- Neumark, D. (2016). *Experimental Research on Labor Market Discrimination*. Cambridge, USA: National Bureau of Economic Research.
- Neumark, D., Burn, I. and Button, P. (2016). Evidence from lab and field experiments on discrimination. Experimental age discrimination evidence and the Heckman critique. *American Economic Review*, 106, 303–308.
- Ng, T. W. H. and Feldman, D. C. (2012). Evaluating six common stereotypes about older workers with meta-analytical data. *Personnel Psychology*, 65, 821–858.
- OECD (2018). Ageing and employment policies—statistics on average effective age of retirement. Available from: <https://www.oecd.org/els/emp/average-effective-age-of-retirement.htm> [Accessed 29 July 2019].
- Oude Mulders, J. et al. (2018). Managers' interview invitation decisions about older job applicants: human capital, economic conditions and job demands. *Ageing and Society*, 38, 839–864.
- Oude Mulders, J. et al. (2014). How likely are employers to re-hire older workers after mandatory retirement? A vignette study among managers. *De Economist*, 162, 415–431.
- Pager, D. and Karafin, D. (2009). Bayesian bigot? Statistical discrimination, stereotypes, and employer decision making. *The Annals of the American Academy of Political and Social Science*, 621, 70–93.
- Pager, D. and Quillian, L. (2005). Walking the talk? What employers say versus what they do. *American Sociological Review*, 70, 355–380.
- Perry, E. L. and Bourhis, A. C. (1998). A closer look at the role of applicant age in selection decisions. *Journal of Applied Social Psychology*, 28, 1670–1697.
- Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, 62, 659–661.
- Principi, A., Fabbietti, P. and Lamura, G. (2015). Perceived qualities of older workers and age management in companies: does the age of HR managers matter? *Personnel Review*, 44, 801–810.
- Riach, P. A. (2015). A field experiment investigating age discrimination in four European labour markets. *International Review of Applied Economics*, 29, 608–619.
- Rich, J. (2014). What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000. *IZA Discussion Paper*, 8584, 1–71.

- Rosen, B. and Jerdee, T. H. (1976). The influence of age stereotypes on managerial decisions. *Journal of Applied Psychology*, 61, 428–432.
- Rupp, D. E., Vodanovich, S. J. and Credé, M. (2006). Age bias in the workplace: the impact of ageism and causal attributions. *Journal of Applied Social Psychology*, 36, 1337–1364.
- Schaeffer, M., Höhne, J. and Teney, C. (2015). Income advantages of poorly qualified immigrant minorities: why school dropouts of Turkish origin earn more in Germany. *European Sociological Review*, 32, 93–107.
- Snape, E. and Redman, T. (2003). Too old or too young? The impact of perceived age discrimination. *Human Resource Management Journal*, 13, 78–89.
- Stangor, C. and McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: a review of the social and social developmental literatures. *Psychological Bulletin*, 111, 42–61.
- Steinberg, M. et al. (1996). Attitudes of employees and employers towards older workers in a climate of anti-discrimination. *Australasian Journal on Ageing*, 15, 154–158.
- Taylor, P. and Walker, A. (1998). Policies and practices towards older workers: a framework for comparative research. *Human Resource Management Journal*, 8, 61–76.
- Tijssen, L. et al. (2019). Discrimination against Turkish minorities in Germany and the Netherlands: field experimental evidence on the effect of diagnostic information on labour market outcomes. *Journal of Ethnic and Migration Studies*, 1–18.
- Van Dalen, H. P., Henkens, K. and Schippers, J. (2007). *Oudere werknemers door de lens van de werkgever*, Report 74. Den Haag: Netherlands Interdisciplinary Demographic Institute.
- Van der Lippe, T. et al. (2016). *European Sustainable Workforce Survey [ESWS]*. Utrecht: Utrecht University.
- Vickerstaff, S., Cox, J. and Keen, L. (2003). Employers and the management of retirement. *Social Policy & Administration*, 37, 271–287.
- Wallander, L. (2009). 25 years of factorial surveys in sociology: a review. *Social Science Research*, 38, 505–520.
- Weiss, E. M. and Maurer, T. J. (2004). Age discrimination in personnel decisions: a reexamination. *Journal of Applied Social Psychology*, 34, 1551–1562.

**Jelle Lössbroek** is a postdoctoral researcher in the Department of Sociology at Utrecht University, a member of the Research School ICS and a management committee member of COST Action IS1409: Gender, health and extending working lives. After completing his master's degree in Political Science and in Social & Cultural Science (both cum laude), he wrote his PhD dissertation (defense in January 2019) 'Turning grey into gold: Employer–employee interplay in an ageing workforce.' His labour market research focuses on older workers and on technological change.

**Bram Lancee** is associate professor of Sociology at the University of Amsterdam. His current research interests include social capital, ethnic minorities and the labour market, inequality and social participation, attitudes towards immigration, and ethnic diversity in neighbourhoods. His work has been published in journals such as *Social Forces*, *European Sociological Review*, *International Migration Review*, *Journal of Ethnic and Migration Studies*, and *Social Science Research* and as editor of a volume with Routledge.

**Tanja van der Lippe** is a professor of Sociology at Utrecht University, head of the Department of

Sociology, and chair of the Research School ICS. Her research interests are in the area of work–family linkages, for which she has received a number of large-scale grants from Dutch and European Science Foundations. She received an ERC Advanced Grant for her research into 'Investments in a sustainable workforce in Europe'. She is an elected member of the Royal Netherlands Academy of Arts and Sciences, the Royal Holland Society of Sciences and Humanities, and the European Academy of Sociology.

**Joop Schippers** is a professor of Labour Economics at Utrecht University. He completed his PhD thesis on gender inequality in the labour market in 1987. He is a leading Dutch expert on gender and age-related labour market issues and has published numerous books and articles on gender differences, human capital investments, labour market relations, and labour market flexibility and organizational/employers' behaviour with respect to women and older workers. He is a member of the Board of Supervisors of the University of Humanistic Studies in Utrecht and an affiliated researcher of the Netherlands Interdisciplinary Demographic Institute in The Hague.

Appendix

Figure A1. Example of the vignettes

Imagine yourself in the following situation:

Your department has two vacancies. Both vacancies are for jobs that are typical in your department with a matching salary. Six external candidates have applied. You have some information about the candidates in their current jobs, but you do not have the same type of information about each candidate. For each candidate, please indicate how likely it is that you will hire her or him.

<p><b>Candidate A</b></p> <ul style="list-style-type: none"> <li>• Man, 43 years old</li> <li>• Has considerable experience in comparable positions</li> <li>• Average performance in recent assessments</li> <li>• Appears to be reasonably motivated</li> <li>• Took 15 sick days in the past year</li> </ul>	<p><b>Candidate B</b></p> <ul style="list-style-type: none"> <li>• Man, 59 years old</li> <li>• Above-average performance in recent assessments</li> <li>• Had various forms of training relevant to this job</li> <li>• Appears to be reasonably motivated</li> <li>• Intends to retire at the statutory retirement age</li> </ul>	<p><b>Candidate C</b></p> <ul style="list-style-type: none"> <li>• Woman, 51 years old</li> <li>• Has considerable experience in comparable positions</li> <li>• Average performance in recent assessments</li> <li>• Had various forms of training relevant to this job</li> <li>• Appears to be highly motivated</li> </ul>
<p><b>Candidate D</b></p> <ul style="list-style-type: none"> <li>• Woman, 55 years old</li> <li>• Has little experience in comparable positions</li> <li>• Above-average performance in recent assessments</li> <li>• Had various forms of training relevant to this job</li> <li>• Took 2 sick days in the past year</li> </ul>	<p><b>Candidate E</b></p> <ul style="list-style-type: none"> <li>• Man, 63 years old</li> <li>• Average performance in recent assessments</li> <li>• Had various forms of training relevant to this job</li> <li>• Appears to be reasonably motivated</li> <li>• Intends to retire early</li> </ul>	<p><b>Candidate F</b></p> <ul style="list-style-type: none"> <li>• Woman, 47 years old</li> <li>• Has considerable experience in comparable positions</li> <li>• Had no relevant training last year</li> <li>• Appears to be highly motivated</li> <li>• Took 15 sick days in the past year</li> </ul>

How likely is it that you would hire these candidates?

	extremely unlikely										extremely likely
	1	2	3	4	5	6	7	8	9	10	
Candidate A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
Candidate B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
Candidate C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
Candidate D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

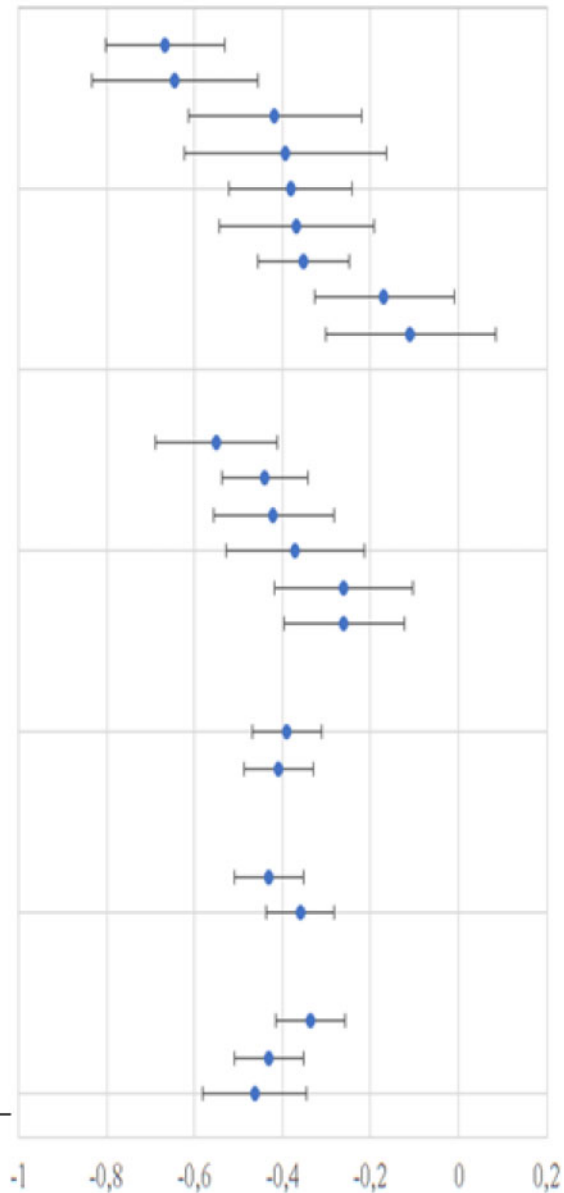
Note: Nearly each respondent received a different vignette set.

Table A1a. Breakdown of the linear age effect for sub-groups, unstandardized coefficients

	Age effect	N
<i>By country</i>		
Bulgaria	-.67***	312
Hungary	-.64***	264
Finland	-.42***	246
Germany	-.39***	234
Portugal	-.38***	402
Spain	-.37***	294
Netherlands	-.35***	744
Sweden	-.17+	210
UK	-.11	186
<i>By sector</i>		
Transport	-.56***	426
Manufacturing	-.44***	846
Health Care	-.42***	480
Financial Services	-.37***	390
Telecommunication	-.26**	330
Higher Education	-.26***	420
<i>% 50+ in department</i>		
Less than 20%	-.39***	1,506
More than 20%	-.41***	1,386
<i>% high-skilled in department</i>		
Less than 50%	-.43***	1,512
More than 50%	-.36***	1,380
<i>Size of department</i>		
Small (1-10)	-.34***	1,122
Medium (11-49)	-.43***	1,290
Large (50+)	-.46***	480

+ =  $p < .10$ , \* =  $p < .05$ , \*\* =  $p < .01$ ,

\*\*\* =  $p < .001$ . Table A1 replicates model Table 4, model M4a, each time splitting the sample into subsamples based on various variables; estimates for constant and gender are not presented due to space limitations.





**Table A1b.** Breakdown of the linear age effect for subgroups, unstandardized coefficients

	Age effect	N	
<i>Candidate profiles are realistic</i>			
(strongly) disagree	-.42***	654	
neutral	-.39***	780	
(strongly) agree	-.40***	1,458	
<i>I have experience taking similar decisions</i>			
(strongly) disagree	-.38***	582	
neutral	-.38***	594	
(strongly) agree	-.42***	1,716	
<i>Decisions were easy to take</i>			
(strongly) disagree	-.37***	1,116	
neutral	-.47***	774	
(strongly) agree	-.38***	1,002	
<i>Assumed purpose of the study [author's coding]</i>			
Empty	-.49***	876	
Unrelated to age	-.41***	888	
Related to age	-.33***	1,110	
<i>Manager age</i>			
49 or younger	-.43***	1,890	
50 or older	-.34***	1,002	
<i>Manager gender</i>			
Male	-.37***	1,770	
Female	-.45***	1,110	
<i>Manager time stress</i>			
Often / Always	-.42***	1,308	
Regularly/sometimes/never	-.38***	1,584	

+ =  $p < .10$ , \* =  $p < .05$ , \*\* =  $p < .01$ ,  
 \*\*\* =  $p < .001$ . Table A1 replicates model  
 Table 4, model M4a, each time splitting the  
 sample into subsamples based on various  
 variables; estimates for constant and gender  
 are not presented due to space limitations.

-0,8      -0,6      -0,4      -0,2      0

**Table A2.** Average effective retirement age in each country per 2016

	Men	Women	Average
Bulgaria	63.1	60.7	61.9
Hungary	63.6	60.7	62.2
Spain	62.2	62.6	62.4
Netherlands	63.5	62.3	62.9
Finland	63.2	62.6	62.9
Germany	63.3	63.2	63.3
Portugal	62.6	64.9	63.8
United Kingdom	65.6	63.2	64.4
Sweden	65.8	64.6	65.2
All nine countries	63.7	62.8	63.2

Source: OECD (2018).