American Psychological Association

Journal of Personality and Social Psychology: Attitudes and Social Cognition

0022-3514/19/$12.00

2019, Vol. 116, No. 6, 899–918
http://dx.doi.org/10.1037/pspa0000162

Learning What to Inhibit: The Influence of Repeated Testing on the Encoding of Gender and Age Information

Tomás A. Palma, Leonel Garcia-Marques, Pedro Marques, and Sara Hagá
University of Lisbon

B. Keith Payne
University of North Carolina at Chapel Hill

Psychological research has devoted considerable attention to the relationship between the multiple category dimensions that can be extracted from faces. In the present studies, we investigated the role of experience and learning on the way the social perceiver deals with multiple category dimensions. Specifically, we tested whether learning which of 2 dimensions is the most relevant to the task at hand influences the encoding and retrieval of both task-relevant and irrelevant dimensions. In our studies, participants went through several cycles, each consisting of a study and a test phase. We manipulated the structure of the tests such that participants were probed on only 1 category dimension (age or gender), despite viewing faces of both category dimensions in all study phases. We hypothesized that when participants were repeatedly tested on 1 dimension, they would proactively control their attention toward that specific dimension and away from the nonrelevant dimension. Five studies demonstrated that: (a) participants learned which dimension was test-relevant such that they gradually became faster and more accurate on that dimension; (b) when the gender dimension was test-relevant, participants were faster and more accurate retrieving information concerning the target faces' gender than age, while the opposite did not happen when the age dimension was test-relevant; and (c) this dominance of the gender dimension is mainly caused by the inhibition of the age dimension. Implications about the importance of previous experience and control for research on social categorization in general and research examining the interplay between gender and age are discussed.

Keywords: social categorization, face perception, stereotypes, learning, task relevance

Supplemental materials: http://dx.doi.org/10.1037/pspa0000162.supp

Imagine the following scenario. You work in a consulting company and your director has chosen you to run a focus group about a supermarket chain. She did not give you any particular instructions about what to look for in the focus group. During the focus group discussions, you took notes about each candidate's views on different topics. After this first focus group discussion, you meet with your director. In this meeting, she asked you several questions about the views expressed by the senior versus the junior members of the group and particularly about how they differed. Will this meeting with the director, namely the nature of the questions she asked you, influence the way you will run the next focus group discussion about the supermarket chain?

This article was published Online First April 18, 2019.

Tomás A. Palma, Leonel Garcia-Marques, Pedro Marques, and Sara Hagá, CICPSI, Faculty of Psychology, University of Lisbon; B. Keith Payne, Department of Psychology & Neuroscience, University of North Carolina at Chapel Hill.

This research was partially supported by grants (SFRH/BPD/92271/2013; PTDC/MHC-PCN/1267/2014) from the Fundação para a Ciência e Tecnologia (FCT). We thank Joana Reis for her valuable help in the construction of the neutral statements used in the first three studies.

Correspondence concerning this article should be addressed to Tomás A. Palma, Faculty of Psychology, University of Lisbon, 1649–013 Lisbon, Portugal. E-mail: tapalma@psicologia.ulisboa.pt

As illustrated in this hypothetical scenario, the purpose of the present research is to examine whether learning the task requirements determines the encoding and retrieval of new social information. In other words, do the requirements of a previous retrieval task help people tune their encoding strategies to the most relevant stimulus dimensions when confronted with similar yet new stimuli? We tested this idea in the context of social categorization from faces. Namely, we examined the encoding and retrieval of two basic social categories for person perception—gender and age (Brewer, 1988; Fiske & Neuberg, 1990). Going back to our scenario, you could learn from the questions asked by the director that the age of the focus group members is a critical dimension and, in the next group discussion, shift your attention (either tacitly or more deliberately) to the age and away from the gender of the group members.

In the next paragraphs, we elaborate on these ideas. Namely, we first briefly review the social cognition literature on the extraction of social categories from facial cues. Then, we describe the studies that have examined the interplay between multiple category dimensions. Finally, we detail the goal and hypothesis of our research.

The Prevalence of Social Categorization

People spontaneously classify other people according to their social category memberships (Allport, 1954; Tajfel & Turner, 1979, 1986). Upon seeing an unfamiliar person, we may, for example, categorize the person as a senior, infer that the person may need help to go up the stairs and consequentially adjust our behavior to help the person. This elegant cognitive strategy dramatically facilitates the perceiver's job as it reduces the demands of dealing with several different people every day. Social categorization is, thus, an essential element of social perception (Allport, 1954; Brewer, 1988; Fiske & Neuberg, 1990; Freeman & Ambady, 2011; Macrae & Bodenhausen, 2000).

One important conclusion that can be drawn from early research is that the encoding of social category memberships typically occurs spontaneously and unintentionally. For example, several studies using the classic "Who said what?" paradigm (Taylor, Fiske, Etcoff, & Ruderman, 1978) have shown that the mere observation of a discussion between members of different social categories is enough to trigger category encoding. This process is reflected by the pattern of recall errors systematically committed by participants when later asked to match each discussion statement with its original speaker. That is, they tend to show a higher likelihood of misattributing a statement to a member of the speaker's category than to a member of a different category (e.g., Blanz, 1999; Klauer, Ehrenberg, & Wegener, 2003; Klauer & Wegener, 1998; Pietraszewski, Cosmides, & Tooby, 2014; Stangor, Lynch, Duan, & Glas, 1992; Taylor et al., 1978; van Twuyver & van Knippenberg, 1995).

From all the social category dimensions described in the literature, race, gender, and age seem to possess a unique cognitive status. According to classic social perception models, these dimensions have such a high level of chronic accessibility—presumably because of its frequent activation in daily life—that one cannot avoid noticing and using them (e.g., Brewer, 1988; Fiske & Neuberg, 1990). Another reason why these particular category dimensions are so salient to the perceiver is their distinct physical markers. They are easily distinguished based on a host of facial features (e.g., skin texture, skin color, hairstyle, and hair color/quantity; Berry & McArthur, 1986; Blair, Judd, & Chapleau, 2004; Brebner, Martin D., & Macrae, 2009; Burt & Perrett, 1995; Johnson & Tassinary, 2005; MacLin & Malpass, 2001; Macrae & Martin, 2007). Actually, categorization accuracy for race, age, and gender typically approaches ceiling levels (George & Hole, 1995; D. Martin & Macrae, 2007; Remedios, Chasteen, Rule, & Plaks, 2011).

Numerous ERP studies have shown that these perceptually basic category dimensions are encoded at very early stages of visual processing regardless of their relevance to the perceiver's task (e.g., Ebner, He, Fichtenholtz, McCarthy, & Johnson, 2011; Freeman, Ambady, & Holcomb, 2010; Ito & Urland, 2003; Kubota & Ito, 2007; Mouchetant-Rostaing & Giard, 2003; for a review, see Amodio, Bartholow, & Ito, 2014). For example, Ebner and colleagues (2011) demonstrated that the age of the target faces (young vs. old) affected electrophysiological responses as soon as 160 ms after stimulus onset.

Furthermore, it appears that the encoding of these basic category dimensions occurs even when faces are displayed in suboptimal conditions. For example, research has shown that manipulations that typically interfere with the encoding of the target's identity, like face inversion, blurring, or brief presentation, have little or no impact on the extraction of category information (e.g., Cloutier & Macrae, 2007; Cloutier, Mason, & Macrae, 2005; Macrae, Quinn, Mason, & Quadflieg, 2005).

How to Deal With Multiple Category Dimensions?

In much of the early research on social categorization, the emphasis was on studying the effects of a single focal category dimension on the person perception process (Macrae & Bodenhausen, 2000). Researchers deliberatively manipulated stimulus faces such that they only varied along the dimension of interest (e.g., gender) while keeping all other possible dimensions constant (e.g., age). Although this approach revealed important insights, real life is much more complex. For example, when actual persons are encountered, multiple category dimensions can be readily extracted from the face alone, and from these dimensions, multiple stereotypical inferences, some of them contradictory, can be drawn. Thus, it is important to understand how perceivers deal with this complexity problem. Do they focus solely on one dimension or do they simultaneously categorize along multiple dimensions (for recent reviews, see Bodenhausen & Peery, 2009; Crisp & Hewstone, 2007)?

According to one theoretical account, perceivers solve the complexity problem by focusing on one relevant category dimension while inhibiting other nonrelevant, but applicable, dimensions (Bodenhausen & Macrae, 1998; Brewer, Ho, Lee, & Miller, 1987; Hewstone, Islam, & Judd, 1993; Roccas & Brewer, 2002). That is, although perceivers might be perceptually sensitive to nonrelevant category dimensions, the first dimension to reach a certain activation threshold will come to dominate social perception. Nonrelevant category dimensions will be actively inhibited so that perceivers can coherently focus on the dimension of interest (Bodenhausen & Macrae, 1998; Macrae, Bodenhausen, & Milne, 1995; Sinclair & Kunda, 1999, 2000). Thus, according to this account, any category dimension can become temporarily dominant depending on its contextual and motivational relevance (for a review of the main determinants of category dominance, see Bodenhausen, Todd, & Becker, 2007).

The perspective that perceivers are able to focus only on one relevant category dimension is, however, inconsistent with a number of findings showing that nonrelevant dimensions can interfere with the processing of relevant dimensions (e.g., Carpinella, Chen, Hamilton, & Johnson, 2015; Freeman, Nakayama, & Ambady, 2013; Hess, Adams, & Kleck, 2004; Johnson, Freeman, & Pauker, 2012; Purdie-Vaughns & Eibach, 2008; Smith & Zarate, 1992; Stroessner, 1996; Zárate & Smith, 1990; for a review, see Johnson, Lick, & Carpinella, 2015). For example, it has been shown that race and gender mutually influence one another, such that manipulating the race of target faces systematically biased gender categorizations (Johnson et al., 2012) while varying the masculinity or femininity of facial cues affected race categorizations (Carpinella et al., 2015).

The Present Research

As demonstrated in the previous paragraphs, research has yet to find a satisfactory solution to the problem of how the social perceiver deals with multiple category dimensions. We argue that a contributing factor for this knowledge gap is the neglect of experience and learning. In our view, the problem of dealing with multiple category dimensions is a problem of control, of establishing processing priorities between these dimensions according to context constraints and goal requirements. Several authors have argued that there are two types of control: proactive and reactive (e.g., Braver, 2012; Purmann, Badde, & Wendt, 2009; Ridderinkhof, 2002). Proactive control is the process by which the sustained and anticipatory maintenance of goal-relevant information is accomplished, and it is required for adaptive cognitive performance. Reactive control, on the other hand, reflects transient conflict detection and unpredicted stimulus-driven interference obstacles derived from task-irrelevant previous learning (Braver, 2012). One typical characteristic of the process of adaptive proactive control is that it requires knowledge about, experience, and practice with the relevant task to be implemented. Typically, social categorization paradigms deal with a single study and test cycle (e.g., Klauer et al., 2003), the residual effects of previous tasks (e.g., Macrae et al., 1995), novel tasks for participants or with surprising requirements (e.g., Freeman & Johnson, 2016), and/or aleatory changes in the required task responses (e.g., Carpinella et al., 2015; Quinn & Macrae, 2005). Under these conditions, reactive dominates proactive control and responses are typically suboptimal, with substantial interference effects from task-irrelevant attributes.

Therefore, in the present research, we sought to further explore how people deal with multiple category dimensions under conditions that allow for proactive control. Specifically, we focused on the relationship between gender and age under several cycles of study and test. We chose these two category dimensions because they have received considerably much less attention when compared with gender and race (e.g., Cloutier, Freeman, & Ambady, 2014; Klauer et al., 2003; Quinn & Macrae, 2005). To do so, we drew on recent developments in memory research suggesting that the act of retrieving information from memory is a powerful learning tool, both in boosting future recallability and in potentiating subsequent encoding (e.g., Arnold & McDermott, 2013; Cho, Neely, Crocco, & Vitrano, 2017; Garcia-Marques, Nunes, Marques, Carneiro, & Weinstein, 2015; Karpicke & Roediger, 2007).

According to recent research by Garcia-Marques and colleagues (2015), the experience of retrieval can serve as a powerful cue for the most effective ways to encode similar information in comparable future study episodes. The idea tested by these researchers was that repeatedly retrieving information from memory furnishes participants with knowledge regarding the specificities of the retrieval task, allowing for the adaptive control of subsequent encoding and/or retrieval strategies. In their study, participants went through several successive cycles, each composed of a study phase followed by a test phase. What was manipulated across cycles was only the structure of the test phases. No instruction-based manipulations were used. In one between-participants condition, the lures were exemplars from different taxonomic categories than the exemplars previously studied (e.g., animal exemplars + exemplars from different categories); in the other condition, lures and studied exemplars all belonged to the same category. Thus, for the participants in the former condition, the structure of the test encouraged the use of a conceptually based encoding strategy in the next cycle's study phase. That is, the identification of the category of the exemplars included in each study list was sufficient to assure good performance in the subsequent recognition test. For the participants in the latter condition, however, the test did not encourage the use of conceptually based encoding strategy. For these, the best encoding strategy was to attend to the distinctive features of the studied exemplars.

More important, in the last cycle, all participants were tested with lures from the same category of the previously studied exemplars. This means that the structure of the test changed for the participants who had been tested in the previous cycles with the lures from a different category. Thus, if these participants learned the structure of the tests and adapted their encoding strategies accordingly, they should have attended more to the category membership of the exemplars and less their distinctive features. If this was the case, then the recognition performance of these participants in this cycle should have dropped abruptly comparing with their performance in the previous cycles. Indeed, this was what the results of this study showed. Actually, their performance was so low that it stayed at the level of the participants who were always tested with lures from the same category.

The hypothesis examined in the current research stems directly from the ideas just described. Specifically, we predicted that the requirements of past and (expected) future tests would determine participants' encoding of age and gender information. Thus, if in previous tests participants were probed on the gender of a set of target faces, we expected them to learn the relevance of gender to the test and gradually shift their attention to the gender of new faces and away from the age of the same faces. As a consequence, they should show better memory performance for gender than age. We expected the reverse pattern when age was the test-relevant dimension, namely when previous tests probed participants regarding the age of the faces. Hence, in line with previous work (see Bodenhausen & Macrae, 1998; Macrae et al., 1995; Sinclair & Kunda, 1999, 2000), our hypothesis also predicts that one category dimension can temporarily dominate social perception. However, such temporary dominance will be the result of proactive control implemented throughout the successive study-test cycles.

Our study differs from previous social categorization studies, particularly those that have used the Who-Said-What paradigm (e.g., Klauer et al., 2003; Stangor et al., 1992; Taylor et al., 1978), in two main aspects. First, we manipulated the relevance of social dimensions in the retrieval phase while leaving the encoding phase untouched. To our knowledge, all the existing studies have oriented participants' encoding processes by delivering their manipulations during the instructions to the task or during encoding phase itself. For example, Klauer et al. (2003) made the participants focus on a given relevant dimension using stereotypical discussion topics of each category dimension.

Second, because we were interested in understanding the dynamic changes in encoding strategies as a result of previous test requirements, we created an experimental paradigm that gives the opportunity for participants to learn from the tests. We believe this is an important feature of our paradigm because it not only resembles the way we acquire new information in everyday life but also seems to allow for proactive control to be established as we explained above.

By integrating research in social categorization with recent ideas coming from research on learning and memory (Garcia-Marques et al., 2015), we hope that new insights will be gained about the way people encode multiple category dimensions from faces.

Sample Sizes, Diversity, and Open Science Practices

Because the true effect size of our predicted effect was unknown, we did not conduct an a priori power analysis. Instead, we relied on previous research using a similar multiple study-test cycles paradigm (see Garcia-Marques et al., 2015) to determine our minimum sample sizes. In that research, we found reliable effects with samples around 40 participants per condition. Thus, in the present studies, we aimed having at least 40 participants per between-subjects condition. Whenever possible (i.e., if there were students still needing course credits or we still had research funds available), we went beyond the minimum predetermined sample size (i.e., 40) and tried to recruit as many participants as we could during the term of the study. The data for each study were collected in one shot without prior statistical analyses. We report all data exclusions, all manipulations, and all measures. All materials and data are available at the Open Science Framework website (osf.io/5ucr9).

The samples used in this research are homogenous in terms of nationality (Portuguese), racial group (White), and socioeconomic characteristics (university students). We chose these samples because our research was guided by the assumption that the processes under study are relatively low level and, thus, likely to be universal. Previous research indicates that the extraction of social category dimensions from faces is robust across different cultural groups (e.g., Cloutier et al., 2014; Freeman et al., 2013; Klauer et al., 2003).

Study 1

To test our hypothesis, we developed a new paradigm that combines features from Garcia-Marques et al.'s (2015) paradigm with features from the "Who Said What?" paradigm (Taylor et al., 1978). Thus, in our paradigm, participants went through four study-test cycles. In the study phases, participants viewed different pairs of picture-statement. The pictures depicted faces of young and older male and female individuals. The statements were about topics irrelevant for gender and age. After each study phase, participants' memory was probed. We manipulated the structure of the test in the first three cycles. Namely, participants were asked to either retrieve the gender (gender-relevant condition) or the age (age-relevant condition) of the faces accompanying each statement. With this manipulation, we expected participants to gradually learn which social dimension was relevant for test performance and which one was not. More important, in the fourth and last cycle, the structure of test was the same in both conditions. Namely, participants were asked to retrieve both the gender and the age of the faces accompanying the statements. For half the statements, they had to retrieve the gender while for the other half they had to retrieve the age. If participants indeed learned the test structure and adapted their encoding strategies accordingly, they should have a better performance (regarding memory accuracy and response time) on the test-relevant dimension than the test-irrelevant dimension.

Method

Participants. We recruited 87 students (66 women and 21 men; $M_{\text{age}} = 23.46$ years, $SD_{\text{age}} = 5.93$ years) from the University of Lisbon. They received monetary compensation for their participation. In all studies reported here, participants gave informed consent before participation. The Ethics Committee of the Faculty of Psychology of the University of Lisbon approved this research.

Materials. The stimulus materials consisted of pictures of faces and statements. Sixty-four pictures of unfamiliar faces were selected from The Center for Vital Longevity Face Database (Minear & Park, 2004). Thirty-two pictures depicted young faces and 32 depicted old faces.[1] Within each age category, half were men and half women. For each picture used in this and in the following studies, the background was removed, as well as any particular feature that could make the face particularly distinctive (e.g., necklaces and earrings). All pictures were then grayscaled and standardized to the approximate size of 140 × 186 pixels.

Sixty-four statements were constructed. These statements were related with four different topics: 16 statements referred to aspects that could be improved in the city of Lisbon; 16 to cinema and cultural preferences; 16 were about traveling destines and habits; and 16 were about cuisine and food preferences. All statements had roughly the same length and were neutral regarding the gender and age dimensions (see Table 25 in the supplemental materials for the complete list of statements used in all studies and details on how these sentences were constructed and selected).

Procedure. The study was run in sessions of up to eight participants at a time. Participants sat in front of CRT computer screens in individual workstations. All instructions and stimuli were presented on the computer. Participants learned that they were going to participate in a study that intended to recreate those everyday situations where one observes a group of unfamiliar people discussing certain topics. They further learned that they would see faces of several individuals, each one accompanied by a statement made by that individual about the topic under discussion, and that their task was simply to pay attention to the different pairs of faces and statements. After this general instruction, participants went through four study-test cycles. At the beginning of each cycle, immediately before the start of the study phase, they saw a screen indicating the topic under discussion in that cycle. The order of presentation of the discussion topics was counterbalanced using a Latin Square design.

Each study phase consisted of 16 faces and 16 statements. In each trial, a face was displayed at the center of the screen with the respective statement underneath for 6,000 ms. The intertrial interval was 500 ms. The order of trials was randomized individually for each participant. Each face and statement appeared only once during the multiple study phases. There were an equal number of young men (4), young women (4), old men (4), and old women (4) in each study phase. Taking into account these constraints, the formation of the pairs of faces and statements was randomly by the computer, individually for each participant.

Following each study phase, participants completed a filler task (math equations) for 1 min, and then received a memory test. In this test, only the statements from the previous study phase were presented. Specifically, in each test trial a statement was shown at

[1] The age of the junior faces used in the five studies ranged from 18 years old to 32 years old while for senior faces the ages ranged from 65 years old to 88 years old.

the center of the screen accompanied by a probe question (displayed at the top of the screen) and by two response options (displayed at the bottom of the screen). In the first three study-test cycles, the test was manipulated between participants. Namely, half of the participants were asked whether the statement was made by a young or an old person (age-relevant condition; n = 43), while the other half was asked whether the statement was made by a male or a female person (gender-relevant condition; n = 44). Depending on the condition, for half of the statements (eight) the correct option was man or young while for the other half it was woman or old. There was no time limit to respond, but participants were instructed to do it without too much thought. After they completed the test phase, a new study phase was immediately initiated.

The fourth and last test phase was identical in both conditions. Namely, for half of the statements, participants were probed on the same focal category dimension of the previous tests, while for the remaining half of the statements they were probed on the nonfocal dimension (i.e., the dimension that had been irrelevant until then). Thus, participants in the age-relevant condition continued making judgments about the age of the faces (on eight statements), exactly as in the previous test phases, but now they also had to retrieve from memory the gender of the faces (on eight statements). Similarly, in the gender-relevant condition, in addition to retrieving the gender of the faces, participants also had to retrieve their age. To guarantee that participants would easily realize that they needed to make two different judgments, we presented the new probe question, and the respective response options, in a different color (green). The assignment of the sentences to the type of probe question was counterbalanced across participants such that all 16 sentences received both types of probes.

More important, none of the instruction screens presented during the study contained any information about the structure of the subsequent test phases. Therefore, participants only realized what their task was when they started responding to the tests. The idea behind this was to let participants learn the structure and requirements of the tests by doing them. In the end of the study, participants were debriefed, paid, and thanked for their participation.

Results

In this and the following studies, we report η_p^2 as effect-sizes for omnibus and planned analysis of variance (ANOVA) tests. For planned tests, we report 95% confidence intervals (95% CI) around η_p^2.[2] The dependent measures were the number of correct responses and the latencies of those responses. For reasons of brevity and clarity, we only report the results that are directly relevant to our hypothesis. Additional statistical analyses, as well as the full ANOVA tables for all analyses reported in this article, can be found in the online supplemental materials.

A preliminary analysis revealed that four participants responded correctly to all trials and, therefore, their data were not included in the main statistical analyses. This procedure left us with a final sample size of 83 participants: 42 in the gender-relevant condition and 41 in the age-relevant condition. The inclusion of these participants' data in the statistical analyses does not change the results described below.

Performance on cycles 1–3. To assess whether participants learned the structure of the first three tests, we analyzed the number of correct responses in each cycle and the median time[3] participants took to correctly respond. A 2 (test structure: gender-relevant, age-relevant) × 3 (study-test cycle: 1–3) mixed model ANOVA on the number of correct responses did not yield any significant results. As can be seen in Table 1, participants' performance was very high already on the first cycle, which did not leave much room for improvement in the following cycles.

The same mixed model ANOVA on the median response times though revealed a strong main effect of study-test cycle, $F(2, 162) = 87.73$, $p < .001$, $\eta_p^2 = .52$. If participants learned the structure of tests, then there should have been a decrease in response times from cycle to cycle. To test this prediction, we ran a linear contrast that was highly significant, $F(1, 81) = 135.26$, $p < .001$, $\eta_p^2 = .62$, 95% CI [.49, .71]. There was no interaction between study-test cycle and test structure, which suggests that the same data pattern was obtained for the two test structure conditions (see Table 1).

Performance on cycle 4. We submitted the number of correct responses on the fourth test to a 2 (test structure: gender-relevant, age-relevant) × 2 (probed dimension: age, gender) mixed-model ANOVA, with test structure manipulated between-participants and probed dimension manipulated within-participant. This analysis revealed a significant interaction between test structure and probed dimension, $F(1, 81) = 4.05$, $p = .047$, $\eta_p^2 = .05$ (see Figure 1). As predicted, in the gender-relevant condition, participants had better memory for the gender than for the age of the faces, $F(1, 81) = 7.54$, $p = .007$, $\eta_p^2 = .08$, 95% CI [.01, .21]. However, in the age-relevant condition, participants' memory performance was equally good for the two dimensions, $F(1, 81) = 0.01$, $p = .908$, $\eta_p^2 = .001$, 95% CI [.00, .03].

We performed the same ANOVA on median response times. This analysis showed a significant interaction between test structure and probed dimension, $F(1, 81) = 11.61$, $p = .001$, $\eta_p^2 = .12$. In the gender-relevant condition, participants were faster to correctly retrieve the gender (M = 3,081 ms, SE = 174 ms) than the age (M = 3,406 ms, SE = 154 ms) of the faces, $F(1, 81) = 8.20$, $p = .005$, $\eta_p^2 = .09$, 95% CI [.01, .22]. In contrast, in the age-relevant condition, there was a trend toward a faster retrieval of age (M = 3,267 ms, SE = 156 ms) than gender (M = 3,494 ms, SE = 176 ms) information, $F(1, 81) = 3.84$, $p = .053$, $\eta_p^2 = .04$, 95% CI [.00, .16]. There were no more statistically significant results.

Discussion

Study 1 tested the hypothesis that participants can learn from the test requirements which category dimensions are the relevant ones and use that knowledge to tune their encoding strategies when confronted with new faces. As predicted, participants became faster from test to test, which suggests that they indeed learned the structure of the tests. However, memory accuracy showed no significant improvement across the three first cycles. This might

[2] Partial eta-squares and 95% CI were calculated using STATA statistical software (StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP).

[3] We used the median in our analysis because the median is less affected by outliers than the mean (Ratcliff, 1993). However, using the mean does not change any of the results reported in this article.

Table 1
Mean Number of Correct Responses and Mean Median Response Times (in Milliseconds), Overall, and by Test Structure Condition, on the Three Study-Test Cycles (Study 1)

Test structure	Cycle 1	Cycle 2	Cycle 3
Gender-relevant	14.83 (.22)	14.74 (.23)	14.88 (.20)
	3,035 (166)	*2,379 (123)*	*2,253 (116)*
Age-relevant	14.63 (.22)	14.95 (.23)	14.93 (.20)
	3,251 (168)	*2,455 (124)*	*2,210 (118)*
Overall	14.73 (.16)	14.84 (.16)	14.90 (.14)
	3,143 (118)	*2,417 (87)*	*2,232 (83)*

Note. Response times are in italics; Standard errors of the means are presented in parentheses.

have been because the task was relatively easy, as demonstrated by the high level of accuracy registered in all cycles. We addressed this issue in the next study.

To test whether the test structure manipulation implemented in the three first cycles impacted encoding of gender and age, we analyzed participants' performance on cycle four. We found that, when the gender was the test-relevant category dimension, participants indeed had better memory for gender than for age. They were also faster retrieving the gender information. However, when the age was the test-relevant dimension, we did not find differences in memory accuracy nor in time between the two dimensions.

Taken together, these results seem to suggest an asymmetry between gender and age categorizations. Namely, when gender was the test-relevant dimension, participants were capable of encoding gender-diagnostic information independently of age. In contrast, when age was the test-relevant dimension, participants encoded the two dimensions to same extent.

Study 2

In Study 2, we sought to replicate and clarify the findings obtained in Study 1. Namely, we tried to understand whether the difference between gender and age in the gender-relevant condition was mainly caused by a more efficient encoding of the test-relevant dimension or, instead, whether it was driven by the inhibition of the test-irrelevant dimension. By inhibition we are not referring to a suppression process operating during retrieval (e.g., Anderson & Bjork, 1994; Anderson, Bjork, & Bjork, 1994), but to an attention control of which information enters working memory during encoding (see Hasher, 2007; Hasher & Zacks, 1988). To help disentangle these two processes (i.e., more efficient encoding of the test-relevance dimension vs. inhibition of the test-irrelevant dimension), we included a new between-participants condition in which all tests required the retrieval of both social dimensions.

We also changed one response label in the age-relevant condition. Namely, because the word "old" in the Portuguese language has a masculine and a feminine version, we decided, in the previous study, to present both versions on the screen (*idoso/a*). However, by doing this we might have unintentionally increased the accessibility of the gender dimension in this condition, which might help explaining the obtained results. Thus, to correct for this problem, in the current study we replaced the label old with "senior," which in the Portuguese language is a gender-neutral word. Additionally, we also added a response deadline to increase the difficulty of the task and, as a consequence, decrease the level of accuracy.

Method

Participants. There were 133 students (93 women and 40 men; M_{age} = 23.35 years, SD_{age} = 6. 06 years; one participant did not report her age) of the University of Lisbon who took part in this study. They received partial course credit in return for their participation.

Materials. In this study, we used 68 pictures of faces, the same 64 of Study 1 plus 4 new ones also selected from The Center for Vital Longevity Face Database (Minear & Park, 2004). Two of these new pictures were of young faces (one man and one woman) and the other two of old faces (one man and one woman). We also constructed 16 new statements, four of each topic, to add to our existing pool; thus, making a total 80 statements. The new faces and statements were used as buffer trials in last study-test cycle, as explained in the section below.

Procedure. The procedure mirrored that of Study 1, with a few differences. The first was a new condition to serve as a baseline in which we probed participants on both categories in all test phases. Namely, for half of the statements, participants had to respond whether it was a man or a woman who said them while for the other half of the statements; they had to indicate whether it was a junior ("jovem") or a senior person who said them. Thus, in Study 2 there were three between-participants conditions: a gender-relevant condition (n = 44), in which participants were probed on gender in the first three cycles and then on gender and age in the fourth cycle; an age-relevant condition (n = 44), where participants were probed on age in the first three cycles and both on age and gender in the last cycle; and a baseline condition (n = 45), in which they were probed on both category dimensions in all cycles.

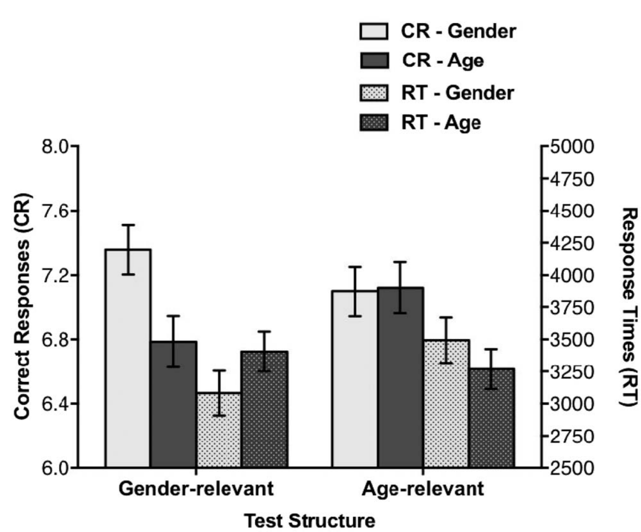

Figure 1. Mean number of correct responses and mean median response times (in milliseconds) as a function of test structure and probed dimension (Study 1). Error bars depict standard errors of the means.

The second difference was that participants were given a response deadline of 4,000 ms; if they did not respond within this time, the following warning appeared at the top of the screen: "The time is over! Please give your response now!". This warning remained on the screen until a response was entered. The computer recorded the responses below and above the 4,000 ms deadline.

The third difference was that the number of trials in the fourth study-test cycle increased from 16 to 20. These extra four pairs were created from the new stimuli described in the Materials section. Namely, in the fourth cycle, participants saw two more young faces (one man and one woman) and two more old faces (one man and one woman) relative to the previous cycles. These faces were paired with four newly created sentences about the topic in question. All 20 pairs were presented in a random order. In the test phase, the four statements from the new pairs were always the first to appear. On two of them, participants were probed on the same focal category of the previous tests, while for the other two they were probed on the nonfocal category. Our goal with the inclusion of these buffer trials was to exclude the possible negative impact of the new test structure on participants' response times. For example, participants might have been surprised with the request to retrieve the irrelevant category, and that might have delayed their responses (for reviews on task switching effects see Kiesel et al., 2010; Monsell, 2003). These trials were not included in the statistical analyses.

Results

A preliminary analysis revealed that one participant from the gender-relevant condition responded incorrectly to more than 50% of all trials (i.e., bellow chance level) and, therefore, her data was not included in the main statistical analyses. This left us with a final sample size of 132 participants. The inclusion of this participant's data in the analyses does not change the results reported below. Next, we report the results for the responses given within the deadline. However, it is important to note that these results do not change when we also include the responses given after the deadline (see the description of these results in the online supplemental materials).

Performance on cycles 1–3. We submitted the number of correct responses provided by each participant within the response deadline to a 3 (test structure: gender-relevant, age-relevant, baseline) × 3 (study-test cycle: 1–3) mixed model ANOVA. This analysis revealed a statistically significant main effect of study-test cycle, $F(2, 258) = 3.61$, $p = .028$, $\eta_p^2 = .03$. Although the linear contrast did not reach statistical significance, $F(1, 129) = 3.40$, $p = .060$, $\eta_p^2 = .03$, 95% CI [.00, .10], the pattern of means (see Table 2, bottom row) suggests that there was some improvement on memory accuracy throughout the cycles. There was also a significant interaction between test-structure and study-test cycle, $F(4, 258) = 2.63$, $p = .035$, $\eta_p^2 = .03$. To interpret this interaction, we ran linear contrasts within each test-structure condition. In the gender-relevant condition, the linear contrast was not significant, $F(1, 129) = 0.11$, $p = .744$, $\eta_p^2 = .00$, 95% CI [.00, .04]. Considering the response deadline, performance was very good in the first cycle (seen in Table 2, first row), which might have left little room for improvement in the subsequent two cycles. Regarding the age-relevant condition, the linear contrast was significant, $F(1, 129) = 14.67$, $p < .001$, $\eta_p^2 = .10$, 95% CI [.02, .21]. Finally,

Table 2
Mean Number of Correct Responses and Mean Median Response Times (in Milliseconds), Overall, and by Test Structure Condition, on the Three Study-Test Cycles (Study 2)

Test structure	Cycle 1	Cycle 2	Cycle 3
Gender-relevant	13.35 (.47)	13.44 (.39)	13.21 (.40)
	2,305 (76)	*2,006 (70)*	*2,071 (76)*
Age-relevant	11.77 (.47)	13.34 (.38)	13.39 (.40)
	2,535 (75)	*2,153 (69)*	*2,003 (75)*
Baseline	11.98 (.46)	12.42 (.38)	11.89 (.40)
	2,583 (74)	*2,325 (68)*	*2,553 (74)*
Overall	12.37 (.27)	13.07 (.22)	12.83 (.23)
	2,474 (43)	*2,161 (40)*	*2,209 (44)*

Note. Response times are in italics; Standard errors of the means are presented in parentheses.

in the baseline condition, the linear contrast was not significant, $F(1, 129) = 0.04$, $p = .831$, $\eta_p^2 = .00$, 95% CI [.00, .03], which suggests that participants' performance remained somewhat stable across cycles.

We again conducted the same analysis on median response times. This analysis revealed a strong main effect of study-test cycle, $F(2, 258) = 42.63$, $p < .001$, $\eta_p^2 = .25$. A linear contrast confirmed that participants became faster across cycles, $F(1, 129) = 76.25$, $p < .001$, $\eta_p^2 = .37$, 95% CI [.24, .48]. There was also a significant interaction between test-structure and study-test cycle, $F(4, 258) = 8.73$, $p < .001$, $\eta_p^2 = .12$. Linear contrasts showed that responses were faster across cycles in the gender-relevant condition, $F(1, 129) = 19.25$, $p < .001$, $\eta_p^2 = .13$, 95% CI [.04, .24] and in the age-relevant condition, $F(1, 129) = 102.37$, $p < .001$, $\eta_p^2 = .44$, 95% CI [.32, .54], but not in the baseline condition, $F(1, 129) = 0.33$, $p = .565$, $\eta_p^2 = .00$, 95% CI [.00, .05] (see Table 2).

Performance on cycle 4. We submitted the number of correct responses given within the response deadline to a 3 (test structure: gender-relevant, age-relevant, baseline) × 2 (probed dimension: age, gender) mixed-model ANOVA. The interaction between test structure and probed dimension was significant, $F(2, 129) = 8.45$, $p < .001$, $\eta_p^2 = .12$ (see Figure 2). To decompose this interaction, we first ran planned contrasts within each test structure condition. Consistent with our hypothesis, participants in the gender-relevant condition had worse memory for age than for gender, $F(1, 129) = 24.58$, $p < .001$, $\eta_p^2 = .16$, 95% CI [.06, .27]. In the age-relevant condition, there was again no difference between gender and age, $F(1, 129) = 0.27$, $p = .601$, $\eta_p^2 = .00$, 95% CI [.00, .04]. Regarding the baseline condition, participants showed similar memory for gender and age, $F(1, 129) = 0.35$, $p = .553$, $\eta_p^2 = .00$, 95% CI [.00, .05]. As clearly illustrated in Figure 2, the difference between gender and age in the gender-relevant condition seems to be driven by a memory loss for age (the irrelevant dimension) rather than a memory gain for gender (the relevant dimension). Indeed, participants in the gender-relevant condition had worse performance for age than participants in the baseline condition, $F(1, 129) = 14.21$, $p < .001$, $\eta_p^2 = .10$, 95% CI [.02, .20]. For the gender category dimension, memory performance was the same between gender-relevant and baseline conditions, $F(1, 129) = 0.02$, $p = .894$, $\eta_p^2 = .00$, 95% CI [.00, .02].

A 3 (test structure: gender-relevant, age-relevant, baseline) × 2 (probed dimension: age, gender) mixed-model ANOVA on median

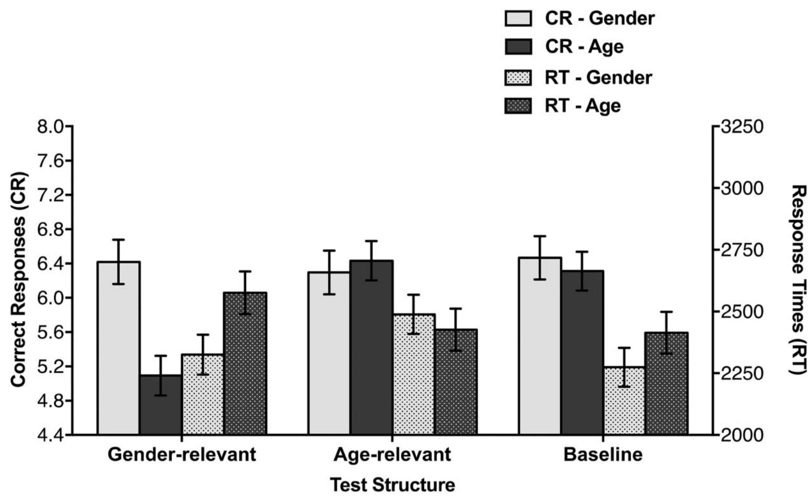

Figure 2. Mean number of correct responses and mean median response times (in milliseconds) as a function of test structure and probed dimension (Study 2). Error bars depict standard errors of the means.

response times also yielded a significant interaction between test structure and probed dimension, $F(2, 129) = 8.44$, $p < .001$, $\eta_p^2 = .12$ (see Figure 2). Consistent with our hypothesis, and with Study 1, participants in the gender-relevant condition took longer to retrieve age than gender information, $F(1, 129) = 20.81$, $p < .001$, $\eta_p^2 = .14$, 95% CI [.05, .25], while in the age-relevant condition there was only a trend in the opposite direction, $F(1, 129) = 1.30$, $p = .255$, $\eta_p^2 = .01$, 95% CI [.00, .07]. Regarding the baseline condition, participants were significantly faster to retrieve gender than age information, $F(1, 129) = 6.76$, $p = .010$, $\eta_p^2 = .05$, 95% CI [.00, .14]. We also compared the time participants took to retrieve gender information in gender-relevant condition versus the baseline condition as well as the time taken to retrieve age information in the same two conditions. None of these comparisons reached statistical significance (all Fs < 3.66 and ps $> .058$).

Discussion

The results of Study 2 replicated and extended those of Study 1. Namely, in the first three cycles, we again observed a decrease in response times, suggesting that participants learned the structure of the tests. Regarding correct responses, the data pattern indicates some improvement (though nonsignificant) across cycles. Although the inclusion of a response deadline resulted in more errors, its impact was not what we expected as the accuracy levels were still very high. Thus, we believe we would have observed a significant improvement if the task had been more difficult. In the next study, we introduced some modifications to our paradigm with the goal making participants' task even more difficult.

Regarding performance on the fourth cycle, we found that participants were more accurate and faster for gender than age when gender was the test-relevant category dimension. As in the previous study, we again did not find any differences between these two dimensions when age was the test-relevant dimension. More important, the results showed that the obtained difference between gender and age in the gender-relevant condition was because of a memory performance decrease for age and not by a boost in memory performance for gender. These results suggest that repeated testing did not lead to an improvement on the encoding of the test-relevant dimension (gender), but rather to an inhibition of the test-irrelevant dimension (age).

Study 3

In Study 2, we found evidence suggesting that participants learned throughout the cycles to inhibit the age of the faces rather than becoming better at encoding their gender. One possible concern about this study, however, is that this result might reflect a ceiling effect for gender information. That is, there might not have been a memory enhancement for gender information because the performance was already so high that it did not leave room for improvement. To address this concern, in Study 3 we modified our paradigm to increase the task difficulty and, thus, decrease participants' performance. If the results of Study 2 were because of a ceiling effect, then reducing participants' performance should give rise to a memory enhancement of gender information.

Method

Participants. There were 151 students (103 women and 48 men; $M_{\text{age}} = 23.23$ years, $SD_{\text{age}} = 4.83$ years; 2 participants did not report their age) of the University of Lisbon took who part in this study. They received a monetary compensation for their participation.

Materials. We added 30 new pictures of faces to the pool of 68 used in the previous studies. The new faces were also taken from The Center for Vital Longevity Face Database (Minear & Park, 2004). Thus, in this study we used 48 young and 48 old faces, half of which were men and half were women. Two extra faces, a young female and a young male, were used as buffers in the last cycle.

We used the same 80 statements of the previous studies to which we added 24 newly constructed statements; thus, making a total 104 statements. There were 26 statements about each topic: aspects that could be improved in the city of Lisbon; cinema and cultural preferences; traveling destines and habits; cuisine and food preferences (see Table 25 in the supplemental materials).

Procedure. The procedure was similar to that of the previous studies, with some differences. The most important difference was that, in each one of the four study phases, the faces were presented before the statements and for a very short period. Namely, each study trial had the following structure: there was a blank screen for 1,000 ms, followed by a fixation mark (+) for 500 ms; then, a face was presented for 300 ms; fixation mark and face were located slightly above the center of the screen; after the face disappeared, there was a blank screen for 250 ms; finally, a statement was shown slightly below the center of the screen during 4,000 ms.

The second difference was that we increased the number of face-statement pairs in each cycle. Namely, participants saw 24 different pairs in the first three cycles and 26 pairs in the fourth cycle. The two extra pairs in the fourth cycle were used as buffers and were not included in the statistical analyses. In the study phase, the two buffer pairs were presented in a random order together with the other pairs. In the test phase, the sentences that composed the buffer pairs were always the first to be presented. On one of these sentences, we probed participants on the same focal category of the previous tests, while on the other we probed them on the nonfocal category.

The third and last difference relative to previous studies was that, in the test phases, the statement remained on the screen for only 4,000 ms. If participants did not provide a response within this deadline, they received a feedback message on the screen ("Respond faster!"), followed by a new statement.

Results

A preliminary analysis revealed that the overall performance of six participants was bellow chance level and, therefore, their data was not included in the main statistical analyses. This procedure left us with a final sample size of 145 participants: 49 participants in the gender-relevant condition, 49 in the age-relevant condition, and 47 in the baseline condition. Including the data from these six participants in the analyses does not change the results reported next.

Performance on cycles 1–3. A 3 (test structure: gender-relevant, age-relevant, baseline) × 3 (study-test cycle: 1–3) mixed model ANOVA on the number of correct responses revealed a statistically significant effect of study-test cycle, $F(2, 284) = 39.70$, $p < .001$, $\eta_p^2 = .22$. To test whether the data showed the predicted positive linear pattern, we ran a linear contrast that was highly significant, $F(1, 142) = 67.44$, $p = .001$, $\eta_p^2 = .32$, 95% CI [.20, .43]. There was no interaction between test structure and study-test cycle, which suggests that there was a performance improvement in all test structure conditions (see Table 3).

The same mixed model ANOVA on response times also showed a main effect of study-test cycle, $F(2, 284) = 54.39$, $p < .001$, $\eta_p^2 = .28$. As expected, there was a significant linear pattern between the first and the third cycles, $F(1, 142) = 103.50$, $p = .001$, $\eta_p^2 = .42$, 95% CI [.30, .52]. Again, there was no interaction between test structure and study-test cycle (see Table 3).

Performance on cycle 4. A 3 (test structure: gender-relevant, age-relevant, baseline) × 2 (probed dimension: age, gender) mixed-model ANOVA on the number of correct responses given within the response deadline revealed that the interaction between test structure and probed dimension was once again significant, $F(2, 142) = 4.08$, $p = .019$, $\eta_p^2 = .05$ (see Figure 3). As predicted, participants performed worse for age than for gender in the gender-relevant condition, $F(1, 142) = 7.41$, $p = .007$, $\eta_p^2 = .05$, 95% CI [.004, .13]. The opposite did not happen in the age-relevant condition, namely participants performed equally well for gender and age, $F(1, 142) = 0.55$, $p = .459$, $\eta_p^2 = .00$, 95% CI [.00, .05]. In the baseline condition, participants also showed similar memory performance for both social dimensions, $F(1, 142) = 0.67$, $p = .413$, $\eta_p^2 = .00$, 95% CI [.00, .05]. As in the previous study, participants in the gender-relevant condition performed worse on the age dimension (the irrelevant dimension) than did participants in the baseline condition, $F(1, 142) = 18.15$, $p < .001$, $\eta_p^2 = .11$, 95% CI [.03, .21]. For the gender dimension, memory performance was similar between participants in the gender-relevant and baseline conditions, $F(1, 142) = 1.20$, $p = .275$, $\eta_p^2 = .01$, 95% CI [.00, .06].

Regarding the response times, the 3 (test structure) × 2 (probed dimension) mixed-model ANOVA did not show a significant interaction between test structure and probed dimension, $F(2, 142) = 0.74$, $p = .477$, $\eta_p^2 = .01$. As can be seen on Figure 3, participants were faster for gender than for age in all conditions.

Table 3
Mean Number of Correct Responses and Mean Median Response Times (in Milliseconds), Overall, and by Test Structure Condition, on the Three Study-Test Cycles (Study 3)

Test structure	Cycle 1	Cycle 2	Cycle 3
Gender-relevant	16.45 (.45)	18.59 (.42)	19.18 (.43)
	2,086 (60)	*1,972 (61)*	*1,863 (55)*
Age-relevant	17.90 (.47)	19.59 (.42)	20.12 (.43)
	2,131 (60)	*1,938 (61)*	*1,797 (55)*
Baseline	17.02 (.45)	19.30 (.43)	19.36 (.44)
	2,461 (62)	*2,197 (63)*	*2,129 (56)*
Overall	17.12 (.26)	19.16 (.24)	19.55 (.25)
	2,226 (35)	*2,036 (36)*	*1,930 (32)*

Note. Response times are in italics; Standard errors of the means are presented in parentheses.

Discussion

The results of Study 3 corroborate the key findings from Studies 1 and 2. Specifically, we replicated the finding that gender information is better recalled than age information in the gender-relevant condition. Again, there was no evidence of a memory advantage of age over gender information in the age-relevant condition. Importantly, we also replicated the finding from Study 2 showing that the difference between gender and age is caused by a memory performance decrease for the age information. Because the modifications introduced to the paradigm significantly reduced participants' performance, this result rules out the possible concern that the obtained results reflect a ceiling effect for gender information. Surprisingly, in this study, the response times did not accompany memory accuracy as we did not find an interaction between test structure and probed dimension on response times. Participants were faster for gender information in the three conditions.

Regarding performance on the three first cycles, we found evidence of test structure learning not only in response times but also in memory accuracy. That is, we found that participants

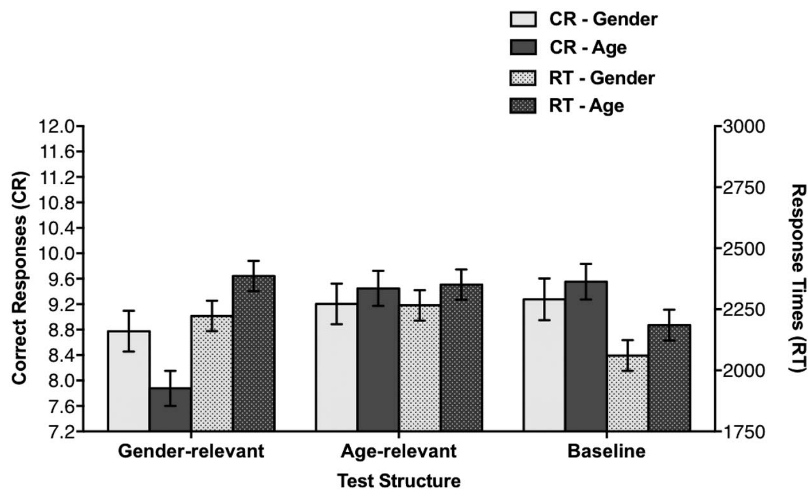

Figure 3. Mean number of correct responses and mean median response times (in milliseconds) as a function of test structure and probed dimension (Study 3). Error bars depict standard errors of the means.

became faster and more accurate throughout the cycles. In the next study, we used a new task to further test if the age dimension was indeed inhibited was a result of its test-irrelevance.

Study 4

Studies 1–3 provided converging evidence for an asymmetrical relationship between gender and age, such that gender dominates face processing when it is test-relevant, and it remains highly salient when it is test-irrelevant. Studies 2 and 3 found evidence suggesting that the gender dominance pattern obtained when gender is test-relevant is driven by the inhibition of age information (the test-irrelevant dimension) rather than by an increased memory for gender information. In Study 4, we further examined the conclusion that participants inhibited the age dimension.

One signature of category activation is the enhanced accessibility of its corresponding knowledge structures (e.g., stereotypes; Brewer, 1988; Devine, 1989; Fiske & Neuberg, 1990). As widely demonstrated, accessible stereotypes provide the perceiver with expectancies that shape subsequent information processing (for a review, see Macrae & Bodenhausen, 2000). For example, people take less time reading expectancy-consistent than expectancy-inconsistent information following stereotype activation (e.g., Brewer, Dull, & Lui, 1981; Stern, Marrs, Millar, & Cole, 1984; for other similar expectancy-based effects on reading times, see, e.g., Bargh & Thein, 1985; Belmore, 1987; O'Brien, Rizzella, Albrecht, & Halleran, 1998; Pérez, Joseph, Bajo, & Nation, 2016; Ziegler & Burger, 2011).

Thus, if the age dimension becomes inhibited as a result of our test-structure manipulation, we predict that the corresponding stereotypes will become less accessible to participants and, therefore, will have no impact on subsequent information processing. Put simply, if one cannot recall whether someone is young or old, one will not be influenced by the corresponding age stereotypes. To test for the presence of inhibition, we designed a task inspired by the extensive research showing expectancy-based effects on reading times. Namely, after the three study-test cycles where participants either had to retrieve the gender or the age of faces, they were presented with a new task. They saw new faces paired with behavioral sentences. These sentences described behaviors that were congruent or incongruent with the age stereotype and simultaneously neutral regarding the gender stereotype gender. The participants' task was to establish a mental connection between each pair of face-behavior. The task was self-paced, and we measured the time spent attending to each pair. If the age dimension indeed got inhibited throughout the study-test cycles, it should not influence performance in this task. Namely, one should not observe the typical reading time difference between stereotype congruent (i.e., lower reading times) and incongruent information (i.e., higher reading times). On the other hand, if the age category was still active, then the time difference between stereotype congruent and incongruent information should emerge (see Brewer et al., 1981; Stern et al., 1984).

Note that this task is independent of the test-structure manipulation implemented in the first three study-test cycles, which supposedly is causing the inhibition of the age dimension. Thus, unless the age dimension was inhibited in memory, its corresponding stereotypes should influence the processing of stereotype congruent and incongruent information (for a similar rationale, see, e.g., Anderson & Spellman, 1995). This is a crucial feature of our paradigm because "in order to make a strong claim in any study about the presence or absence of inhibition, or about variations in the magnitude of inhibition as a function of condition or population, it is necessary to include an independent probe of the impaired items' accessibility" (Anderson & Levy, 2007, p. 82).

Method

Participants. There were 128 students (108 women and 20 men; M_{age} = 19.10 years, SD_{age} = 3.14 years; 2 female participants did not report their age) of the University of Lisbon who took part in this study. They received partial course credit for their participation.

Materials. We used 84 faces from the pool used in the previous studies (Minear & Park, 2004). From these, 42 were young faces (21 men and 21 women), and 42 were old faces (21 men and

21 women). In this study, we used statements from only three discussion topics: 20 about aspects that could be improved in the city of Lisbon; 20 about traveling destines and habits; and 20 about cuisine and food preferences. Besides, we also used sentences describing stereotypical behaviors of young (e.g., Got a tattoo this week) and old (e.g., Took the medications during lunch) people. These behaviors were neutral regarding the gender dimension. See the supplemental materials (Table 26) for more details on development and selection of these stimuli.

Procedure. The procedure was similar to that of Studies 1 and 2, with an important difference. Again, participants read that the study intended to recreate the everyday situations where one observes a group of unknown people discussing specific issues. Then, participants went through three study-test cycles where they saw, in the study phases, 20 different pairs of face-statement (for 5,000 ms each). In the test phases, participants had 4,000 ms to answer to each statement. The computer recorded only the responses provided within this deadline. The discussion topics used in these three cycles were: aspects that could be improved in the city of Lisbon; traveling destines and habits; cuisine and food preferences. The order of presentation of the topics throughout the three cycles was counterbalanced using a Latin Square design.

Different from the previous studies, there was no fourth cycle. Instead, participants were given a new task. As explained before, the goal of this task was to measure the activation of the age dimension. Namely, they received an instruction on the computer screen saying they would see several pictures of faces, each accompanied by a sentence describing a behavior performed by that person. Instructions informed them they should pay attention to each pair because in a later phase they were going to receive a memory test about the pairs. They were told to press the spacebar key when they felt they had learned the association between the person in the picture and the respective behavior described in the sentence. They had a maximum of 15 s to learn each pair, after which a new pair would appear on the computer screen. We recorded the time that participants took learning each pair, that is, the time elapsed between the onset of a pair and the participant's spacebar press. Participants saw a total of 26 pairs: 6 with young faces (3 men and 3 women) and stereotypical behaviors of young people; 6 with young faces (3 men and 3 women) and counterstereotypical behaviors of young people (i.e., stereotypical of old people); 6 with old faces (3 men and 3 women) and stereotypical behaviors of old people; and 6 with old faces (3 men and 3 women) and counterstereotypical behaviors of old people (i.e., stereotypical of old young). The composition of these pairs was randomly determined by the computer, individually for each participant. After seeing all pairs, the study ended without their memory being tested. They were then adequately debriefed, paid, and thanked for their participation.

Results

A preliminary analysis revealed that, on the fourth cycle, three participants pressed the spacebar faster than 1,000 ms in more than 50% of the trials, which suggests they did not try to establish a connection between faces and behaviors as requested by the instructions. The data from these participants were, thus, excluded from all analyses. This procedure left us with a final sample size of 125 participants: 65 participants in the gender-relevant condition and 63 in the age-relevant condition. The inclusion of these participants' data does not change the results reported next.

Performance on cycles 1–3. A 2 (test structure: gender-relevant, age-relevant) × 3 (study-test cycle: 1–3) mixed model ANOVA on the number of correct responses revealed only a significant main effect of study-test cycle, $F(2, 246) = 31.76$, $p < .001$, $\eta_p^2 = .20$. As predicted, participants' performance across cycles exhibited a positive linear shape, $F(1, 123) = 45.30$, $p < .001$, $\eta_p^2 = .27$, 95% CI [.14, .38]. We conducted the same analysis on response times and the results showed again only a significant main effect of study-test cycle, $F(2, 246) = 27.88$, $p < .001$, $\eta_p^2 = .18$. Replicating previous results, the linear contrast revealed that participants became significantly faster throughout the cycles, $F(1, 123) = 42.65$, $p < .001$, $\eta_p^2 = .26$, 95% CI [.13, .37] (see Table 4).

Reading times task. Participants pressed the spacebar in 99.93% ($SD = 3.66\%$) of the trials. We calculated the median time each participant took to press the spacebar in each condition. A 2 (test structure: gender-relevant, age-relevant) × 2 (type of pair: congruent, incongruent) × 2 (target age: young, old) mixed-model ANOVA on the median response times unveiled the following results. Consistent with previous research, there was a main effect of type of pair, $F(1, 123) = 10.03$, $p = .002$, $\eta_p^2 = .07$, such that participants took longer to press the spacebar when confronted with incongruent ($M = 3{,}752$ ms, $SE = 236$ ms) than with congruent pairs ($M = 3{,}585$ ms, $SE = 223$ ms). The main effect of test structure was also significant, $F(1, 123) = 7.12$, $p = .009$, $\eta_p^2 = .05$, namely, participants in the age-relevant condition processed the face-behavior pairs faster ($M = 3{,}242$ ms, $SE = 225$ ms) than participants in the gender-relevant condition ($M = 4{,}095$ ms, $SE = 227$ ms). The predicted interaction between test structure and type of pair, however, did not reach statistical significance, $F(1, 123) = 1.05$, $p = .308$, $\eta_p^2 = .01$. Despite the nonsignificant interaction, and to provide a direct test to our hypotheses, we conducted the corresponding planned contrasts. Contrasts showed a significant difference between congruent and incongruent pairs in the age-relevant condition ($M_{congruent} = 3{,}131$ ms, $SE_{congruent} = 314$ ms vs. $M_{incongruent} = 3{,}353$ ms, $SE_{incongruent} = 332$ ms), $F(1, 123) = 8.85$, $p = .003$, $\eta_p^2 = .07$, 95% CI [.01, .16], but not in the gender-relevant condition ($M_{congruent} = 4{,}038$ ms, $SE_{congruent} = 316$ ms vs. $M_{incongruent} = 4{,}152$ ms, $SE_{incongruent} = 335$ ms), $F(1, 123) = 2.28$, $p = .134$, $\eta_p^2 = .02$, 95% CI [.00, .09]. The three-way interaction between test structure, type of pair, and target age did

Table 4
Mean Number of Correct Responses and Mean Median Response Times (in Milliseconds), Overall, and by Test Structure condition, on the Three Study-Test Cycles (Study 4)

Test structure	Cycle 1	Cycle 2	Cycle 3
Gender-relevant	15.23 (.30)	16.89 (.25)	16.76 (.30)
	2,133 (55)	*1,996 (48)*	*1,957 (50)*
Age-relevant	15.25 (.29)	16.44 (.25)	16.97 (.29)
	2,169 (54)	*1,967 (48)*	*1,887 (50)*
Overall	15.24 (.21)	16.67 (.18)	16.86 (.21)
	2,151 (38)	*1,982 (33)*	*1,922 (35)*

Note. Response times are in italics; Standard errors of the means are presented in parentheses.

not reached statistical significance, $F(1, 123) = 3.48$, $p = .064$, $\eta_p^2 = .03$ (see Table 21 in the supplemental materials for the complete ANOVA table).

Discussion

In Study 4, we further tested the hypothesis that the age dimension was inhibited in the gender-relevant condition, as a result of the test structure manipulation employed in the first three cycles. To do so, we included a task designed to access the activation of age stereotypes. We reasoned that, if the age dimension was indeed inhibited in the gender-relevant condition, one should not observe a manifestation of age stereotypes on the reading times task. Conversely, in the age-relevant condition, we expected stereotype activation effects on reading times. Consistent with previous research showing expectancy-based effects on reading times (e.g., Brewer et al., 1981; Stern et al., 1984), results showed longer reading times for incongruent than congruent pairs. More important, there was a significant difference between congruent and incongruent pairs in the age-relevant condition, but that difference failed to emerge in the gender-relevant condition (i.e., participants spent the same time attending to congruent and incongruent pairs). This pattern of results is consistent with our hypothesis, although it should be carefully interpreted given that the interaction was not significant.

Study 5

In the previous study, we tested the hypothesis that the age dimension is inhibited in memory as a consequence of its continued irrelevance for the tests in the three initial cycles of the experiment. Inhibition was accessed immediately after the cycles phase using a different task, namely a reading time task. Our inhibition hypothesis predicted an asymmetrical interaction between test structure and type of pair, such that one should have observed a reading time difference between congruent and incongruent pairs in the age-relevant condition but not in the gender-relevant condition. Although the obtained reading times pattern is consistent with our hypothesis, two factors limit the conclusions one can draw from the results. The first is that the interaction between test structure and type of pair did not reach statistical significance. The second is the absence of a baseline condition in which neither gender nor age is test-relevant. Without such condition, it is difficult to understand whether the reading time advantage for the congruent pairs over incongruent pairs is caused by the enhanced activation of the age dimension when the age of the target faces is test-relevant or by inhibition of the age dimension when age is test-irrelevant. To argue for the presence of inhibition, one would need to show that the reading time advantage for congruent pairs in age-relevant and baseline conditions are of similar magnitude and both are larger than an analogous result in the gender-relevant condition. To overcome these limitations, in Study 5 we added a baseline condition consisting only of the reading time task. That is, participants were not exposed to our test structure manipulation. This way we could measure the activation of the age dimension when none of the category dimensions were previously made relevant.

Method

Participants. There were 210 students (170 women and 40 men; $M_{age} = 22.55$ years, $SD_{age} = 6.29$ years) of the University of Lisbon who participated in this study in return for monetary compensation.

Materials. We used the exact same materials of Study 4.

Procedure. The procedure was similar to that of Study 4, with one modification: a new condition consisting only of the reading times task. Thus, participants were randomly assigned to one of three between-participants conditions: gender-relevant condition, age-relevant condition, or baseline condition. As in Study 4, participants assigned to the gender- and age-relevant conditions first went through three initial study-test cycles (where they were always tested on the gender or age dimension, depending on the condition) and then completed the reading times task. In the baseline condition, participants completed only the reading times task. In the reading times task, they were asked to press the spacebar key once they felt they had learned the association between the person in the picture and the behavior described in the sentence presented underneath. Half of the pairs of faces and behaviors were congruent with the age stereotype, and half were incongruent.

Results

The data from five participants were excluded from all analyses because of the reasons detailed next. We excluded the data from two participants in the gender-relevant condition because one did not respond to any trial in cycle 1 and the other did not press the spacebar in any trial of the reading times task. In the age-relevant condition, we also removed the data of two participants because one of them did not respond to any trial in cycle 1 and the other pressed the spacebar faster than 1,000 ms in more than 50% of the trials in the reading times task. Finally, we removed the data of one participant from the baseline condition for pressing the spacebar in less than 50% of the trials. More important, including these participants' data does not change the results reported next. The final sample consisted of 205 participants: 68 participants in the gender-relevant condition, 68 in the age-relevant condition, and 69 in the baseline condition.

Performance on cycles 1–3. This analysis was conducted only on data from the gender- and age-relevant conditions, as the baseline condition did not include the three study-test cycles. A 2 (test structure: gender-relevant, age-relevant) × 3 (study-test cycle: 1–3) mixed model ANOVA on the number of correct responses showed only a significant main effect of study-test cycle, $F(2, 268) = 25.97$, $p < .001$, $\eta_p^2 = .16$. As expected, performance across cycles exhibited a positive linear trend, $F(1, 134) = 53.10$, $p < .001$, $\eta_p^2 = .28$, 95% CI [.16, .39]. No other effects were significant. We conducted the same analysis on response times and the results also revealed a significant main effect of study-test cycle, $F(2, 268) = 48.62$, $p < .001$, $\eta_p^2 = .27$. Consistent with the results from previous studies, the linear contrast showed that participants became significantly faster throughout the cycles, $F(1, 134) = 81.43$, $p < .001$, $\eta_p^2 = .38$, 95% CI [.25, .48] (see Table 5).

Reading times task. Participants pressed the spacebar in 98.43% ($SD = 3.25\%$) of the trials. We submitted participants' reading times to a 3 (test structure: gender-relevant, age-relevant,

Table 5
Mean Number of Correct Responses and Mean Median Response Times (in Milliseconds), Overall, and by Test Structure condition, on the Three Study-Test Cycles (Study 5)

Test structure	Cycle 1	Cycle 2	Cycle 3
Gender-relevant	15.40 (.30)	16.31 (.28)	16.78 (.29)
	2,166 (58)	*2,006 (56)*	*1,893 (57)*
Age-relevant	14.54 (.30)	16.21 (.28)	16.66 (.29)
	2,181 (58)	*1,969 (56)*	*1,847 (57)*
Overall	14.97 (.21)	16.26 (.20)	16.72 (.20)
	2,174 (41)	*1,988 (40)*	*1,870 (41)*

Note. Response times are in italics; Standard errors of the means are presented in parentheses.

baseline) × 2 (type of pair: congruent, incongruent) × 2 (target age: young, old) mixed-model ANOVA. This analysis revealed three significant effects. Replicating previous results, there was a main effect of type of pair, $F(1, 202) = 21.19$, $p < .001$, $\eta_p^2 = .09$, such that participants took longer to press the spacebar for incongruent ($M = 4{,}661$ ms, $SE = 204$ ms) than congruent face-behavior pairs ($M = 4{,}391$ ms, $SE = 198$ ms). The main effect of test structure was also significant, $F(2, 123) = 7.12$, $p = .009$, $\eta_p^2 = .05$. Post hoc tests (Sheffé) showed that participants in the baseline condition took longer processing the face-behavior pairs ($M = 4{,}095$ ms, $SE = 227$ ms) than participants in the age-relevant condition ($M = 3{,}688$ ms, $SE = 242$ ms; $p < .001$) and participants in the gender-relevant condition ($M = 3{,}964$ ms, $SE = 242$ ms; $p < .001$). There was no significant difference between participants in the age-relevant and gender-relevant conditions ($p = .722$). The predicted interaction between test structure and type of pair was significant, $F(2, 202) = 3.68$, $p = .030$, $\eta_p^2 = .03$ (see Figure 4). As hypothesized, participants were faster pressing the spacebar for congruent than incongruent face-behavior pairs in both the age-relevant condition, $F(1, 202) = 15.65$, $p < .001$, $\eta_p^2 = .07$, 95% CI [.02, .15], and baseline condition, $F(1, 202) = 12.77$, $p < .001$, $\eta_p^2 = .06$, 95% CI [.01, .13]. More important, such reading time advantage for congruent pairs was not observed in the gender-relevant condition, $F(1, 202) = 0.20$, $p = .653$, $\eta_p^2 = .00$, 95% CI [.00, .03]. There was no significant three-way interaction between test structure, type of pair, and target age, $F(2, 202) = 0.42$, $p = .660$, $\eta_p^2 = .00$.

To directly compare the magnitude of the reading time advantage of congruent relative to incongruent pairs across the three conditions, we subtracted, for each participant, the time spent on incongruent pairs from the time spent on congruent pairs and submitted these difference scores to an ANOVA, with test structure as the only factor. This analysis revealed a significant main effect of test structure, $F(2, 202) = 4.18$, $p = .017$, $\eta_p^2 = .04$. In line with our hypothesis, post hoc tests (Sheffé) showed that the difference scores obtained in the age-relevant ($M_{\text{cong-incong}} = -409$ ms, $SE = 100$ ms) and baseline ($M_{\text{cong-incong}} = -407$ ms, $SE = 99$ ms) conditions were of similar magnitudes ($p > .999$), and that both were significantly larger than the difference score obtained in the gender-relevant condition ($M_{\text{cong-incong}} = -54$ ms, $SE = 100$ ms; $p = .045$).

Discussion

In Study 5, we sought to address two limitations of Study 4, namely the nonsignificant interaction between test structure and type of pair and the lack of a baseline condition in which neither gender nor age is test-relevant. To this end, we used the same experimental design of Study 4 with one modification: we added a baseline condition in which participants only completed the reading time task. Thus, contrary to what occurred in the other two conditions, in this condition participants were not exposed to the test structure manipulation. Consistent with our hypothesis, we found that participants took less time on congruent than on incongruent face-behavior pairs both when the age-dimension was test-relevant (age-relevant condition) and when neither age nor gender were test-relevant (baseline condition). In contrast, when the gender dimension was test-relevant (gender-relevant condition), participants took approximately the same time on congruent and incongruent pairs.

Two other effects were obtained. We again obtained longer reading times for incongruent than congruent face-behavior pairs; thus, replicating the results of Study 4 and previous findings in the literature (e.g., Brewer et al., 1981; Stern et al., 1984). There was also a main effect of test structure, with longer reading times for the baseline condition relative to the two other conditions. Such a result might be explained by the absence of initial study-test cycles in the baseline condition. One possibility is that reading times were longer in the baseline condition because of less practice. Another possibility is that this difference was instead motivated by the speedup of participants in age-relevant and gender-relevant conditions who had to go through the three study-test cycles before reaching the reading times task. Past research suggests that longer response times often have larger *SD*s, which can complicate interpretations of response time differences (e.g., Wagenmakers & Brown, 2007). However, in this task the *SD* in the baseline condition ($SD = 1{,}886$ ms) was slightly smaller than the *SD*s in the other two conditions ($SD_{\text{age-relevant}} = 1{,}996$ ms, $SD_{\text{gender-relevant}} = 1{,}996$ ms), so this does not compromise the interpretation of our

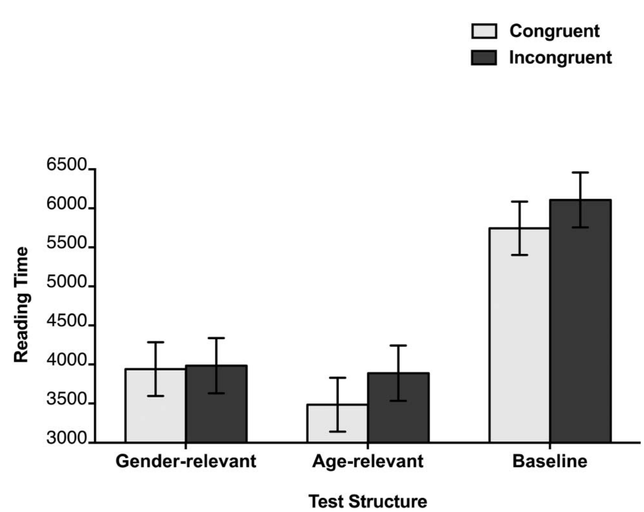

Figure 4. Mean median reading times (in milliseconds) as a function of type of test structure and type of pair. Error bars depict standard errors of the means.

results. Moreover, the observation that longer response times have larger *SD*s is based on means, which are sensitive to skewed distributions, but we have used median reading times in all analyses, which are not sensitive to skew. Thus, taken together, these results provide compelling support to the hypothesis that the age-dimension is inhibited in memory when it is test-irrelevant.

General Discussion

The goal of the current research was to investigate the role of experience and learning on the encoding and retrieval of basic social category dimensions such as age and gender. Drawing on recent memory research (Garcia-Marques et al., 2015), we hypothesized that repeated testing would provide participants with knowledge about the relevance of each category dimension and allow them to control their subsequent encoding strategies accordingly. Specifically, if participants were repeatedly tested on one category dimension (age or gender), they should learn to focus their attention on that dimension and away from the other nonrelevant dimension. We obtained a robust pattern of results across studies. Consistent with our prediction, in Studies 1 to 5 participants became faster at retrieving information regarding the category dimension tested on cycles 1 to 3. In Studies 3 and 4, we also obtained a significant increase in memory accuracy from cycle 1 to cycle 3, while in Studies 1 and 2, because the overall performance was near ceiling level, we only obtained a trend in the expected direction. We interpret these results as strong evidence that participants learned the structure of the tests (i.e., the relevant dimension), as the target materials changed from cycle to cycle, and the test requirements were kept constant.

Regarding the consequences of learning the test structure on the encoding and retrieval of the age and gender category dimensions (i.e., performance on cycle 4), we found different results depending on which dimension was test-relevant on the previous cycles. Namely, when the gender dimension was test-relevant on cycles 1 to 3, participants were faster and more accurate retrieving information concerning the gender than the age of the target faces. On the other hand, when the age dimension was test-relevant on cycles 1 to 3, participants were equally fast and accurate retrieving information concerning age and gender of the faces. We obtained this pattern in Studies 1 to 3. More important, results of Studies 2 and 3 suggest that the difference between gender and age in the gender-relevant condition is mainly because of the inhibition of age information (test-irrelevant information) rather than to an improvement on memory for the gender information.

In Study 4, we provided a new test to the hypothesis that the age dimension was inhibited in the gender-relevant condition. For this purpose, we measured the activation of age stereotypes following three cycles where we manipulated the test-relevance of the age or the gender dimensions. If the age dimension was in fact inhibited in the gender-relevant condition, one should not observe the typical reading time difference between stereotype congruent and incongruent information in this condition. If, however, the age dimension was still active, then one should obtain the time difference between stereotype congruent and incongruent information (e.g., Brewer et al., 1981; Stern et al., 1984). The obtained results are consistent with our predictions, as we found a reading time difference between congruent and incongruent pairs of faces and behaviors in the age-relevant condition but not in the gender-relevant condition. However, the predicted interaction did not reach statistical significance, which constrains our interpretation of the results. Additionally, because we did not include a baseline condition in which neither age nor gender was test-relevant, we did not know what mechanism caused the reading time difference between congruent and incongruent pairs. That is, if it was mainly driven by the activation of the age dimension when age was test-relevant or by the inhibition of the age dimension when age was test-irrelevant.

To address the issues mentioned above, in Study 5 we added a baseline condition to the design used in Study 4. In this condition, participants completed only the reading times task, which allowed us to access the activation of the age dimension in the absence of test structure manipulation. If the age dimension was inhibited in the gender-relevant condition, then the results of the baseline condition should match the results of the age-relevant condition (i.e., longer reading times for incongruent than congruent pairs), and differ from the results of the gender-relevant condition. Indeed, these were the results of Study 5. These results support the hypothesis that participants inhibited the age dimension when the age of the target faces was test-irrelevant. Taken together, the current set of studies provides initial support for the application of the distinction of reactive versus proactive control framework to multiple categorization and social cognition in general.[4]

Implications for Multiple Categorization Research

The present research was guided by the general principle that the complexities of multicategorical targets faced by social perceivers are dealt by them in real-time, in an incremental and trial-and-error manner that leads to adaptation and, thus, facilitated subsequent performance. Moreover, the success in overcoming these complexities greatly depends on the specific requirements of the relevant social tasks and of the nature of the social context, allowing for more or less learning. Overcoming the complexities of multicategorical information is, therefore, a matter of cognitive control. Thus, we argue that dynamic learning and cognitive control are powerful determinants of temporary category dominance.

Although previous research has provided great insights into the complexities of multiple categorization of social actors, we contend that it also has greatly overlooked dynamic learning and cognitive control. The current research aimed to overcome this relative neglect and, therefore, it diverges from previous research in several important ways. First, by including successive cycles of learning and test, our experimental paradigm reproduces the dynamic nature of social reality, allowing the nature of the test to inform subsequent encoding processes. Second, the cycles of

[4] In a similar vein, recent research by Amodio and Swencionis (2018) has shown the effect of proactive control on implicit racial bias. To do so, the authors manipulated response interference as a means to induce proactive control. Namely, they varied the proportion of congruent and incongruent prime-target pairings within different blocks of trials in the Weapons Identification Task (Payne, 2001). Low-interference blocks consisted of more stereotype congruent (Black– gun and White–tool) than incongruent trials (Black–tool and White– gun), while high-interference blocks consisted of more incongruent than congruent trials. The idea was that high interference blocks, because of their experienced difficulty, should engage a proactive mode of control that enhances focus on the primary task of identifying the object (gun or tool), thereby reducing racial bias. The results were congruent with this prediction.

learning and test provide sequential data that illustrate learning processes and the gradual modifications of control processes. Third, by changing the requirements of the final test, our approach is also able to document the costs of previous learning when test requirements change. The present paradigm is, thus, meant to offer evidence of the cognitive mastering of specific tasks involving social knowledge and its associated costs in performance on different subsequent tasks with requirements.

A handful of studies on social categorization have examined the interplay between age and gender categorizations from faces, and the results are mixed. Specifically, there is evidence showing an asymmetrical relationship between the two dimensions. On the one hand, in some tasks, the nonrelevant age dimension interferes with the processing of the task relevant gender dimension. For example, using a Garner interference task, Quinn and Macrae (2005) found that when the goal was to categorize the target faces by gender, participants' performance was worse when the age of the targets also varied relative to when it remained constant. In contrast, when the goal was to categorize the faces according to their age, it did not matter if the gender of faces was kept constant or not (for similar findings, see Wiese, Schweinberger, & Neumann, 2008).

On the other hand, a recent study by Cloutier and colleagues (2014) showed precisely the opposite relationship between age and gender. Namely, when participants were asked to make age categorizations of faces for younger and older, male and female, individuals, their hand movements were attracted toward the applicable nonrelevant gender category option. However, when the goal was to categorize the faces according to their gender, the authors did not find any evidence of attraction to the applicable nonfocal age category.

There is also evidence for a symmetrical relationship between age and gender. Namely, D. Martin and colleagues (2015) found that participants were slower and less accurate when, in two consecutive trials, the task-relevant dimension was repeated (e.g., younger face–younger face) and the nonrelevant dimension switched (e.g., male–female) relative to when the nonrelevant dimension also repeated (e.g., male–male). Conversely, when the task relevant dimension switched (e.g., younger face–older face), their performance was more hindered when the nonrelevant dimension repeated (e.g., male–male) than when it switched (e.g., male–female). More important, this pattern was obtained both when age and gender were relevant to the task. Thus, these results suggest that age and gender mutually influence one another.

Finally, a study by Klauer et al. (2003) using the "Who Said What?" paradigm obtained some evidence showing that age and gender can dominate social perception independently of one another. Namely, these authors manipulated the relevance of age and gender by using discussion topics stereotypical of each category dimension. In the three studies in which the age of the speakers was made relevant via discussion topic, participants showed higher recall for the age of the speakers than for its gender in two of these studies. In the two studies in which gender was made relevant, participants' recall performance was always better for the gender of the speakers.

The results of the present studies showed an asymmetrical relationship in the encoding of gender and age information, with gender prevailing over age information and, thus, they seem consistent with the findings of Cloutier and colleagues (2014).

Relations to Existing Models of Selective Attention and Category Learning

Our approach to multiple categorizations shares some ideas with existing models of selective attention and category learning, but it also diverges from these in several key aspects. Namely, research on selective attention has consistently shown that the selection of specific stimulus dimensions can be driven by the low-level perceptual characteristics of a stimulus (e.g., the salience or distinctiveness of its features) or by our expectations and goals for the task at hand (e.g., Folk & Remington, 1998; Ruz & Lupianez, 2002; Theeuwes, 2010; Wolfe, 1994; Yantis, 2002). In line with this, we argue that when one is exposed for the first time to a task in which one is asked to judge an unfamiliar individual in some dimension, category selection might result from the interaction of both bottom-up and top-down attentional processes (for a similar idea, see Freeman & Ambady, 2011). However, when given the opportunity to learn the structure and requirements of the task, category selection is mainly determined by task-relevance. That is, we propose that bottom-up selective attention is gradually replaced and constrained by top-down selective attention as a function of the learning accumulated by the experience with previous tasks.

Regarding category learning models, there are also similarities between our approach and some of these models', particularly to those that attribute a central role to selective attention in preserving and facilitating learning (e.g., Kruschke, 2001). However, there are many relevant differences to these models' proposals. For example, instead of repeating pairs of stimuli (e.g., cues and outcomes) multiple times and then give feedback on whether participants correctly learned the association (for a review, see Kruschke, 2011), we always present new stimuli in each cycle while maintaining the test structure constant. Thus, our focus is on whether participants can learn the relevance of some dimensions based on the structure of tests and not on reinforcing accurate performance per se. Learning dimension relevancy must be a cognitive precursor of accurate performance, but no associative learning model that we know of deal with both processes. However, of course, it is easy to think of extensions of current associative learning models to encompass relevance and accuracy learning. Another essential feature of our research is that we used highly meaningful stimuli (age and gender categories) about which people possess a rich network of previous knowledge (e.g., stereotypes). This feature is nonnegligible as previous knowledge can bias information processing (e.g., Macrae & Bodenhausen, 2000) and, thus, facilitate or interfere with manipulations aimed at activating or inhibiting category encoding.

Possible Reasons for the Dominance of Gender Categorizations

How to explain the prevalence of gender categorizations observed in Studies 1–3? There might be multiple explanations for this finding. Pinpointing exactly what (and how) is facilitating the encoding of gender information is beyond the scope of the present work. Below we discuss the possibilities that we believe to be the most relevant ones and are hopeful future research will clarify which may best explain the current findings.

One possible explanation is related with the diagnosticity of the facial features associated with each category dimension. Namely, one of the most diagnostic features to discriminate between men and women is the person's hairstyle (Brown & Perrett, 1993; Burton, Bruce, & Dench, 1993). For example, it has been shown that the hair alone is sufficient to trigger category and stereotype activation (Freeman, Ambady, Rule, & Johnson, 2008; Macrae & Martin, 2007). An interesting find, among the many facial features that support the distinction between young and older people is also the hair, particularly the color and quantity of hair (Berry & McArthur, 1986). Thus, although the hair is a diagnostic feature for both dimensions, the style (i.e., long or short) is more critical to gender while the color is more critical to age. This aspect might be of extreme relevance for our results as we used grayscale photographs and, therefore, unintentionally reduced the relative importance of the hair color for age categorizations. In fact, this aspect may have also played a role in other studies (e.g., Cloutier et al., 2014).

A second potential explanation is language. Language shapes our cognitive understanding of the world around us (Boroditsky, 2001; Boroditsky, Schmidt, & Phillips, 2003; Whorf, 1956). Specifically, the grammatical gender of a language can affect cognition such that differences across languages lead to differences in cognition (e.g., Boroditsky et al., 2003; Konishi, 1993; Sera et al., 2002; Vigliocco, Vinson, Paganelli, & Dworzynski, 2005). For example, Boroditsky and colleagues (2003) discovered that Spanish native-speakers described a "key" (a feminine marked noun) with feminine characteristics (e.g., golden, little) whereas German speakers described key (a masculine marked noun in the German language) with masculine characteristics (e.g., hard, heavy). Thus, the fact that the Portuguese language is a gendered language (i.e., most nouns and adjectives are assigned to masculine or feminine gender), might have contributed to maintain the gender dimension active in the age-relevant condition, even though we used gender-neutral labels (i.e., young and senior). Despite the existing studies showing an influence of language and cognition, there is also research questioning the pervasiveness of this influence (e.g., Bender, Beller, & Klauer, 2018; Ramos & Roberson, 2011; Sera et al., 2002). For example, research by Ramos and Roberson (2011) showed that the grammatical gender of object nouns biases the categorization judgments of Portuguese speakers, but only when the task instruction makes explicit reference to gender. Removing explicit references to gender eliminates its impact on categorization judgments.

Thus, if the absence of hair color in the photographs and the fact that the Portuguese language is gendered are contributing (separately or jointly) to the encoding of gender information irrespective of its test-relevance, then manipulating the hair color (brown vs. white) and/or replicating these studies with native speakers of a less gendered language (e.g., English speakers) should make the encoding of gender dependent of its test-relevance, just like we observed for age.

Another possibility is that despite both category dimensions are fundamental for person perception, gender is a more fundamental dimension as it is learned (and overlearned) very early in life. Infants as young as 3 to 4 months of age differentiate faces based on gender (Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002) and by 12 months they are able to integrate visual and auditory gendered information about a speaker (e.g., Poulin-Dubois, Serbin, Kenyon, & Derbyshire, 1994). Moreover, multiple developmental studies show that preschool-aged children spontaneously use gender information to guide their social inferences and preferences about other individuals (e.g., Dunham, Baron, & Banaji, 2016; Maccoby, 2002; C. L. Martin & Ruble, 2004; Shutts, Banaji, & Spelke, 2010; Yee & Brown, 1994). For example, preschoolers assign more positive than negative traits to children of their own gender (e.g., Albert & Porter, 1983); when presented with information about unfamiliar children, they say that they would prefer to be friends with other children of their own gender (e.g., C. L. Martin, 1989; C. L. Martin et al., 1999); and they use gender as a cue to remember faces in the "Who said what?" paradigm (Weisman, Johnson, & Shutts, 2015). Developmental research suggests that increasing the perceptual discriminability between boys and girls by means of gender-stereotypic markers (e.g., hairstyles, clothing; Wild et al., 2000) or exposing children to environments that are characterized by gender-based sorting (e.g., boys and girls sports' teams and bathrooms; Bigler & Liben, 2006) are examples of social practices that causally contribute to the early salience of binary gender categorizations and to the development of corresponding gender stereotypes and prejudices (for a recent discussion on this see Hyde, Bigler, Joel, Tate, & van Anders, 2019).

Research on evolutionary psychology also suggests that gender might be more crucial than age. According to the evolutionary approach, our ancestors inhabited a social world in which encoding the gender of an individual would have enabled a large variety of useful inferences and decisions such as, for example, quickly decide who was a potential mate, friend, or rival. Thus, natural selection favored our cognitive machinery to encode gender in an automatic and mandatory fashion (e.g., Kurzban, Tooby, & Cosmides, 2001; Pietraszewski et al., 2014). In line with this reasoning, it has been shown that gender categorizations are not disrupted by factors that typically erase race categorizations, as, for example, the presence of cues that increased the relevance of a cross-cutting coalitional category (e.g., two rival basketball teams; Kurzban et al., 2001, Experiments 3 and 4). Although we recognize the plausibility of this explanation, it is hard to reconcile it with existing findings (described above) showing that gender categorizations decrease when gender is not task-relevant (Quinn & Macrae, 2005; Wiese et al., 2008). Additionally, recent research shows that the findings obtained by Kurzban et al. (2001; Studies 3 and 4) might have been caused by a confound related with the particular coalitional category used (basketball teams), such that when this issue is removed, gender categorizations are shown to be greatly reduced in the presence of situational cues that decrease their relevance (Klauer, Hölzenbein, Calanchini, & Sherman, 2014; Study 5). Thus, based on this research, we have reasons to believe that gender categorization might be more malleable than assumed by the evolutionary approach.

Finally, yet another possibility is that gender categorizations are easier because they typically involve only two alternatives (man or woman) while age categorizations are more complex as they involve a wide range of options. To minimize for the eventual impact of this issue, we selected our stimulus faces such that young and senior faces would be easily distinguishable regarding their age category and our task required participants to respond whether the statements were said by a junior or senior person. Thus, although we do not think this issue played a particular role in our studies, the "gender binary" may facilitate the everyday encoding of gender information, in contrast to age information.

If the pervasiveness of gender categorizations is because of its evolutionary significance and/or to the gender binary, then manipulations like the ones outlined above (or others) should have little or no impact on gender encoding, which should always occur irrespective of its test-relevance.

Conclusion

To date, most findings on the relation between multiple category dimensions were obtained under conditions that restricted participants' ability to learn the task requirements and, thus, to proactively control their processing strategies. In the current research, we proposed that, when given the opportunity, people are capable of learning the relevance of each category dimension and controlling their subsequent encoding strategies accordingly. The results of five studies supported our hypothesis such that they showed that the encoding of age information is dependent of task-relevance, while the encoding of gender information seems relatively independent of task-relevance. Taken together, the current findings provide initial evidence that control processes shape that category dimension will momentarily dominate person perception.

References

Albert, A. A., & Porter, J. R. (1983). Age patterns in the development of children's gender-role stereotypes. *Sex Roles, 9,* 59–67. http://dx.doi.org/10.1007/BF00303110

Allport, G. W. (1954). *The nature of prejudice*. New York, NY: Basic books.

Amodio, D. M., Bartholow, B. D., & Ito, T. A. (2014). Tracking the dynamics of the social brain: ERP approaches for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience, 9,* 385–393. http://dx.doi.org/10.1093/scan/nst177

Amodio, D. M., & Swencionis, J. K. (2018). Proactive control of implicit bias: A theoretical model and implications for behavior change. *Journal of Personality and Social Psychology, 115,* 255–275. http://dx.doi.org/10.1037/pspi0000128

Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long-term memory: A new taxonomy. In D. Dagenbach & T. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 265–325). New York: Academic Press.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063–1087. http://dx.doi.org/10.1037/0278-7393.20.5.1063

Anderson, M. C., & Levy, B. J. (2007). Theoretical issues in inhibition: Insights from research on human memory. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 81–102). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/11587-005

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102,* 68–100. http://dx.doi.org/10.1037/0033-295X.102.1.68

Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 940–945. http://dx.doi.org/10.1037/a0029199

Bargh, J. A., & Thein, R. D. (1985). Individual construct accessibility, person memory, and the recall-judgment link: The case of information overload. *Journal of Personality and Social Psychology, 49,* 1129–1146. http://dx.doi.org/10.1037/0022-3514.49.5.1129

Belmore, S. M. (1987). Determinants of attention during impression formation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 480–489. http://dx.doi.org/10.1037/0278-7393.13.3.480

Bender, A., Beller, S., & Klauer, K. C. (2018). Gender congruency from a neutral point of view: The roles of gender classes and conceptual connotations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 1580–1608. http://dx.doi.org/10.1037/xlm0000534

Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin, 100,* 3–18. http://dx.doi.org/10.1037/0033-2909.100.1.3

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. *Advances in Child Development and Behavior, 34,* 39–89. http://dx.doi.org/10.1016/S0065-2407(06)80004-2

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15,* 674–679. http://dx.doi.org/10.1111/j.0956-7976.2004.00739.x

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

Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. *Advances in Social Cognition, 11,* 1–52.

Bodenhausen, G. V., & Peery, D. (2009). Social categorization and stereotyping in vivo: The VUCA challenge. *Social and Personality Psychology Compass, 3,* 133–151. http://dx.doi.org/10.1111/j.1751-9004.2009.00167.x

Bodenhausen, G. V., Todd, A. R., & Becker, A. P. (2007). Categorizing the social world: Affect, motivation, and self-regulation. In B. H. Ross & A. B. Markman (Eds.), *Psychology of learning and motivation* (Vol. 47, pp. 123–155). Amsterdam, the Netherlands: Elsevier.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43,* 1–22. http://dx.doi.org/10.1006/cogp.2001.0748

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and cognition* (pp. 61–79). Cambridge, MA: MIT.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences, 16,* 106–113. http://dx.doi.org/10.1016/j.tics.2011.12.010

Brebner, J. L., Martin, D., & Macrae, C. N. (2009). Dude looks like a lady: Exploring the malleability of person categorization. *European Journal of Social Psychology, 39,* 1109–1119. http://dx.doi.org/10.1002/ejsp.600

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *A dual process model of impression formation* (pp. 1–36). Hillsdale, NJ: Erlbaum.

Brewer, M. B., Dull, V., & Lui, L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality and Social Psychology, 41,* 656–670. http://dx.doi.org/10.1037/0022-3514.41.4.656

Brewer, M. B., Ho, H. K., Lee, J. Y., & Miller, N. (1987). Social identity and social distance among Hong Kong school children. *Personality and Social Psychology Bulletin, 13,* 156–165. http://dx.doi.org/10.1177/0146167287132002

Brown, E., & Perrett, D. I. (1993). What gives a face its gender? *Perception, 22,* 829–840. http://dx.doi.org/10.1068/p220829

Burt, D. M., & Perrett, D. I. (1995). Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society of London, Series B: Biological Sciences, 259,* 137–143. http://dx.doi.org/10.1098/rspb.1995.0021

Burton, A. M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception, 22,* 153–176. http://dx.doi.org/10.1068/p220153

Carpinella, C. M., Chen, J. M., Hamilton, D. L., & Johnson, K. L. (2015). Gendered facial cues influence race categorizations. *Personality and Social Psychology Bulletin, 41,* 405–419. http://dx.doi.org/10.1177/0146167214567153

Cho, K. W., Neely, J. H., Crocco, S., & Vitrano, D. (2017). Testing enhances both encoding and retrieval for both tested and untested items. *The Quarterly Journal of Experimental Psychology, 70,* 1211–1235.

Cloutier, J., Freeman, J. B., & Ambady, N. (2014). Investigating the early stages of person perception: The asymmetry of social categorization by sex vs. age. *PLoS ONE, 9,* e84677.

Cloutier, J., & Macrae, C. N. (2007). Who or what are you? Facial orientation and person construal. *European Journal of Social Psychology, 37,* 1298–1309. http://dx.doi.org/10.1002/ejsp.433

Cloutier, J., Mason, M. F., & Macrae, C. N. (2005). The perceptual determinants of person construal: Reopening the social-cognitive toolbox. *Journal of Personality and Social Psychology, 88,* 885–894. http://dx.doi.org/10.1037/0022-3514.88.6.885

Crisp, R. J., & Hewstone, M. (2007). Multiple social categorization. *Advances in Experimental Social Psychology, 39,* 163–254. http://dx.doi.org/10.1016/S0065-2601(06)39004-1

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18. http://dx.doi.org/10.1037/0022-3514.56.1.5

Dunham, Y., Baron, A. S., & Banaji, M. R. (2016). The development of implicit gender attitudes. *Developmental Science, 19,* 781–789. http://dx.doi.org/10.1111/desc.12321

Ebner, N. C., He, Y., Fichtenholtz, H. M., McCarthy, G., & Johnson, M. K. (2011). Electrophysiological correlates of processing faces of younger and older individuals. *Social Cognitive and Affective Neuroscience, 6,* 526–535. http://dx.doi.org/10.1093/scan/nsq074

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23,* 1–74. http://dx.doi.org/10.1016/S0065-2601(08)60317-2

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 847–858. http://dx.doi.org/10.1037/0096-1523.24.3.847

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118,* 247–279. http://dx.doi.org/10.1037/a0022327

Freeman, J. B., Ambady, N., & Holcomb, P. J. (2010). The face-sensitive N170 encodes social category information. *NeuroReport, 21,* 24–28. http://dx.doi.org/10.1097/WNR.0b013e3283320d54

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General, 137,* 673–690. http://dx.doi.org/10.1037/a0013875

Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences, 20,* 362–374. http://dx.doi.org/10.1016/j.tics.2016.03.003

Freeman, J. B., Nakayama, K., & Ambady, N. (2013). Finger in flight reveals parallel categorization across multiple social dimensions. *Social Cognition, 31,* 792–805. http://dx.doi.org/10.1521/soco.2013.31.6.792

Garcia-Marques, L., Nunes, L. D., Marques, P., Carneiro, P., & Weinstein, Y. (2015). Adapting to test structure: Letting testing teach what to learn. *Memory, 23,* 365–380. http://dx.doi.org/10.1080/09658211.2014.889166

George, P. A., & Hole, G. J. (1995). Factors influencing the accuracy of age estimates of unfamiliar faces. *Perception, 24,* 1059–1073.

Hasher, L. (2007). Inhibition: Attentional regulation in cognition. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 291–294). New York, NY: Oxford University Press.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). New York, NY: Academic Press.

Hess, U., Adams, R. B., Jr., & Kleck, R. E. (2004). Facial appearance, gender, and emotion expression. *Emotion, 4,* 378–388. http://dx.doi.org/10.1037/1528-3542.4.4.378

Hewstone, M., Islam, M. R., & Judd, C. M. (1993). Models of crossed categorization and intergroup relations. *Journal of Personality and Social Psychology, 64,* 779–793. http://dx.doi.org/10.1037/0022-3514.64.5.779

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist, 74,* 171–193. http://dx.doi.org/10.1037/amp0000307

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85,* 616–626. http://dx.doi.org/10.1037/0022-3514.85.4.616

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology, 102,* 116–131. http://dx.doi.org/10.1037/a0025335

Johnson, K. L., Lick, D. J., & Carpinella, C. M. (2015). Emergent research in social vision: An integrated approach to the determinants and consequences of social categorization. *Social and Personality Psychology Compass, 9,* 15–30. http://dx.doi.org/10.1111/spc3.12147

Johnson, K. L., & Tassinary, L. G. (2005). Perceiving sex directly and indirectly: Meaning in motion and morphology. *Psychological Science, 16,* 890–897. http://dx.doi.org/10.1111/j.1467-9280.2005.01633.x

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57,* 151–162.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136,* 849–874. http://dx.doi.org/10.1037/a0019842

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,* 332–354. http://dx.doi.org/10.1016/S0022-1031(03)00017-9

Klauer, K. C., Hölzenbein, F., Calanchini, J., & Sherman, J. W. (2014). How malleable is categorization by race? Evidence for competitive category use in social categorization. *Journal of Personality and Social Psychology, 107,* 21–40. http://dx.doi.org/10.1037/a0036609

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychology, 75,* 1155–1178. http://dx.doi.org/10.1037/0022-3514.75.5.1155

Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of Psycholinguistic Research, 22,* 519–534. http://dx.doi.org/10.1007/BF01068252

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology, 45,* 812–863. http://dx.doi.org/10.1006/jmps.2000.1354

Kruschke, J. K. (2011). Models of attentional learning. In E. M. Pothos & A. J. Wills (Eds.), *Formal models in categorization* (pp. 120–152). England: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511921322.006

Kubota, J. T., & Ito, T. A. (2007). Multiple cues in social perception: The time course of processing race and facial expression. *Journal of Experimental Social Psychology, 43,* 738–752. http://dx.doi.org/10.1016/j.jesp.2006.10.023

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences, 98,* 15387–15392.

Maccoby, E. E. (2002). Gender and group process: A developmental perspective. *Current Directions in Psychological Science, 11,* 54–58. http://dx.doi.org/10.1111/1467-8721.00167

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law, 7,* 98–118. http://dx.doi.org/10.1037/1076-8971.7.1.98

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51,* 93–120. http://dx.doi.org/10.1146/annurev.psych.51.1.93

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69,* 397–407. http://dx.doi.org/10.1037/0022-3514.69.3.397

Macrae, C. N., & Martin, D. (2007). A boy primed Sue: Feature-based processing and person construal. *European Journal of Social Psychology, 37,* 793–805. http://dx.doi.org/10.1002/ejsp.406

Macrae, C. N., Quinn, K. A., Mason, M. F., & Quadflieg, S. (2005). Understanding others: The face and person construal. *Journal of Personality and Social Psychology, 89,* 686–695. http://dx.doi.org/10.1037/0022-3514.89.5.686

Martin, C. L. (1989). Children's use of gender-related information in making social judgments. *Developmental Psychology, 25,* 80–88. http://dx.doi.org/10.1037/0012-1649.25.1.80

Martin, C. L., Fabes, R. A., Evans, S. M., & Wyman, H. (1999). Social cognition on the playground: Children's beliefs about playing with girls versus boys and their relations to sex segregated play. *Journal of Social and Personal Relationships, 16,* 751–771. http://dx.doi.org/10.1177/0265407599166005

Martin, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science, 13,* 67–70. http://dx.doi.org/10.1111/j.0963-7214.2004.00276.x

Martin, D., & Macrae, C. N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology, 37,* 806–816. http://dx.doi.org/10.1002/ejsp.445

Martin, D., Swainson, R., Slessor, G., Hutchison, J., Marosi, D., & Cunningham, S. J. (2015). The simultaneous extraction of multiple social categories from unfamiliar faces. *Journal of Experimental Social Psychology, 60,* 51–58. http://dx.doi.org/10.1016/j.jesp.2015.03.009

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments & Computers, 36,* 630–633. http://dx.doi.org/10.3758/BF03206543

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7,* 134–140. http://dx.doi.org/10.1016/S1364-6613(03)00028-7

Mouchetant-Rostaing, Y., & Giard, M. H. (2003). Electrophysiological correlates of age and gender perception on human faces. *Journal of Cognitive Neuroscience, 15,* 900–910. http://dx.doi.org/10.1162/089892903322370816

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1200–1210. http://dx.doi.org/10.1037/0278-7393.24.5.1200

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81,* 181–192. http://dx.doi.org/10.1037/0022-3514.81.2.181

Pérez, A., Joseph, H. S., Bajo, T., & Nation, K. (2016). Evaluation and revision of inferential comprehension in narrative texts: An eye movement study. *Language, Cognition and Neuroscience, 31,* 549–566. http://dx.doi.org/10.1080/23273798.2015.1115883

Pietraszewski, D., Cosmides, L., & Tooby, J. (2014). The content of our cooperation, not the color of our skin: An alliance detection system regulates categorization by coalition and race, but not sex. *PLoS ONE, 9,* e88534. http://dx.doi.org/10.1371/journal.pone.0088534

Poulin-Dubois, D., Serbin, L. A., Kenyon, B., & Derbyshire, A. (1994). Infants' intermodal knowledge about gender. *Developmental Psychology, 30,* 436–442. http://dx.doi.org/10.1037/0012-1649.30.3.436

Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles, 59,* 377–391. http://dx.doi.org/10.1007/s11199-008-9424-4

Purmann, S., Badde, S., & Wendt, M. (2009). Adjustments to recent and frequent conflict reflect two distinguishable mechanisms. *Psychonomic Bulletin & Review, 16,* 350–355. http://dx.doi.org/10.3758/PBR.16.2.350

Quinn, K. A., & Macrae, C. N. (2005). Categorizing others: The dynamics of person construal. *Journal of Personality and Social Psychology, 88,* 467–479. http://dx.doi.org/10.1037/0022-3514.88.3.467

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception, 31,* 1109–1121. http://dx.doi.org/10.1068/p3331

Ramos, S., & Roberson, D. (2011). What constrains grammatical gender effects on semantic judgements? Evidence from Portuguese. *Journal of Cognitive Psychology, 23,* 102–111. http://dx.doi.org/10.1080/20445911.2011.466795

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114,* 510–532. http://dx.doi.org/10.1037/0033-2909.114.3.510

Remedios, J. D., Chasteen, A. L., Rule, N. O., & Plaks, J. E. (2011). Impressions at the intersection of ambiguous and obvious social categories: Does gay+ Black= likable? *Journal of Experimental Social Psychology, 47,* 1312–1315. http://dx.doi.org/10.1016/j.jesp.2011.05.015

Ridderinkhof, K. R. (2002). Micro- and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research, 66,* 312–323. http://dx.doi.org/10.1007/s00426-002-0104-7

Roccas, S., & Brewer, M. B. (2002). Social identity complexity. *Personality and Social Psychology Review, 6,* 88–106. http://dx.doi.org/10.1207/S15327957PSPR0602_01

Ruz, M., & Lupianez, J. (2002). A review of attentional capture: On its automaticity and sensitivity to endogenous control. *Psicológica, 23,* 283–309.

Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodríguez, W., & Dubois, D. P. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General, 131,* 377–397. http://dx.doi.org/10.1037/0096-3445.131.3.377

Shutts, K., Banaji, M. R., & Spelke, E. S. (2010). Social categories guide young children's preferences for novel objects. *Developmental Science, 13,* 599–610. http://dx.doi.org/10.1111/j.1467-7687.2009.00913.x

Sinclair, L., & Kunda, Z. (1999). Reactions to a black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77,* 885–904. http://dx.doi.org/10.1037/0022-3514.77.5.885

Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin, 26,* 1329–1342. http://dx.doi.org/10.1177/0146167200263002

Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review, 99,* 3–21. http://dx.doi.org/10.1037/0033-295X.99.1.3

Stangor, C., Lynch, L., Duan, C., & Glas, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Person-*

ality and Social Psychology, 62, 207–218. http://dx.doi.org/10.1037/0022-3514.62.2.207

Stern, L. D., Marrs, S., Millar, M. G., & Cole, E. (1984). Processing time and the recall of inconsistent and consistent behaviors of individuals and groups. *Journal of Personality and Social Psychology, 47,* 253–262. http://dx.doi.org/10.1037/0022-3514.47.2.253

Stroessner, S. J. (1996). Social categorization by race or sex: Effects of perceived non-normalcy on response times. *Social Cognition, 14,* 247–276. http://dx.doi.org/10.1521/soco.1996.14.3.247

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago, IL: Nelson-Hall.

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

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica, 135,* 77–99. http://dx.doi.org/10.1016/j.actpsy.2010.02.006

van Twuyver, M., & van Knippenberg, A. (1995). Social categorization as a function of priming. *European Journal of Social Psychology, 25,* 695–701. http://dx.doi.org/10.1002/ejsp.2420250608

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology, 134,* 501–520. http://dx.doi.org/10.1037/0096-3445.134.4.501

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review, 114,* 830–841.

Weisman, K., Johnson, M. V., & Shutts, K. (2015). Young children's automatic encoding of social categories. *Developmental Science, 18,* 1036–1043. http://dx.doi.org/10.1111/desc.12269

Whorf, B. L. (1956). In J. B. Carroll (Ed.), *Language, thought, and reality: Selected writings of Benjamin Lee Whorf.* Cambridge, MA: MIT Press.

Wiese, H., Schweinberger, S. R., & Neumann, M. (2008). Perceiving age and gender in unfamiliar faces: Brain potential evidence for implicit and explicit person categorization. *Psychophysiology, 45,* 957–969.

Wild, H. A., Barrett, S. E., Spence, M. J., O'Toole, A. J., Cheng, Y. D., & Brooke, J. (2000). Recognition and sex categorization of adults' and children's faces: Examining performance in the absence of sex-stereotyped cues. *Journal of Experimental Child Psychology, 77,* 269–291. http://dx.doi.org/10.1006/jecp.1999.2554

Wolfe, J. M. (1994). Guided Search 2.0. A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238. http://dx.doi.org/10.3758/BF03200774

Yantis, S. (2002). Stimulus-driven and goal-directed attentional control. In V. Cantoni, M. Marinaro, & A. Petrosino (Eds.), *Visual attention mechanisms* (pp. 125–134). Boston, MA: Springer.

Yee, M., & Brown, R. (1994). The development of gender differentiation in young children. *British Journal of Social Psychology, 33,* 183–196. http://dx.doi.org/10.1111/j.2044-8309.1994.tb01017.x

Zárate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition, 8,* 161–185. http://dx.doi.org/10.1521/soco.1990.8.2.161

Ziegler, R., & Burger, A. M. (2011). Mood and the impact of individuating information on the evaluation of ingroup and outgroup members: The role of mood-based expectancies. *Journal of Experimental Social Psychology, 47,* 1000–1006. http://dx.doi.org/10.1016/j.jesp.2011.04.002

Received June 4, 2018
Revision received February 28, 2019
Accepted March 11, 2019 ■