RESEARCH ARTICLE

WILEY

Age categorization and stereotyping at work

Diane Pecher ⓘ | **Heleen van Mierlo** ⓘ | **C. M. Romers** | **K. Feersma Hoekstra** | **Naomi Possel** | **Samantha Bouwmeester** ⓘ

Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands

Correspondence
Diane Pecher, Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Postbus 1738, Rotterdam 3000 DR, The Netherlands.
Email: pecher@essb.eur.nl

Abstract

Age stereotyping at the workplace can lead to discrimination. We investigated stereotype influence on recall. In the who-said-what paradigm participants studied older and younger speakers making stereotypical and counterstereotypical statements about their work, followed by memory tests for the statements and speakers. Statements were more likely attributed to a speaker from the same age category than the other age category, replicating earlier findings. Stereotypicality had no effect on statement recognition but participants were more likely to guess a stereotypical speaker even though speakers made equal numbers of stereotypical and counterstereotypical statements (Experiments 1 and 3). When speakers made more counterstereotypical statements (Experiment 2) this bias was reduced but less so than the actual proportion of stereotypical statements warranted. Speaker judgments were also influenced by prior stereotypes. We conclude that participants used prior stereotypes to supplement recall from memory. Measures to reduce stereotypes should include procedures to improve memory accuracy.

KEYWORDS
age stereotypes, source guessing, source monitoring, stereotypes, who-said-what paradigm

1 | AGE CATEGORIZATION AND STEREOTYPING AT WORK

A common stereotype about older people is that they are more costly employees than younger people. Not only because their salaries on average are higher than those of younger people, but also, and perhaps even more important, because their performance is assumed to be poorer (Finkelstein et al., 2013; Ng & Feldman, 2012). People may think that older people have lost their ambitions, are unable or unwilling to learn new things, have more health problems, and have more problems maintaining a balance between work and family, among other things. Although there seems to be little truth in these stereotypes (Ng & Feldman, 2012) they are still commonly held (Kite et al., 2005; Zaniboni et al., 2019) and influence decision making in

areas such as hiring practices and training and development (Derous & Decoster, 2017; Finkelstein et al., 2013; Hebl et al., 2020; Posthuma & Campion, 2009; Rupp et al., 2006; but see Murphy & DeNisi, 2022). In the current study we investigated how memory limitations may contribute to incorrect use of stereotypes in remembering and evaluating people.

Stereotyping is the result of a more general cognitive process of categorization and representation of categories as prototypes. When people encounter an exemplar of a particular category, they activate knowledge about that category that represents its properties. Such knowledge may be based on previously encountered exemplars and on explicit statements about the category. For example, when people encounter an object that looks like a tree, they may activate previous experiences with trees to induce that the object will be sturdy and

that there will be birds in it, and they may activate previously told knowledge to induce that the object needs water and sunlight. Without such inductive skills it would be very cumbersome to deal with novel situations, such as deciding whether a novel object that looks like a chair can be sat on or whether an animal is likely to attack. Thus, people are very good at using previously acquired knowledge to make predictions about novel situations by attributing prototypical features to new exemplars of a category. While this process is usually quite helpful, it also leads to problems. Category exemplars vary in the extent to which they share features with the prototype, and thus predictions might be wrong. This is often the case with social stereotypes, not only because a prototype might contain incorrect information (e.g., "Older people are less productive than younger people") but also because even accurate prototypical information will not be true for each individual.

In the present study we investigated how age categorization and age stereotypes might play a role in the workplace. The who-said-what paradigm (Taylor et al., 1978) is a useful tool to measure how memory failures contribute to categorization and stereotyping. In this paradigm, participants observe a discussion between a group of speakers. Typically, the speakers are chosen such that they could be categorized into two groups, for example younger and older. After the discussion has been observed, participants are asked to recall which speaker said which statement. Many studies have shown that when participants do not recall correctly who said what, they are more likely to erroneously recall a wrong speaker from the same group as the original speaker than to recall a wrong speaker from the other group (Blanz & Aufderheide, 1999; Klauer et al., 2003; Klauer & Wegener, 1998; Stangor et al., 1992; Taylor et al., 1978; van Twuyver, 1996; Wegener & Klauer, 2004). In other words, participants are more likely to make within group errors than between group errors. For example, if a statement was made by a woman, participants who make an error in recalling who made the statement are more likely to choose another woman than one of the men. This suggests that participants categorize speakers into groups, and if they do not remember the identity of the speaker, they might still remember the speaker's category.

The interesting aspect of this paradigm is that participants are not asked explicitly about their social categorization and stereotyping, and that, therefore, effects of social desirability seem unlikely (Stangor et al., 1992; Wegener & Klauer, 2004). Besides being an elegant tool to measure social categorization, the paradigm resembles everyday work situations, such as team meetings or selection procedures, where managers observe many people and make continuous, often unconscious, judgments about them based on these observations. Here also it seems likely that categorization processes occur, such that people implicitly use stereotypes when making impactful decisions based on who said or did what. Here we investigate how errors in recall may result in stereotypical decisions.

The who-said-what paradigm has revealed that various types of social grouping influence source memory, suggesting that participants spontaneously categorize speakers based on gender (Blanz, 1999; Stangor et al., 1992; Taylor et al., 1978; van Twuyver, 1996), ethnicity (Stangor et al., 1992; Taylor et al., 1978), university major (Blanz & Aufderheide, 1999), or occupation (Nolan et al., 1999). Although it seems likely that age is also used to categorize people, only a few studies have investigated age categorization in the who-said-what paradigm. In two studies (Klauer et al., 2003; Pietraszewski et al., 2015), age was crossed with another dimension such as gender or political affiliation. These studies showed that age categorization occurs although in some cases it may depend on whether age is relevant for the discussion topic. Karasawa et al. (2014) found that Japanese but not Italian participants were sensitive to age. In their study, however, the age categories were rather close (16–18 vs. 24–26 years old). The participants' ages were between the ages of these two categories. For Japanese participants whether a person is younger or older than themselves was a culturally significant distinction, with linguistic implications. For example, for younger persons, a familiar form is appropriate, but for older persons a polite form is appropriate, even if the age difference is just 1 year. For the Italian participants a few years difference in age was not a significant distinction for choosing how to address the other person. Thus, Karasawa et al. results suggest that age categorization happens if it is relevant to the participant. However, in their study the absolute age difference was very small, and the relevant dimensions for categorization may have been younger/older than the participant rather than absolute age.

In the present study, we were interested to see if western European participants would show age categorization for speakers whose ages were further apart than in the Karasawa et al. study. Unlike Japanese, Europeans may not be sensitive to small age differences but the large number of studies on age discrimination and stereotypes suggest that age is relevant and plays a role in many situations, including the workplace. In particular, we investigated if age categorization resulted in remembering the speakers as more stereotypical than they were. Speakers and statements may have high or low normative fit. If speakers make mostly statements that fit with stereotypical expectations for their category, the normative fit is high. Several studies have shown that the tendency to confuse speakers from the same category rather than from different categories is stronger if speakers make many statements that fit with the stereotype of their category (i.e., high normative fit) than if they make few of such stereotypical statements (i.e., low normative fit). The more stereotypical a speaker's statements, the more salient their category membership and the stronger the social categorization effect (Blanz & Aufderheide, 1999; Klauer & Ehrenberg, 2005; van Twuyver, 1996; Wegener & Klauer, 2004). However, this finding may simply indicate that participants noticed such correlation in the experiment because stereotypicality was manipulated at the group level. That is, in conditions with larger stereotype fit, all speakers made more stereotypical than counterstereotypical statements. Participants may use this information to guess speaker category not because they are using prior stereotypes but because they noticed that a specific type of statement was always or often made by a speaker from a specific group. This has been suggested by the probability-matching account (Spaniol & Bayen, 2002). Moreover, greater fit in the experiment

increases similarity between group members and decreases similarity between groups, which may promote categorization (Wegener & Klauer, 2004). Although they did not manipulate stereotypicality systematically, Klauer et al. (2003) found that participants may have used statement content to guess speaker category. Such reconstructive memory processes will lead to memory outcomes that are more stereotypical than the information that was originally presented to the participants. Therefore, in the present study we manipulated stereotypicality at the statement level. Each speaker made both stereotypical and counterstereotypical statements and we investigated how age-stereotypicality of the individual statement affected recall performance for the speaker.

Performance in source memory tasks such as the who-said-what paradigm is the result of a combination of memory sensitivity and bias (see Bayen et al., 1996; DeCarlo, 2003; Dodson et al., 2008; Klauer, 2010; Klauer & Wegener, 1998; Pazzaglia et al., 2013; Schütz & Bröder, 2011 for formal models). Stereotypicality may contribute to the pattern of errors in the who-said-what paradigm by influencing either of these processes. First, stereotypicality may affect memory sensitivity for the statement itself or for the speaker of the statement. On the one hand, stereotypical information may fit better into existing memory structures and as a result may be easier to encode than counterstereotypical information. Consistent with this idea, several studies have shown that memory, especially recall, is better for information that is consistent with prior knowledge than for information that is inconsistent with it (Brewer & Treyens, 1981; Cantor & Mischel, 1979; Nolan et al., 1999). On the other hand, because counterstereotypical information is unexpected and therefore more salient, people may process it more deeply during encoding than stereotypical information. Supporting this view, many studies have demonstrated better memory for inconsistent information (Bardach & Park, 1996; Bell et al., 2015; Ehrenberg & Klauer, 2005; Graesser et al., 1980; Kleider et al., 2008; Küppers & Bayen, 2014; Pezdek et al., 1989; Wegener & Klauer, 2004).

The memory benefit for counterstereotypical information will likely depend on the strength of stereotype violation (Schaper et al., 2019). For example, Ehrenberg and Klauer (2005) presented statements about two persons who were described as a friendly, helpful social worker and an unfriendly, aggressive skinhead. Statements describing friendly behavior would therefore strongly violate expectancies about the person previously described as unfriendly. In a typical workplace situation, however, behaviors and opinions might be considered more typical for older or younger employees, but not constitute strong violations of expectancies. Thus, the memory benefit of expectancy violations might be absent in this case.

A second way in which stereotypicality may contribute to errors is by influencing guessing for the speaker category (Ehrenberg & Klauer, 2005; Klauer & Ehrenberg, 2005; Wegener & Klauer, 2004). In that case, people use prior knowledge to fill gaps in their memory for the specific study items (Hemmer & Steyvers, 2009; Huttenlocher et al., 2000; Poirier et al., 2017). In fact, guessing biases might completely explain better memory performance for stereotypical information in some paradigms (Hicks & Cockman, 2003; Schaper

et al., 2019; Sherman & Bessenoff, 1999). When people use prior knowledge or schemas, they should be more likely to guess a stereotypical than a counterstereotypical category. In an experiment with only two speakers, Ehrenberg and Klauer (2005) showed that speaker memory was better for inconsistent information but speaker guessing was better for consistent information. Compelling evidence for such a guessing bias is provided by studies in which participants hear about the speaker's category only after they have studied a list of items (Hicks & Cockman, 2003; Kuhlmann et al., 2016; Marsh et al., 2006; Pyszczynski et al., 1987). For example, Kuhlman et al. presented statements made by two speakers, half stereotypical and half counterstereotypical. The speakers differed in age, but their ages were only revealed after the study phase. Stereotypicality therefore had no effect during the encoding phase, and thus the stereotype effect could only have been the result of guessing during retrieval. Marsh et al. found that bias was not affected by attention and therefore argued that the guessing bias is an automatic process. Kleider et al. (2008), see also Graesser et al. (1980) found that category guessing was increased after a delay between the presentation and the recall phase, which suggests that category guessing is the result of memory failure.

Wegener and Klauer (2004), see also Ehrenberg & Klauer (2005), Klauer & Ehrenberg (2005) showed that category guessing was higher when all category members conformed to category-consistent stereotypes (i.e., high normative fit: e.g., A homosexual person making progressive statements about gender roles) than when they did not (i.e., low normative fit: e.g., A homosexual person making conservative statements about gender roles). They also showed that speaker memory and category memory were higher under high fit than under low fit. In a final experiment, they presented mostly statements from one stereotype (politically progressive or conservative) for both speaker groups (members of opposite political parties). This way, one group of speakers had high normative fit and the other had low normative fit, but there was no correlation between statement content and group, nor a difference in within category similarity between high and low fit. In this experiment Wegener and Klauer found that category memory was higher for counterstereotypical statements than for stereotypical statements, but category guessing was higher for stereotypical than counterstereotypical statements. Thus, these results are consistent with the idea that counterstereotypical information is better encoded, possibly due to its higher salience, but that there is a guessing bias for stereotypical information, even in the absence of a true correlation (see also Küppers & Bayen, 2014). This increase in memory for counterstereotypical information should be related to the strength of the inconsistency. One could argue that in Wegener and Klauer's Experiment 4 the inconsistency was quite salient, because all members of one group produced statements that were opposite to what would be expected based on their political party affiliation. Thus, these studies suggest that category memory might be better for counterstereotypical information and category guessing might be better for stereotypical information.

In the present experiments we used a who-said-what paradigm to measure the effect of age stereotypicality in a work-related context

on speaker memory. We investigated if the recall errors that participants make in this paradigm would result in recall of the speakers that was more stereotypical than the information that was presented originally. Kuhlmann et al. (2016) found that age stereotypicality influenced guessing. In their study, however, they only had one speaker for each age group and only revealed age after the study phase. Therefore, they could not assess within group speaker confusion. In the present experiment, participants observed a group of four younger and four older speakers make several statements about their work. To prevent subtyping by gender or race (Klauer et al., 2003; Stangor et al., 1992), all speakers were white women. Each speaker made an equal number of stereotypical and counterstereotypical statements so that the experimental materials did not introduce a bias toward stereotypicality. After they had seen all statements, participants performed an old-new recognition memory task that included all statements that had been said by the speakers ("old statements") as well as an equal number of statements that had not been said by the speakers ("new statements"). This memory task required participants to indicate the speaker for all statements that they recognized as "old".

In most studies that investigated normative fit the speakers made statements that expressed opinions about topics that would be expected to be relevant for their group, such as politically progressive or conservative statements made by speakers from different political groups. Such explicit topics might prime categorization and stereotypes. In the present study we used statements that did not express opinions about a stereotype relevant topic, but speakers said various things about their work that could be considered more typical for older or younger workers. Thus, stereotypicality was implicit rather than explicit. Moreover, rather than manipulating normative fit at the group level and looking at the overall results, we investigated memory for stereotypical and counterstereotypical statements separately. This design might be closer to the situation at a workplace where people show a mix of stereotypical and counterstereotypical behavior.

As in prior studies, we expected that participants would make more within category errors than between category errors in speaker recall. In addition, we expected that the within category error rates would be higher for stereotypical than for counterstereotypical statements because participants may have a tendency to guess stereotypical speakers. For statement recognition performance, however, we had no specific expectations because prior studies show mixed results.

2 | EXPERIMENT 1

2.1 | Method

2.1.1 | Participants

Sixty-four employees at various organizations[1] participated voluntarily to help undergraduate students with their thesis project. They were recruited from the students' personal network and received no compensation in exchange for their participation. The average age was 43.0 (SD = 12.9, range = 18–65). Thirty were female and two did not disclose their gender. Sample size was determined before any data analysis. We calculated power for main effects and interactions in a within-participant 2 (stereotypicality) x^2 (error type) ANOVA. The power to obtain a significant medium size effect (partial $\eta^2 = .06$) was .99.

2.1.2 | Materials

Eight black and white photographs of four older and four younger women were used to represent the eight speakers in the group discussion. Each person wore a neutral black top and was photographed against a white background. Two example photographs are shown in Figure 1. In a separate study, 60 participants estimated the older women to be substantially older (M = 51.2, SD = 6.8) than the younger women (M = 27.1, SD = 3.5). Based on literature on age stereotypes (Ng & Feldman, 2012; Posthuma & Campion, 2009) we created 48 statements that are stereotypical for older workers and 48 statements that are stereotypical for younger workers. The statements were made about six different work-related topics: Learning and Adjustment, Technology, Health, Ambition, Work Ethics and Experience, and Energy and Impulsivity. For each topic eight statements were stereotypical for older people and eight were stereotypical for younger people. For both stereotypes we created statements with positive and statements with negative connotations. All statements are presented in Appendix A, Table A1. The statements were divided into 16 sets, each consisting of three stereotypical old and three stereotypical young statements. In each set there was one statement about each topic. Following these restrictions assignment of statements to sets was semirandom, with the additional restriction that statements within a set could not be contradictory. Each speaker made the six statements from one of the sets. Sets were counterbalanced across speakers and across studied (targets) and non-studied (distractors) conditions.

A separate group of 41 participants rated the stereotypicality of the statements. Each participant rated half of the statements on two separate 7-point likert scales (from "absoluut niet van toepassing" [absolutely unlikely] to "absoluut van toepassing" [absolutely likely]), one for old typicality and one for young typicality. They were asked whether they thought the statements would generally be considered typical for older and younger people and to judge how likely the statement would be made by an older or younger speaker. The stereotypical old statements were all rated as more typical for older (M = 5.16) than younger people (M = 3.24) and the stereotypical young statements were all rated as more typical for younger (M = 5.49) than for older people (M = 3.38).

2.1.3 | Procedure

Participants were tested individually in a quiet room near their workspace. They were told that they would see a chat discussion about

FIGURE 1 Examples of photographs used in Experiment 1.

work among a group of people. They were instructed to form an impression of the group (following Klauer & Wegener, 1998). No mention was made of the memory test that would follow. Each statement was presented in a text balloon next to the photograph of a speaker for 10 s, followed by a blank screen of 500 ms. Statements were grouped by topic, such that each speaker was presented once with a statement about the first topic, followed by each speaker presented once with a statement about the next topic, and so on, to make it look more like a real discussion. Thus, eight statements were made per topic. The order of topics was the same for all participants and the order of speakers within each topic was randomized. At the end of the study phase each of the eight speakers had made a statement about six different topics, such that in total the participant had seen 48 statements. Each speaker made three stereotypically old statements and three stereotypically young statements, such that 24 statements were presented in the stereotypical condition (e.g., a young speaker making a statement that would be considered typical for a young speaker) and 24 in the counterstereotypical condition (e.g., an old speaker making a statement that would be considered typical for a young speaker). Because statements were counterbalanced across speakers, they were also counterbalanced across stereotypicality.

After all 48 statements had been presented for study, the memory test started. All statements, 48 old (studied; that is, presented in the group "discussion") and 48 new (not studied; that is, not presented in the group "discussion") were presented one by one in random order. Below the statement were two boxes, one contained the text "old," and the other contained the text "new." Participants made their response by clicking "old" if they remembered seeing the statement in the previous phase, or "new" if they did not remember seeing the statement. After being clicked, the box turned from gray to green for 500 ms to indicate that a click had been registered. If they clicked "new," this indicated they did not remember seeing the statement and would not be able to remember speaker or category. In that case, the next statement was presented. If they clicked "old," they did

remember seeing the statement. In that case, the statement was presented again, and below the statement the eight photographs were shown. Participant clicked on the photograph of the speaker who they thought made the statement. After they clicked a photograph it increased in size for 500 ms, to show that a click was registered, after which the next statement was shown. The photographs were always presented in the same position on the screen for a particular participant, photographs were counterbalanced across positions across participants. The photographs were mixed and not grouped by age category.

After the memory test participants were asked their age, gender, and a description of the strategy they had used to perform the task.

2.2 | Results

In these studies, we report all measures, manipulations, and exclusions. The data files that were used for the analyses reported below for all experiments can be found on https://osf.io/32gt6/.

2.2.1 | Statement accuracy

We calculated the proportion correct "old" responses to the statements. Accuracy did not differ between statements made by a speaker from the stereotypical ($M = .79$, SEM $= .02$) or counterstereotypical age group ($M = .81$, SEM $= .02$), $t(63) = 1.55$, $p = .127$, $BF_{01} = 3.21$.[2] The false alarm rate was .10 (SEM $= .02$).

2.2.2 | Speaker errors

Because the number of incorrect speakers is larger between categories (4) than within categories (3, that is, excluding the correct
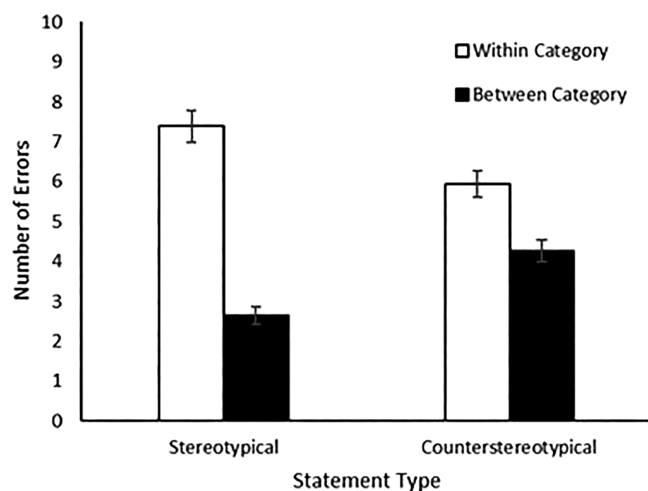
FIGURE 2 Adjusted mean number of errors in Experiment 1. Error bars represent standard error of the mean difference.

speaker), we adjusted the between category errors by multiplying the number of errors by .75 (following Taylor et al., 1978). The number of adjusted incorrect responses per condition is shown in Figure 2. A 2 (error type) by 2 (stereotypicality) ANOVA confirmed that participants made more errors within than between categories, $F(1,63)$ = 95.85, $p < .001$, partial $\eta^2 = .60$, $BF_{10} > 10,000$. The null-hypothesis that the number of errors was not influenced by stereotypicality could not be rejected and the BF also favored the H0 over H1, $F(1,63) = .17$, $p = .679$, partial $\eta^2 < .005$, $BF_{01} = 6.89$. A significant interaction between error type and stereotypicality, $F(1,63) = 27.32$, $p < .001$, partial $\eta^2 = .30$, $BF_{10} > 10,000$, indicated the difference between within and between category errors was larger for stereotypical than for counterstereotypical statements.

2.2.3 | Exit questions

After the experiment we asked participants to describe their strategy in the memory task. Of all 64 participants, 9 mentioned age but did not specify how they used it, 4 mentioned that they noticed that statements varied in how they fit the speaker's age, 2 said they used age in a counterstereotypical way, and 1 said they used age in a stereotypical way. Thus, the majority of participants did not mention age as a relevant dimension, and only a few mentioned that they used stereotypicality of the statement to select the speaker.

To summarize the results of Experiment 1, we found that recognition accuracy for the statements was not influenced by stereotypicality. The absence of a main effect of stereotypicality suggests that speaker memory was also not affected by stereotypicality, although performance may have been a mix of memory and guessing. We will return to this issue in the General Discussion. Most important, participants made more within category errors when the statement was stereotypical than when it was counterstereotypical. These findings suggest that recall accuracy for the

statements was not biased, but speaker guessing was biased toward the age stereotype.

3 | EXPERIMENT 2

The results of Experiment 1 indicated that participants activated a priori age stereotypes and that their guessing performance was biased toward assigning speakers in a stereotypical way. This effect was most likely due to a priori stereotypes because there was no normative fit in the experimental design. Each speaker made three stereotypical statements and three counterstereotypical statements. It is possible that participants did not notice the absence of normative fit. Indeed, some studies have found that stereotype-based guessing occurs only when participants cannot pay attention to the actual ratio of expected and unexpected items (Bayen & Kuhlmann, 2011; Kuhlmann et al., 2012; but see Ehrenberg & Klauer, 2005; Küppers & Bayen, 2014; Wegener & Klauer, 2004). In the next experiment we tested if participants noticed the actual ratio and used that to guess speakers. If speakers from both categories make mostly counterstereotypical statements guessing might show a bias toward counterstereotypical categories. This would be evident in the types of errors, because if participants would guess toward the counterstereotypical category, they would make more within-category errors for counterstereotypical statements and more between-category errors for stereotypical statements. We thus explored whether category guessing would show an opposite pattern from that of Experiment 1 if all speakers were predominantly counterstereotypical. In Experiment 2 each speaker made five counterstereotypical and one stereotypical statement.

3.1 | Method

3.1.1 | Participants

Sixty-four employees from the same organizations as in Experiment 1 participated voluntarily. None had participated in Experiment 1. The average age was 41.5 ($SD = 12.9$, range = 18–62). Twenty-nine were female.

3.1.2 | Stimuli and procedure

The same materials and procedure were used as in Experiment 1, except that the composition of the sets of statements was adjusted. Eight sets consisted of five statements stereotypical for old and one stereotypical for young workers, and the other eight sets consisted of five statements stereotypical for young and one stereotypical for old workers. Four stereotypical old sets were assigned to the young speakers, and four stereotypical young sets were assigned to the old speakers. As in Experiment 1, set assignment to targets and distractors was counterbalanced across participants, as was assignment of sets to speakers within each age category.
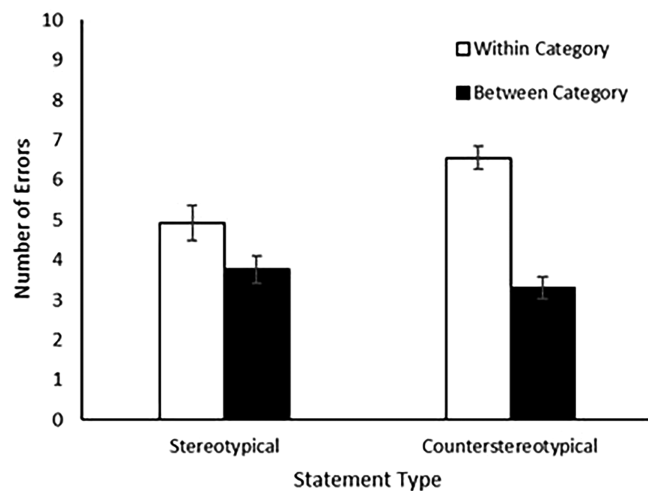
FIGURE 3 Adjusted mean number of errors in Experiment 2. Error bars represent standard error of the mean.

3.2 | Results

3.2.1 | Statement accuracy

We calculated the proportion correct "old" responses to the statements. Accuracy did not differ between statements made by a speaker from the stereotypical ($M = .74$, SEM = .02) or counterstereotypical age group ($M = .77$, SEM = .02), $t(63) = 1.60$, $p = .114$, $BF_{01} = 2.95$. The false alarm rate was .13 (SEM = .01).

3.2.2 | Speaker errors

The number of adjusted incorrect responses per condition is shown in Figure 3. A 2 (error type) by 2 (stereotypicality) ANOVA confirmed that participants made more errors within than between categories, $F(1,63) = 33.23$, $p < .001$, partial $\eta^2 = .35$, $BF_{10} > 10,000$. The number of errors was influenced by stereotypicality, $F(1,63) = 7.47$, $p = .008$, partial $\eta^2 = .11$, although $BF_{01} = 2.08$, and interacted with error type, $F(1,63) = 7.64$, $p = .007$, partial $\eta^2 = .11$, $BF_{10} = 20.68$, indicating that the difference between within and between category errors was larger for counterstereotypical than stereotypical statements. A 2 (experiment) by 2 (error type) by 2 (stereotypicality) ANOVA of the combined Experiment 1 and 2 data showed a significant three-way interaction $F(1,126) = 29.09$, $p < .001$, partial $\eta^2 = .19$, $BF_{10} > 10,000$, indicating that the stereotypicality by error type interactions differed between experiments.

3.2.3 | Exit questions

After the experiment we asked participants to describe their strategy in the memory task. Of all 64 participants, 5 mentioned age but did not specify how they used it, 1 said they used age but related it to

valence of the statement, 4 said they used age in a counterstereotypical way, and 3 said they used age in a stereotypical way.

The results of Experiments 1 and 2 show that participants' recognition accuracy for the statements was not influenced by stereotypicality. When they did not recall the speaker, however, their errors were influenced by stereotypicality. When the speakers were neutral in terms of the stereotypicality of the statements they made, as in Experiment 1, participants seemed to have a bias toward widely held age stereotypes. As shown in Experiment 2, however, when all speakers made mostly counterstereotypical statements, participant seemed to shift their bias to conform to this counterstereotypical state of affairs. Even in this experiment, however, the proportion of stereotypical choices (within category errors for stereotypical statements and between category errors for counterstereotypical statements) was still larger than the proportion of stereotypical statements during the discussion. Thus, although participants were influenced by the actual stereotypicality of the speakers, they still relied on prior stereotypes when trying to recall who said what.

An important question is whether the bias in speaker recall extends to a more general impression that participants have of the speakers. In work related evaluations such as during performance assessments it is relevant that a manager remembers whether a candidate has shown particular attitudes or behavior, for example good work ethics, even if they do not correctly remember a verbatim statement made by the employee. When candidate profiles are evaluated in terms of employability, explicit age cues (such as a photograph of the candidate) influence raters' decisions (Kleissner & Jahn, 2021; Lössbroek et al., 2021). In the next experiment we therefore also measured the participant's impression of the speakers on all topics that were discussed in the statements. We wanted to know to what degree these impressions were influenced by the statement content or by the stereotype of the speaker's age. We also asked participants to rate each speaker on their suitability as a team manager.

4 | EXPERIMENT 3

4.1 | Method

4.1.1 | Participants

Sixty volunteers participated in this study (45 female; $M_{age} = 23.9$, $SD = 7.04$, range = 18–58). Of these, 32 worked at a large professional services firm, 17 as accountants, and 15 as HR workers. The remaining 28 were students at the Erasmus University Rotterdam, The Netherlands, with at least some work experience. All participants were recruited from one of the author's personal network and received no compensation in exchange for their participation.

4.1.2 | Materials and procedure

The materials and procedure were the same as those used in Experiment 1 with the exception that the order of topics was
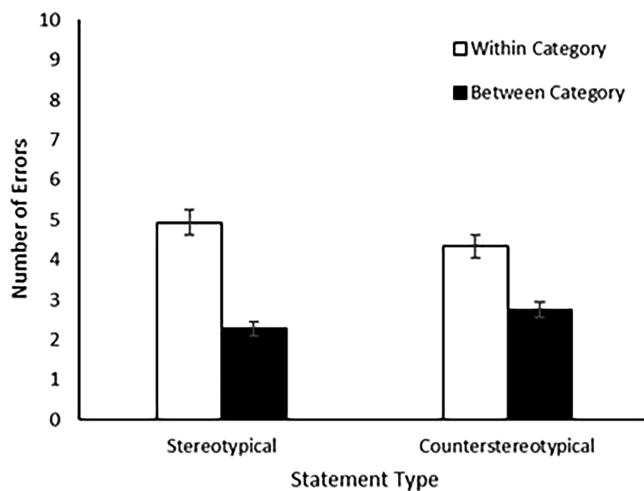
FIGURE 4 Adjusted mean number of errors in Experiment 3. Error bars represent standard error of the mean.

FIGURE 5 Adjusted speaker ratings (higher ratings are more toward young stereotype) in Experiment 3. Error bars represent standard error of the mean.

counterbalanced across participants. In addition, a list of 10 rating scales was created on which each speaker had to be rated. Nine questions corresponded to the subtopics of the statements, and the final question asked how suitable the speaker was for the job of team manager. The questions are listed in the Appendix, Table A2.

The rating phase started with the instruction that all speakers were candidates for a team manager job. Participants were asked to rate each speaker on all 10 questions on a scale of 0 (not applicable) to 10 (very applicable). Participants would rate one speaker on all 10 questions before moving to the next speaker. The photograph was presented at the top of the screen with the question below it. After participants typed a number between 0 and 10 to indicate their rating, the next question was shown below the photograph. Speakers were shown in random order but the questions were always presented in the same order.

After the rating phase, participants were asked their age, gender, and a description of the strategy they had used to perform the task.

4.2 | Results

4.2.1 | Statement accuracy

We calculated the proportion correct "old" responses to the statements. Accuracy did not differ between statements made by a speaker from the stereotypical ($M = .71$, SEM $= .02$) or counterstereotypical age group ($M = .71$, SEM $= .02$), $t(59) = .22$, $p = .823$, $BF_{01} = 9.62$. The false alarm rate was .06 (SEM $= .01$).

4.2.2 | Speaker errors

The numbers of between category errors were multiplied by .75 to adjust for the difference in number of incorrect speakers between categories (4) and within categories (3). The number of adjusted incorrect responses per condition is shown in Figure 4. A 2 (error type) by
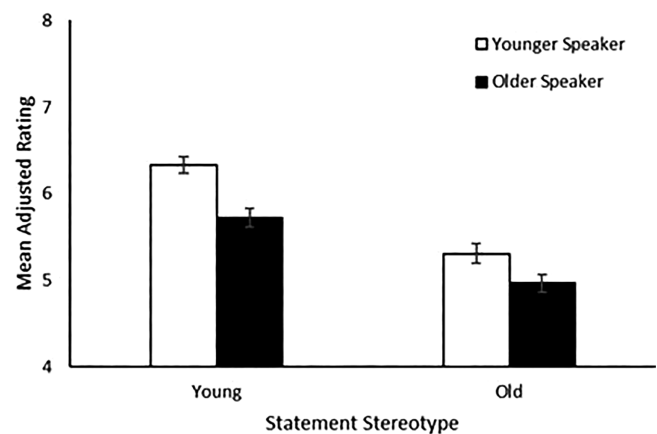
2 (stereotypicality) ANOVA confirmed that participants made more errors within than between categories, $F(1,59) = 55.04$, $p < .001$, partial $\eta^2 = .48$, $BF_{10} > 10,000$. The null-hypothesis that the number of errors was not influenced by stereotypicality could not be rejected and the BF also favored the H0 above H1, $F(1,59) = .11$, $p = .739$, partial $\eta^2 < .005$, $BF_{01} = 7.14$, but a small significant interaction between error type and stereotypicality, $F(1,59) = 4.87$, $p = .031$, partial $\eta^2 = .08$, $BF_{10} = 2.19$, indicated the difference between within and between category errors was larger for stereotypical than for counterstereotypical statements. The interaction effect has the same direction but is smaller than in Experiment 1. This was an unexpected finding. A potential explanation for this difference is that about half of the participants in this experiment were students, whereas the participants in Experiment 1 were all employees. Perhaps students, who probably have less experience in workplace situations, hold weaker age stereotypes than participants who are full time employees. An additional analysis with student/employee status as a between subjects factor, however, revealed no interaction between group, stereotypicality, and error type, $F(1,58) = .34$, $p = .560$. Given that this difference in strength of the interaction was unexpected we refrain from drawing conclusions. Albeit with a smaller effect, the results show the same pattern as in Experiment 1.

4.2.3 | Speaker ratings

Although all speakers were rated on all scales, ratings were excluded if the speaker had not made a statement on that subtopic. Subtopics differed whether the young stereotype would result in a higher rating (e.g., energy, technology) or a lower rating (e.g., experience, loyalty) than the old stereotype. Therefore, ratings on older stereotypes were reversed such that on all scales, higher adjusted ratings always meant stereotypical for young. For each speaker, we calculated ratings separately for topics on which they made a stereotypically young statement and topics for which they made a stereotypically old statement. These adjusted ratings are shown in Figure 5.

An ANOVA showed that adjusted ratings were higher for younger speakers than for older speakers, $F(1,59) = 21.99$, $p < .001$, partial $\eta^2 = .27$, $BF_{10} > 10,000$, and that adjusted ratings were higher for speakers who made a stereotypically young statement on that topic than for speakers who made a stereotypically old statement on that topic, $F(1,59) = 58.30$, $p < .001$, partial $\eta^2 = .50$, $BF_{10} > 10,000$. There was no significant interaction effect between speaker age and statement stereotype, $F(1,59) = 3.23$, $p = .077$, partial $\eta^2 = .05$, $BF_{01} = 2.38$. Participants rated the older speakers as more suitable ($M = 6.59$) for the job of team manager than the younger speakers ($M = 6.24$), $t(59) = 2.51$, $p = .015$, Cohens' $d = .97$, $BF = 2.48$. Thus, although participants did use the speakers' statements in their ratings, they also used the speakers' age.

4.2.4 | Exit questions

After the experiment we asked participants to describe their strategy in the memory task. Of the 60 participants, 5 mentioned age but did not specify how they used it, 2 mentioned that they noticed that statements varied in how they fit the speaker's age, 3 said they noticed age in a counterstereotypical way, and 6 said they noticed age in a stereotypical way. The remaining 44 participants did not mention age as a relevant dimension.

5 | GENERAL DISCUSSION

In three experiments we measured statement and speaker memory performance for age related stereotypical and counterstereotypical statements about work. Our results, larger error rates within than between speaker category, show that participants were more likely to confuse speakers from the same age group than from different age groups. When the speakers were not presented in a stereotypical manner (i.e., made the same number of stereotypical and counterstereotypical statements, Experiments 1 and 3) this difference was larger for stereotypical statements than for counterstereotypical statements. This effect reversed, however, when the speakers were presented in a counterstereotypical manner (i.e., made mostly counterstereotypical statements, Experiment 2). The effects of stereotypicality on error type shows that stereotypes are used when people retrieve information about persons. In Experiments 1 and 3 there was no fit between stereotypicality of the statement and the speaker's age category. That participants made more within category errors for stereotypical than for counterstereotypical statements indicates that they used prior knowledge of stereotypes to guess who said what. This finding is consistent with prior studies that have shown stereotype consistent guessing (Ehrenberg & Klauer, 2005; Graesser et al., 1980; Hicks & Cockman, 2003; Kleider et al., 2008; Kuhlmann et al., 2016; Marsh et al., 2006; Pyszczynski et al., 1987; Schaper et al., 2019; Sherman & Bessenoff, 1999). In our study, the pattern of errors that participants made might have been a mixture of differences in memory for stereotypical and counterstereotypical speaker category and a bias to select

a speaker from the stereotypical category. Prior studies have used models to disentangle the two (Dodson et al., 2008; Klauer & Wegener, 1998; Pazzaglia et al., 2013), but unfortunately our data did not have enough degrees of freedom to fit a single model in which memory sensitivity and bias could both be estimated. That statement recognition performance and the overall number of speaker errors were not affected by stereotypicality suggests, however, that memory for the statements and the speakers did not differ between stereotypical and counterstereotypical items. Moreover, previous studies using similar paradigms have shown consistent effects of expectancy-based guessing. Therefore, it seems likely that our results are at least partly caused by stereotype-consistent guessing. No matter what the exact contribution of memory and bias to the results is, we can observe that speakers were recalled as being more stereotypical than they were originally presented. The practical implication of this finding is that in situations where a group of persons is being evaluated by someone, for example a manager, the manager will likely, even without being aware of it, engage in such reconstructive memory processes. As a consequence, behavior of one person may be erroneously recalled as being the behavior of another person and individuals may be recalled as being more similar to the stereotype than they are.

That the effect of stereotypicality on type of error reversed in Experiment 2 indicates that participants did notice that the speakers were not stereotypical at all, which probably counteracted the effect of prior knowledge, consistent with the probability-matching account (Kuhlmann et al., 2012; Spaniol & Bayen, 2002). Instead, participants based their guessing, at least partly, on the reversed normative fit that they observed for the experimental set of materials. Even in this case, however, participants guessed a stereotypical speaker more often than the actual proportion of stereotypical statements. This indicates that prior stereotypes influence recall performance even when this contradicts the presented information. When all speakers made five counterstereotypical statements and only one stereotypical statement it would be more rational to guess a counterstereotypical speaker every time speaker memory failed. Indeed, some studies have found that expectancy-based guessing was absent when participants could pay attention to the actual ratio of expected and unexpected items during encoding (Bayen & Kuhlmann, 2011; Kuhlmann et al., 2012; but see Ehrenberg & Klauer, 2005; Küppers & Bayen, 2014; Wegener & Klauer, 2004). In these studies the speaker category was indicated by an explicit label, indicating that speaker category was very important. Our manipulation of age was less explicit and the speaker photos provided more information than just age. In addition, unlike other studies, the statements in our study provided information about the person rather than explicit statements about age. Together, these differences may have caused our participants to pay less explicit attention to age stereotypicality, reducing the extent to which they noticed the ratio of stereotypical and counterstereotypical items. As a result, prior stereotypes may have had a larger effect.

The speaker rating results in Experiment 3 also showed that stereotypes influenced the impression that participants had of the speakers. The ratings were made on topics on which the speakers had made explicit statements about themselves. Across participants the

same statements were made by older and younger speakers and all speakers made the same number of stereotypical and counterstereo-typical statements. Nevertheless, younger speakers were rated more toward the young stereotype than older speakers, indicating that participants rated the speakers at least in part based on their age category instead of on what they said. Overall, older speakers were rated as (slightly) better suited for a job as "team manager" than younger speakers. Possibly, the team manager job reflects an age-typed job that is broadly believed to more typically held by older workers (cf. Perry & Finkelstein, 1999). Other studies, however, suggest older workers are commonly considered less qualified for management positions (e.g., Cheung & Woo, 2021) or found that such positions are rated as equally accessible for younger and older workers (Derous & Decoster, 2017). The stereotypicality bias in speaker recall and speaker ratings are consistent with biases in memory performance for object pictures (Hemmer & Steyvers, 2009; Huttenlocher et al., 2000; Poirier et al., 2017). These studies also show that recall performance for a stimulus dimension, such as picture size, is biased toward the category average, where the average can be based on the items that are shown during the experiment and on the prior knowledge of the participants.

In contrast, recognition accuracy for the statements was not affected at all by stereotypicality. This result stands in contrast to the results of many studies that obtained better memory for inconsistent (i.e., counterstereotypical) information (Bardach & Park, 1996; Bell et al., 2015; Ehrenberg & Klauer, 2005; Graesser et al., 1980; Kleider et al., 2008; Pezdek et al., 1989; Wegener & Klauer, 2004). Schaper et al. (2019) have proposed that the memory advantage for inconsistent information may depend on the degree of inconsistency. Unexpected information, such as seeing an oven in a bathroom (Schaper et al., 2019) or a person who was described as a friendly, helpful social worker showing aggressive behavior (Ehrenberg & Klauer, 2005) might be more salient and as a result will be encoded better than expected information. In a typical workplace situation, however, behaviors and opinions might be more typical for older or younger employees, but not constitute strong violations of expectancies. For example, an older employee saying "When there is an opportunity to learn new things I grab it" would not be typical for an older person, but might not be completely unexpected either. Thus, the memory benefit of expectancy violations might be reduced or even absent in this case.

In the present study we specifically addressed stereotyping in the workplace. Statement content was work-related. The who-said-what paradigm resembles a variety of work situations in which different people are observed by others such as their managers. In actual work settings, in many cases, these managers will subsequently evaluate the employee at least partly based on memory of these observations and will likely make memory errors in this process, such as the reconstructive memory processes that lead to the present results. These errors will bias the evaluation process. Our results suggest that managers will tend to confuse an employee's behavior with the behavior of co-workers from the same age group. Moreover, even in the absence of more stereotypical than counterstereotypical behavior,

employees will be recalled as more stereotypical than they really are. Only when the entire age group shows mostly counterstereotypical behavior will recall to some degree be biased in a counterstereotypical direction.

In a realistic work situation, grouping will be based on other characteristics as well, such as gender and ethnicity. To what extent age grouping will result in recall errors might also depend on the perceived relevance of age (Klauer et al., 2003; Van Twuyver, 1996). In any case, if the present results can be extended to the workplace, this suggests the undesirable scenario that people might attribute properties (ideas, skills, behaviors) to the wrong person, and moreover, that this might happen based on the social group that the person is perceived to be in. This is especially undesirable in light of research evidence showing there is little truth in the large majority of workplace age stereotypes (Ng & Feldman, 2012).

Our manipulation of stereotypicality may be regarded as more subtle and perhaps more realistic than that in some previous studies that manipulated normative fit (Blanz & Aufderheide, 1999; Van Twuyver, 1996; Wegener & Klauer, 2004). In previous studies, statements were opinions about topics that were relevant for the social groups, such as gender roles in a discussion among men and women. We presented statements that people made about their work. These statements may resemble more naturalistic settings such as employees talking to each other or to their manager. That we obtained similar results as these previous studies indicates that stereotypes influence how people recall conversations even when these stereotypes have not been primed by the topic of the conversation. This suggests that stereotype bias in recall may occur in all kinds of everyday situations.

Our results also indicate that stereotype bias is due to guessing in cases where memory fails. When asked about how they performed the memory task, the large majority of our participants did not mention the ages of the speakers, and only a few of the participants mentioned age-related stereotypes. This suggests that most participants were not aware that the information that they recalled was more stereotypical than the information that they studied. Thus, they were unaware that their performance was the result of biased reconstructive memory. That statement recognition accuracy was not affected by stereotypes may be good news for measures aimed at reducing stereotype bias. If memory can be improved or supported by external records, bias may be reduced.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data files that were used for the analyses can be found on the Open Science Framework page https://osf.io/32gt6/.

ORCID

Diane Pecher https://orcid.org/0000-0003-0976-6080
Heleen van Mierlo https://orcid.org/0000-0003-3086-0258
Samantha Bouwmeester https://orcid.org/0000-0003-1771-0546

ENDNOTES

[1] A home improvement supply store, a town council, a police station, a shipping company, an insurance company, a professional education college, a nursing service, and workers in food technology.

[2] Bayes Factor (BF), which is the ratio of $p(D \mid H0)$ and $p(D \mid H1)$, reflects the probabilities of observing the data under the null hypothesis and the alternative hypothesis, respectively. The Bayes Factor thus provides a relative measure of the extent to which the data provide evidence for the null hypothesis of no effect or the alternative hypothesis (Rouder et al., 2009). Throughout this paper we report BF_{01} if the evidence is in favor of H_0 and BF_{10} if the evidence is in favor of H_1. Bayes factors were calculated using JASP software (JASP Team, 2018). The prior had an r scale fixed effects of .5 and an r scale random effects of 1.0.

REFERENCES

Bardach, L., & Park, B. (1996). The effect of in-group out-group status on memory for consistent and inconsistent behavior of an individual. *Personality and Social Psychology Bulletin*, 22(2), 169–178. https://doi.org/10.1177/0146167296222006

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source-item contingency and schematic knowledge on source monitoring tests of the probability-matching account. *Journal of Memory and Language*, 64(1), 1–17. https://doi.org/10.1016/j.jml.2010.09.001

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology-Learning Memory and Cognition*, 22(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Bell, R., Mieth, L., & Buchner, A. (2015). Appearance-based first impressions and person memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 41(2), 456–472. https://doi.org/10.1037/xlm0000034

Blanz, M. (1999). Accessibility and fit as determinants of the salience of social categorizations. *European Journal of Social Psychology*, 29(1), 43–74. https://doi.org/10.1002/(SICI)1099-0992(199902)29:1<43::AID-EJSP913>3.0.CO;2-Z

Blanz, M., & Aufderheide, B. (1999). Social categorization and category attribution: The effects of comparative and normative fit on memory and social judgment. *British Journal of Social Psychology*, 38, 157–179. https://doi.org/10.1348/014466699164103

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, 13(2), 207–230. https://doi.org/10.1016/0010-0285(81)90008-6

Cantor, N., & Mischel, W. (1979). Prototypicality and personality: Effects on free recall and personality impressions. *Journal of Research in Personality*, 13(2), 187–205. https://doi.org/10.1016/0092-6566(79)90030-8

Cheung, S. Y., & Woo, L. (2021). Age stereotypes and the job suitability of older workers from hotel managers' perspectives. *International Journal of Hospitality Management*, 95, 102932. https://doi.org/10.1016/j.ijhm.2021.102932

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology*, 47(3), 292–303. https://doi.org/10.1016/S0022-2496(03)00005-1

Derous, E., & Decoster, J. (2017). Implicit age cues in resumes: Subtle effects on hiring discrimination. *Frontiers in Psychology*, 8, 1321. https://doi.org/10.3389/fpsyg.2017.01321

Dodson, C. S., Darragh, J., & Williams, A. (2008). Stereotypes and retrieval-provoked illusory source recollections. *Journal of Experimental Psychology-Learning Memory and Cognition*, 34(3), 460–477. https://doi.org/10.1037/0278-7393.34.3.460

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology*, 41(4), 369–387. https://doi.org/10.1016/j.jesp.2004.08.001

Finkelstein, L. M., Ryan, K. M., & King, E. B. (2013). What do the young (old) people think of me? Content and accuracy of age-based metastereotypes. *European Journal of Work and Organizational Psychology*, 22(6), 633–657. https://doi.org/10.1080/1359432X.2012.673279

Graesser, A. C., Woll, S. B., Kowalski, D. J., & Smith, D. A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning and Memory*, 6(5), 503–515. https://doi.org/10.1037/0278-7393.6.5.503

Hebl, M., Cheng, S. K., & Ng, L. C. (2020). Modern discrimination in organizations. In F. Morgeson (Ed.), *Annual review of organizational psychology and organizational behavior* (Vol. 7, pp. 257–282). Annual Reviews. https://doi.org/10.1146/annurev-orgpsych-012119-044948

Hemmer, P., & Steyvers, M. (2009). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin & Review*, 16(1), 80–87. https://doi.org/10.3758/PBR.16.1.80

Hicks, J. L., & Cockman, D. W. (2003). The effect of general knowledge on source memory and decision processes. *Journal of Memory and Language*, 48(3), 489–501. https://doi.org/10.1016/S0749-596X(02)00537-5

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology-General*, 129(2), 220–241. https://doi.org/10.1037//0096-3445.129.2.220

JASP Team, T. (2018). JASP (version 0.9) [computer software].

Karasawa, M., Maass, A., Rakic, T., & Kato, A. (2014). The emergent nature of culturally meaningful categorization and language use: A Japanese-Italian comparison of age categories. *Journal of Cross-Cultural Psychology*, 45(3), 431–451. https://doi.org/10.1177/0022022113509882

Kite, M. E., Stockdale, G. D., Whitley, B. E., Jr., & Johnson, B. T. (2005). Attitudes toward younger and older adults: An updated meta-analytic review. *Journal of Social Issues*, 61(2), 241–266. https://doi.org/10.1111/j.1540-4560.2005.00404.x

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, 75(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Klauer, K. C., & Ehrenberg, K. (2005). Social categorization and fit detection under cognitive load: Efficient or effortful? *European Journal of Social Psychology*, 35(4), 493–516. https://doi.org/10.1002/ejsp.266

Klauer, K. C., Ehrenberg, K., & Wegener, I. (2003). Crossed categorization and stereotyping: Structural analyses, effect patterns, and dissociative effects of context relevance. *Journal of Experimental Social Psychology*, 39(4), 332–354. https://doi.org/10.1016/S0022-1031(03)00017-9

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychology*, 75(5), 1155–1178.

Kleider, H. M., Pezdek, K., Goldinger, S. D., & Kirk, A. (2008). Schema-driven source misattribution errors: Remembering the expected from a witnessed event. *Applied Cognitive Psychology*, 22(1), 1–20. https://doi.org/10.1002/acp.1361

Kleissner, V., & Jahn, G. (2021). Implicit and explicit age cues influence the evaluation of job applications. *Journal of Applied Social Psychology*, *51*(2), 107–120. https://doi.org/10.1111/jasp.12720

Kuhlmann, B. G., Bayen, U. J., Meuser, K., & Kornadt, A. E. (2016). The impact of age stereotypes on source monitoring in younger and older adults. *Psychology and Aging*, *31*(8), 875–889. https://doi.org/10.1037/pag0000140

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology-Learning Memory and Cognition*, *38*(5), 1365–1376. https://doi.org/10.1037/a0028147

Küppers, V., & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *Quarterly Journal of Experimental Psychology*, *67*(10), 2042–2059. https://doi.org/10.1080/17470218.2014.904914

Lössbroek, J., Lancee, B., van der Lippe, T., & Schippers, J. (2021). Age discrimination in hiring decisions: A factorial survey among managers in nine European countries. *European Sociological Review*, *37*(1), 49–66. https://doi.org/10.1093/esr/jcaa030

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). Gender and orientation stereotypes bias source-monitoring attributions. *Memory*, *14*(2), 148–160. https://doi.org/10.1080/09658210544000015

Murphy, K. R., & DeNisi, A. S. (2022). Do age stereotypes predict personnel decisions? The state of the evidence. *Work, Aging and Retirement*, *8*(4), 323–330. https://doi.org/10.1093/workar/waab019

Ng, T. W. H., & Feldman, D. C. (2012). Evaluating six common stereotypes about older workers with meta-analytical data. *Personnel Psychology*, *65*(4), 821–858. https://doi.org/10.1111/peps.12003

Nolan, M. A., Haslam, S. A., Spears, R., & Oakes, P. J. (1999). An examination of resource-based and fit-based theories of stereotyping under cognitive load and fit. *European Journal of Social Psychology*, *29*(5–6), 641–663.

Pazzaglia, A. M., Dube, C., & Rotello, C. M. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin*, *139*(6), 1173–1203. https://doi.org/10.1037/a0033044

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. https://doi.org/10.1016/S1053-4822(99)00010-8

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 587–595. https://doi.org/10.1037/0278-7393.15.4.587

Pietraszewski, D., Curry, O. S., Petersen, M. B., Cosmides, L., & Tooby, J. (2015). Constituents of political cognition: Race, party politics, and the alliance detection system. *Cognition*, *140*, 24–39. https://doi.org/10.1016/j.cognition.2015.03.007

Poirier, M., Heussen, D., Aldrovandi, S., Daniel, L., Tasnim, S., & Hampton, J. A. (2017). Reconstructing the recent visual past: Hierarchical knowledge-based effects in visual working memory. *Psychonomic Bulletin & Review*, *24*(6), 1889–1899. https://doi.org/10.3758/s13423-017-1277-9

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. https://doi.org/10.1177/0149206308318617

Pyszczynski, T., Laprelle, J., & Greenberg, J. (1987). Encoding and retrieval effects of general person characterizations on memory for incongruent and congruent information. *Personality and Social Psychology Bulletin*, *13*(4), 556–567. https://doi.org/10.1177/0146167287134012

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rupp, D. E., Vodanovich, S. J., & Credé, M. (2006). Age bias in the workplace: The impact of ageism and causal attributions. *Journal of Applied Social Psychology*, *36*(6), 1337–1364. https://doi.org/10.1111/j.0021-9029.2006.00062.x

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019). Metamemory expectancy illusion and schema-consistent guessing in source monitoring. *Journal of Experimental Psychology-Learning Memory and Cognition*, *45*(3), 470–496. https://doi.org/10.1037/xlm0000602

Schütz, J., & Bröder, A. (2011). Signal detection and threshold models of source memory. *Experimental Psychology*, *58*(4), 293–311. https://doi.org/10.1027/1618-3169/a000097

Sherman, J. W., & Bessenoff, G. (1999). Stereotypes as source-monitoring cues: On the interaction between episodic and semantic memory. *Psychological Science*, *10*(2), 106–110. https://doi.org/10.1111/1467-9280.00116

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology-Learning Memory and Cognition*, *28*(4), 631–651. https://doi.org/10.1037/0278-7393.28.4.631

Stangor, C., Lynch, L., Duan, C. M., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, *62*(2), 207–218. https://doi.org/10.1037/0022-3514.62.2.207

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*(7), 778–793. https://doi.org/10.1037/0022-3514.36.7.778

Van Twuyver, M. (1996). *Factors affecting social categorization processes in memory*. [Doctoral Dissertation]. University of Nijmegen https://hdl.handle.net/2066/146208

Wegener, I., & Klauer, K. C. (2004). Inter-category versus intra-category fit: When social categories match social context. *European Journal of Social Psychology*, *34*(5), 567–593. https://doi.org/10.1002/ejsp.217

Zaniboni, S., Kmicinska, M., Truxillo, D. M., Kahn, K., Paladino, M. P., & Fraccaroli, F. (2019). Will you still hire me when I am over 50? The effects of implicit and explicit age stereotyping on resume evaluations. *European Journal of Work and Organizational Psychology*, *28*(4), 453–467. https://doi.org/10.1080/1359432X.2019.1600506

APPENDIX A

TABLE A1   Statements used in the experiments listed by stereotype and theme.

Statement	Approximate translation
Stereotypical young	
Learning and adjustment	
Ik heb niet zo veel ervaring met dit werk, er is nog veel wat ik moet leren.	I have little experience in this job, I still need to learn a lot.
Ik was vrijwel meteen gewend aan het nieuwe logistieke systeem.	I got used to the new logistic system practically immediately.
Is er een nieuw beleid, oké geen probleem, ik pas mij wel aan.	Do we have a new policy, okay no problem, I will adjust.
Als er een kans is om nieuwe dingen te leren pak ik die.	When there is an opportunity to learn new things I grab it.
De nieuwe baas heeft een andere visie op de werkwijze binnen het bedrijf, gelukkig houd ik wel van verandering.	The new boss has a different view on the procedures in this company, fortunately I like change.
Ik word altijd heel enthousiast als ik iets heel nieuws te doen krijg.	I always get very enthusiastic when I get something new to do.
Als ik nieuwe taken krijg leer ik deze altijd erg snel.	When I get new tasks I am always quick to learn them.
Ik werk in een dynamisch bedrijf en het bestuur blijft vernieuwen. Een leuke uitdagende baan dus!	I work for a dynamic company and the board keeps innovating. A fun and challenging job!
Technology	
Ik sla al mijn bestanden op in de Cloud, dan kan ik er overal bij.	I store all my files in the cloud so that I can access them from anywhere.
E-mailen, appen en sms'en is geen probleem voor mij, dat kan ik nog met mijn ogen dicht.	E-mail, app or texting is no problem for me, I can do it with my eyes closed.
Ik snap niet hoe mensen zonder mobiele telefoon kunnen leven.	I do not understand how people can live without a mobile phone.
Morgen krijgen we een nieuw robotarm binnen. Mooi spul hoor, ik heb al helemaal uitgezocht hoe hij werkt.	Tomorrow a new robot arm will arrive. Really great stuff, I've already sorted out how it works.
Iedereen zou met computers moeten werken, het is snel en gemakkelijk.	Everyone should work on computers, it is fast and easy.
Als ik iets heb afgesproken zet ik het vaak in mijn telefoon.	When I have agreed to something I often note in in my phone.
Ik ben vrij handig met de nieuwe technologische apparatuur op het werk.	I am pretty skilful with new technical equipment at work.
Ik vind het veiliger om mijn gegevens op een computer te zetten dan op een los papiertje.	I consider storing my information on a computer safer than putting it on a piece of paper.
Health	
Ik heb een goedkope zorgverzekering genomen, ik heb toch vrijwel nooit ergens last van.	I took a cheap health care insurance, since I never have any trouble.
Ik ben erg fit en kan goed lange dagen maken.	I'm in good shape and can easily make long days.
Mijn gezondheid is prima, het belemmert mij nooit in wat ik wil doen.	My health is excellent, it never stops me in what I want to do.
Ik heb geen medicijnen nodig, ik heb zelfs geen paracetamol in huis.	I do not need medication, I do not even have paracetamol at home.
Ik ben eigenlijk nooit ziek, hooguit wat hoofdpijn na een gezellig weekend.	I really am never sick, at the most some headache after a fun weekend.
Ik heb een hoge weerstand en als ik toch ziek word duurt het nooit lang.	I have good immunity and when I still get sick it never lasts long.
Ik ben dit jaar nog niet ziek geweest.	I have not been sick yet this year.
Als mensen om mij heen ziek zijn heb ik meestal nergens last van.	When people around me are sick I usually have no trouble at all.
Ambition	
Ik word steeds ambitieuzer in mijn werk, ik wil graag succesvol zijn.	I am becoming more and more ambitious in my work, I like to be successful.
In de komende jaren wil ik mijzelf opwerken tot directeur van dit bedrijf.	In the coming years I want to work my way up to manager of this company.
Ik werk graag in een omgeving waar innovatie belangrijk wordt gevonden.	I like to work in an environment where innovation is considered important.
Ik besteed meer tijd aan mijn carrière dan aan mijn familie.	I spend more time on my career than on my family.
Als ik hard werk, kan ik vast hogerop in dit bedrijf komen.	If I work hard I can surely go higher up in this company
Ik hou er van om uitdagingen aan te gaan.	I like to accept challenges.

(Continues)

TABLE A1 (Continued)

Statement	Approximate translation
Als er een vacature voor een hogere positie binnen het bedrijf vrijkomt, solliciteer ik meteen.	If a vacancy becomes available for a higher position in the company I will apply immediately.
Ik werk hard om beter te worden in mijn werk.	I work hard to get better at my job.
Work ethics and experience	
Als een ander bedrijf mij een baan aanbiedt zou ik dat zeker overwegen.	If another company offered me a job I would certainly consider it.
Ik ben wel eens te laat op mijn werk, maar wat maakt het uit als ik 10 minuutjes later ben.	Sometimes I am late for work, but what does it matter that I am 10 minutes late.
Als het even kan meld ik mij gewoon ziek.	Whenever possible I report sick.
Ik moet nog veel leren over dit bedrijf en het beleid.	I still need to learn a lot about this company and its policy.
Als collega's mij een vraag stellen over werk moet ik ze altijd doorsturen naar iemand met meer ervaring.	When colleagues ask me questions about work I always need to send them to someone with more experience.
Ik heb minder vakkennis, dus soms weet ik niet waar mijn collega's het over hebben.	I have less professional knowledge, so sometimes I have no idea what colleagues are talking about.
Ik weet niet zo goed wie ik moet benaderen als ik iets nodig heb.	I am not sure whom to approach when I need something.
Ik kom elke dag weer dingen tegen die ik nog niet eerder heb gedaan.	Every day I encounter things I have never done before.
Energy and impulsivity	
Aan het eind van een werkdag ga ik graag nog iets leuks doen.	At the end of the work day I like to do something fun.
Ik ga ieder weekend uit, ik vind het fijn om actief te zijn.	I go out every weekend, I like being active.
Mijn bijnaam is stuiterbal, omdat ik nooit stilzit.	My nickname is bouncing ball, because I never sit still.
Collega's vragen vaak hoe ik altijd zo energiek kan zijn.	Colleagues often ask me how I can always be so energetic.
Hoe sneller hoe beter, dat is wat ik vind.	The quicker the better, that is what I think.
Als ik iets wil hebben koop ik het meteen.	When I want something I buy it immediately.
Ik neem veel beslissingen in een opwelling.	I make many decision on a whim.
Als ik iets denk flap ik het er vaak gelijk uit.	If I have a thought I often just blurt it out right away.
Stereotypical old	
Learning and adjustment	
Ik ben het productiefst als ik mijn vaste routine kan volgen.	I am most productive when I can follow my set routines.
Ik heb vorig jaar een training gevolgd, maar ik blijf liever op de oude manier werken.	I took a training last year, but I prefer to work in the old way.
Ik doe mijn werk goed, ik denk niet dat er nog iets is wat ik moet leren.	I do my job well, I do not think I need to learn anything anymore.
Ik denk niet dat ik ooit zal wennen aan het nieuwe beleid.	I do not think I will ever get used to the new policy.
Ik raak echt gefrustreerd van al die hypes waar we zo nodig "iets mee moeten" van de directie.	I really get frustrated by all those hypes that the management wants us to "do something with".
Ik merk aan mijzelf dat ik het lastig vind nieuwe dingen bij te leren.	I notice that I find it hard to learn new things.
Ik ben al genoeg opgeleid, ik heb geen behoefte aan het leren van nieuwe dingen.	I have been educated enough, I do not feel the need to learn new things.
Ondanks de training lukt het me niet om het nieuwe administratiesysteem onder de knie te krijgen.	Despite the training I cannot manage to master the new administrative system.
Technology	
Mijn collega moet me vaak helpen om bestanden terug te vinden.	My colleague often has to help me to recover files.
Ik houd al mijn afspraken netjes bij in een papieren agenda.	I keep all my appointments neatly in a paper diary.
We kregen laatst een nieuwe printer. Wat een onmogelijk ding, ik kan er niet mee werken.	The other day we got a new printer. What a disaster, I cannot work with it.
Ik moet toegeven dat werken met computers niet mijn sterkste punt is.	I have to admit that working with computers is not my strong point.
Als ik een e-mail moet versturen duurt het soms erg lang, omdat ik vaak niet weet waar ik op moet drukken.	When I have to send an email it sometimes takes long because I often do not know where to click.
Ik vind het lastig om met computers te werken, het is altijd zo ingewikkeld.	I find it tricky to work with computers, it often is very complicated.

TABLE A1 (Continued)

Statement	Approximate translation
Ik vind het onnodig dat iedereen altijd een mobiele telefoon bij zich heeft.	I do not think everyone always needs to have their mobile phone with them.
Ik kon alleen een afspraak maken via de website, wat een gedoe, geef mij maar gewoon een telefoonnummer om te bellen.	I could only book an appointment on the website, what a hassle, I prefer a simple phone number that I can call.

Health

Iemand met een snotneus hoeft maar mijn kant op te kijken of ik ben al verkouden.	Someone with a runny nose only has to look my way and I get a cold.
Ik heb altijd een doosje paracetamol in mijn tas zitten voor het geval ik mij niet lekker voel.	I always carry a box of paracetamol in my bag just in case I am not feeling too well.
Ik maak jaarlijks veel ziektedagen.	I use a lot of sick days every year.
Als er een virus heerst ben ik vaak één van de eersten die ziek wordt.	When a virus goes around I'm usually one of the first to get sick.
Echt vervelend, ik ben deze maand al 3 keer ziek geweest.	Really bothersome, I've been sick three times already this month.
Ik let op wat ik eet, ik heb namelijk snel last van mijn maag.	I watch what I eat, because my stomach gets upset easily.
Ik ga vrijwel altijd over mijn eigen risico voor ziektekosten heen.	I usually exceed the deductible for health care costs.
Door mijn chronische rugpijn moet ik vaak eerder naar huis.	Because of my chronic back pain I often have to go home early.

Ambition

Ik ben al een professional op dit gebied, dus daar hoef ik niet veel meer aan te doen.	I already am a professional in this field, so I do not have to do much about that anymore.
Er was een vacature voor een hogere positie binnen het bedrijf, maar ik heb niet gesolliciteerd, ik ben tevreden met mijn huidige functie.	A vacancy for a higher position in this company came up but I did not apply, I am happy at my current position.
Ik ben niet erg innovatief ingesteld.	I do not have an innovative mentality.
Ik vind alles wel goed hoe het is.	I think everything is all right the way it is.
Ik vind promotie maken niet zo belangrijk.	I do not care much about getting promotion.
Ik vind dat ik mijzelf genoeg heb ontwikkeld.	I think I have educated myself enough.
Carrière maken? Ik besteed liever meer tijd met mijn familie.	Make a career for myself? I rather spend more time with my family.
Waar ik mezelf zie over vijf jaar? Waarschijnlijk zit ik dan gewoon nog op dezelfde plek.	Where do I see myself in 5 years? I will probably still be in the same spot.

Work ethics and experience

Als er iets mis gaat op de afdeling neem ik vaak de verantwoordelijkheid op me.	When something goes wrong at the department I often take on responsibility.
Ik ben altijd op tijd op mijn werk.	I always arrive on time at work.
Als ze mij een baan aanbieden bij een ander bedrijf zou ik die afwijzen.	If they were to offer me a job at another company I would reject it.
"Werk hard" is mijn motto.	"Work hard" is my motto.
Ik ken dit bedrijf en alle procedures op mijn duimpje.	I know this company and its procedures like the back of my hand.
Ik ken het beleid en weet wie waar verantwoordelijk voor is binnen het bedrijf.	I know the policy and know who is responsible for what in this company.
Mijn collega's vragen mij vaak om mijn kennis met hen te delen.	My colleagues often ask me to share my knowledge with them.
Ik heb veel ervaring met het werk dat ik doe.	I have a lot of experience with the work that I do.

Energy and impulsivity

Als ik 's avonds moet overwerken, val ik bijna in slaap.	When I need to work overtime at night I almost fall asleep.
Na een dag werken ben ik volledig uitgeput.	After a working day I feel completely exhausted.
Ik werk zo langzaam dat mijn collega's mij de turtle (schildpad) noemen.	I work so slowly that my colleagues call me the turtle.
Ik heb het weekend echt nodig om uit te rusten van de werkweek.	I really need the weekend to rest from the work week.
Ik vergeet vaak waar ik mee bezig was.	I often forget what I was doing.
Als ik iets wil kopen kijk ik altijd eerst naar de alternatieven.	When I want to buy something I always look at the alternatives first.
Ik overdenk altijd al mijn acties.	I always reconsider my actions.
Geduld is een schone zaak, dat is in ieder geval mijn mening.	Patience is a virtue, at least, that is my opinion.

TABLE A2 Rating scales used in Experiment 3.

Question	Approximate translation
Deze kandidaat is ambitieus.	This candidate is ambitious.
Deze kandidaat is bereid nieuwe dingen te leren en zich aan te passen.	This candidate is willing to learn new things and adjust themselves.
Deze kandidaat kan goed omgaan met nieuwe technologie.	This candidate is well able to handle new technology.
Deze kandidaat heeft een goede gezondheid.	This candidate has a good health.
Deze kandidaat heeft veel ervaring.	This candidate has lots of experience.
Deze kandidaat is energiek.	This candidate is energetic.
Deze kandidaat is impulsief.	This candidate is impulsive.
Deze kandidaat heeft een positieve werk ethiek.	This candidate has positive work ethics.
Deze kandidaat is loyaal aan haar werkgever.	This candidate is loyal to her employer.
Deze kandidaat is geschikt.	This candidate is suitable.