

Appearance-based inferences bias source memory

Brittany S. Cassidy · Leslie A. Zebrowitz ·
Angela H. Gutchess

Published online: 18 July 2012
© Psychonomic Society, Inc. 2012

Abstract Previous research varying the trustworthiness of appearance has demonstrated that facial characteristics contribute to source memory. Two studies extended this work by investigating the contribution to source memory of babyfacedness, a facial quality known to elicit strong spontaneous trait inferences. Young adult participants viewed younger and older babyfaced and mature-faced individuals paired with sentences that were either congruent or incongruent with the target's facial characteristics. Identifying a source as dominant or submissive was least accurate when participants chose between a target whose behavior was incongruent with facial characteristics and a lure whose face mismatched the target in appearance but matched the source memory question. In Experiment 1, this effect held true when older sources were identified, but not own-age, younger sources. When task difficulty was increased in Experiment 2, the relationship between face-behavior congruence and lure facial characteristics persisted, but it was not moderated by target age even though participants continued to correctly identify fewer older than younger sources. Taken together, these results indicate that trait expectations associated with variations in facial maturity can bias source memory for both own- and other-age faces, although own-age faces are less vulnerable to this bias, as is shown in the moderation by task difficulty.

Keywords Babyfacedness · Source monitoring · Memory · Stereotypes

Parents often advise their children to “never judge a book by its cover.” Overcoming this tendency is difficult, since

people spontaneously rely on facial appearance when forming impressions of others. Appearance-based impressions occur in a seemingly instantaneous way (Willis & Todorov, 2006) and have important outcomes for the actors in question. For instance, competence judgments after brief exposures to faces reliably predict electoral success (Olivola & Todorov, 2010). This suggests that people agree upon initial appraisals of facial characteristics when forming impressions and that these inferences persist in memory in light of learned behavior about individuals. Appearance-based inferences reflect perceptual biases and overgeneralization of personality characteristics on the basis of facial qualities (Zebrowitz & Montepare, 2008). Although such biases may be useful in that they provide insight into potential social interactions, these overgeneralizations may not always accurately predict the personality characteristics of others. This can lead to inaccurate memories of others' behaviors and may have important consequences in everyday life (e.g., deciding who is safe to approach). The present experiments investigated how these perceptual biases may disrupt the ability to correctly attribute source information.

One of the strongest and most well-studied facial features that reliably influence person perception is *babyfacedness*, or the extent to which faces have childlike features (Montepare & Zebrowitz, 1998; Zebrowitz, 1997). People perceive others as being babyfaced when they have more neonatal features, such as bigger eyes and rounder faces (Berry & McArthur, 1986). Regardless of attractiveness, people typify babyfaced individuals as having more childlike traits than their mature-faced peers—particularly, lower dominance (Montepare & Zebrowitz, 1998). Aside from assuming that one possesses childlike traits because of appearance, these overgeneralized impressions critically influence many interpersonal outcomes, including hiring recommendations and job status (Collins & Zebrowitz, 1995; Zebrowitz, Tenenbaum, & Goldstein, 1991), credibility (Brownlow & Zebrowitz, 1990), and attribution of

B. S. Cassidy (✉) · L. A. Zebrowitz · A. H. Gutchess
Brandeis University,
MS 062,
Waltham, MA 02454, USA
e-mail: bcassidy@brandeis.edu

legal responsibility (Berry & Zebrowitz-McArthur, 1988; Zebrowitz & McDonald, 1991). In legal situations, babyfaced defendants are more likely to win cases involving intentional wrongdoing and more likely to lose cases involving negligent actions (Zebrowitz & McDonald, 1991). Parents also recommend harsher punishment for babyfaced children's transgressions, and they assign them less cognitively demanding tasks (Zebrowitz, Kendall-Tackett, & Fafel, 1991). Although babyface stereotypes have implications for the judgment and treatment of others, little is known about how these stereotypes affect memory. When behaviors conflict with impressions based on facial features, it may be difficult to integrate this information into memory. Such an effect would make facial stereotypes less responsive to corrective information, thereby perpetuating them. One way to address how facial appearance influences memory for impressions is through testing memory for sources whose faces are consistent or inconsistent with behaviors.

Source memory refers to the ability to remember the *context* in which information was presented, rather than *what* was presented alone, and source monitoring refers to the mechanisms involved in retrieving such contextual material (Johnson, Hashtroudi, & Lindsay, 1993). Although some source-monitoring errors appear harmless (e.g., misattributing who told you a funny joke), other errors have more sobering outcomes, as when a clever attorney causes an eyewitness to misidentify source information while giving testimony. People also use schematic knowledge when attributing sources (Bayen, Nakamura, Dupuis, & Yang, 2000; Hicks & Cockman, 2003; Marsh, Cook, & Hicks, 2006; Mather et al., 1999). This leads to better performance when people identify schema-consistent over inconsistent sources, which is based on the idea that when people do not remember a source, they identify it using schema-based expectations (Bayen et al., 2000). For instance, stereotypes about gender and sexual orientation bias source attribution, such that male-consistent statements are more likely to be attributed to male over female sources and vice versa, while the reverse is true if individuals are denoted as gay or lesbian (Marsh et al., 2006). Although the pervasive use of stereotypes in categorizing others (e.g., one might expect a mechanic to be handy, but not scholarly) is well studied (Fiske, Neuberg, Beattle, & Milberg, 1987; Hicks & Cockman, 2003), little is known about how appearance-based stereotypes bias memory about sources—specifically, which people have performed particular behaviors.

People use facial characteristics as cues when remembering sources. One recent study showed that people were more likely to remember plausible and positive newspaper headlines as coming from a source with trustworthy facial qualities and implausible and negative headlines from an untrustworthy-looking source. This occurred despite the fact that positive and negative, as well as plausible and

implausible headlines, had been generated equally by both sources (Nash, Bryer, & Schlagh, 2010). Although this study did not directly ask participants to form impressions, the automaticity of stereotyped impressions from facial cues biased source monitoring. Another recent study demonstrated that regardless of race or gender, the similarity between facial features and stereotyped categories also biases source memory, such that individuals with more stereotypically Black features are more likely to be misremembered as criminals than are nonstereotypical faces (Kleider, Cavrak, & Knuycky, 2012). Additionally, Suzuki and Suga (2010) found that after playing an economic game, participants identified more trustworthy- than untrustworthy-looking cheaters, suggesting that face-behavior incongruity increased encoding of impressions into memory, perhaps because memory for potentially misleading trustworthiness could protect people in certain situations.

The present study extends this previous research in several ways. First, we investigated how the babyface stereotype affects source memory for people whose behaviors are incongruent with appearance. After an encoding task in which people were shown face-behavior pairs that implied submissiveness or dominance, they were shown two faces (target and lure) and were asked which one had a particular trait, with the target always being the correct response. We expected congruity effects to bias source memory, such that individuals would make more source memory errors when attempting to identify *targets* whose behaviors were *incongruent* with their faces (e.g., a babyfaced person who performed a dominant behavior), rather than *congruent* (e.g., a babyfaced person who performed a submissive behavior). We also expected that this effect would be moderated by whether the *lures* had similar or differing facial characteristics. More specifically, we predicted that when target and lure faces differed in babyfaceness (e.g., babyfaced target, mature-faced lure) and the behavior originally encoded with the target was incongruent with facial stereotypes (e.g., dominant), participants would make more source memory errors when recalling the target as dominant or submissive than they would when the target's behavior was stereotype congruent. On the other hand, when the target and lure faces were similar in babyfaceness (e.g., both babyfaced), we expected face-behavior congruence to play a lesser role, since the appearance of lures would not provide a cue that could bias source recall.

An alternative hypothesis is provided by the Suzuki and Suga (2010) finding that incongruity improved person memory. This possibility suggests that we would find enhanced source memory when behaviors were incongruent rather than congruent with appearance. However, such an effect is more likely when behavioral information has high arousal value (Sharot & Phelps, 2004), as was true in the finding that people were more likely to recall being cheated by a

trustworthy-looking person. In contrast, when stereotypes are activated in less overt ways, as in the implicit activation in the present study, people remember more expectancy-congruent than incongruent information (Heider et al., 2007).

We also examined the strength of face–behavior congruity effects across in-group and out-group targets. One characteristic, actor age, influences memory such that people show better face recognition for same-age peers (Anastasi & Rhodes, 2006). Such own-age biases have implications for source memory, particularly in the legal system. Havard and Memon (2009) demonstrated that young adults tend to misidentify older adults when indentifying perpetrators in a lineup and that, while older adults perform more poorly overall, they do not show this bias. In the present study, we predicted that our younger participants' source memory would be worse for older than for younger faces, regardless of face–behavior congruence. Additionally, although we predicted overall decreased source memory for incongruent over congruent targets when lure facial characteristics differed from the targets, we also expected that congruity effects would be more apparent when identifying older versus younger sources. This prediction assumes that young adults' greater difficulty identifying older than younger sources may exacerbate any tendency to use the lure as a schema-consistent source decision cue in the case of incongruent older targets.

Finally, we investigated the strength of the congruity effects across two encoding contexts. Previous impression formation work has found that more person information is typically retrieved when information is encoded while using impression formation versus memorization goals (Chartrand & Bargh, 1996; Hamilton, Katz, & Leirer, 1980). Thus, the strength of congruity effects may be modulated by how individuals are instructed to orient to given person information. To test this, we manipulated two encoding contexts. One was a general impression formation encoding context, encouraging evaluation of individuals on the basis of their behaviors. The second was a memorization encoding context, encouraging the rote memorization of behavioral information for a later task. Using different encoding contexts helps to assess the reliability of congruity effects across different situations. Previous work assessing appearance-derived congruity effects (Nash et al., 2010; Suzuki & Suga, 2010) has not utilized these instructions, which are prevalent in the impression formation literature (Cassidy & Gutchess, 2012; Mitchell, Macrae, & Banaji, 2004). By comparing memory performance under impression formation and memorization goals, we may further explore the processes underlying appearance-based congruity effects on source memory. Consistent with suggestions that spontaneous impressions require few attentional resources (Todorov & Uleman, 2003), a broad impression formation goal might

encourage more associations and evaluations, thereby reducing congruity effects, relative to a narrower memorization goal. However, if these spontaneous associations elaborated on the babyface stereotype, impression formation instructions could exacerbate the effects of congruity, especially if a target performed a behavior incongruent with appearance, yielding more source errors than memorization instructions,

To summarize our predictions, we expected that appearance-based impressions would bias source memory, particularly when a target's behavior was incongruent with facial characteristics. We also predicted that the ability to identify sources would be worse for "out-group" targets that differed from participants in age. Finally, we expected that the context in which individuals encoded material would affect source memory, such that an impression formation, relative to a memorization, goal could either exacerbate or ameliorate the effects of congruity when sources were identified.

Experiment 1

Method

Participants

Forty-eight younger adults (18 to 22 years old, 14 males; $M = 19.13$, $SD = 1.06$) recruited from Brandeis University participated.

Stimuli

Faces Sixty-four pictures of faces (evenly distributed across younger/old and male/female) with neutral expressions, drawn from the PAL database (Minear & Park, 2004), were selected for the final data set on the basis of ratings of the babyfacedness and attractiveness of 105 faces by eight younger (M age = 20.75, $SD = 2.05$) and 13 older (M age = 78.85, $SD = 5.98$) adults who did not participate in the full experiment. Faces were rated on two 7-point scales: *extremely mature-faced* (1)–*extremely babyfaced* (7) and *extremely unattractive* (1)–*extremely attractive* (7). Raters viewed faces in blocks of set age–gender categories, and they were asked to judge faces relative to others of the same gender and age. Younger faces were between 18 and 29 years old, and older faces were between 70 and 94 years old. Faces selected for the final data set were rated at the most extreme ends of a 7-point scale (*extremely mature-faced* [1]–*extremely babyfaced* [7]). Because we ran the full experiment in blocks split by age–gender category, we compared the babyfacedness and attractiveness of the 64 selected faces within each category that had been designated as babyfaced and mature-faced on the basis of the preliminary ratings. As was

expected, there were significant differences in babyfacedness (Table 1a), but not attractiveness (Table 1b), within each age–gender group.

Sentences Sixty-four unique sentences (15 drawn from an often-cited database used for impression formation experiments [Uleman, 1988] and 49 sentences created in the lab), distributed equally among positive and negative valences, were selected for the final data set. Eight young adults and 8 of 13 older adults, who also rated the faces but did not complete the memory experiments, rated 140 sentences on a 7-point scale (*extremely submissive* [1]–*extremely dominant* [7]). We selected sentences for the final data set that were rated at the most extreme ends of the scale, with 32 rated as *dominant* and 32 rated as *submissive*.

We ran a 2 × 2 ANOVA using dominance (dominant, submissive) and valence (positive, negative) as factors to assess differences in the ratings of the final 64 sentences. Sentences selected for the dominance set ($M = 5.93, SD = 0.27$) were rated as more dominant than sentences selected for the submissive set ($M = 2.72, SD = 0.33$), $F(1, 60) = 1,905.40, p < .001, \eta_p^2 = .97$. Dominance ratings did not differ for positive ($M = 4.34, SD = 1.72$) and negative ($M = 4.31, SD = 1.59$) sentences, $F(1, 60) = 0.12, p = .73$. Although there was a significant interaction between dominance and valence, $F(1, 60) = 4.19, p = .05, \eta_p^2 = .07$, negative dominant behaviors ($M = 5.99, SD = 0.17$) and positive dominant behaviors ($M = 5.87, SD = 0.34$) did not differ significantly in rated dominance, $t(30) = 1.32, p = .19$, and neither did negative submissive behaviors ($M = 2.63, SD = 0.29$) and positive submissive behaviors ($M = 2.81, SD = 0.35$), $t(30) = 1.56, p = .13$.

Face–behavior pairs The sixty-four faces (evenly distributed across younger/older, male/female, and babyface/mature-

faced) were randomly paired with stereotype-congruent (e.g., babyfaced–submissive) or stereotype-incongruent (e.g., babyfaced–dominant) sentences (evenly distributed across positive/negative valence and dominance/submissiveness). There were eight task versions that counterbalanced the 64 babyfaced and mature-faced face–behavior pairs for stereotype congruence and sentence valence (16 stereotype-congruent face–behavior pairs for each face age and 16 stereotype-incongruent face–behavior pairs for each face age).

Encoding manipulation

Twenty-four participants were given impression formation instructions, and 24 were given memorization instructions. Participants given impression formation instructions were told, “You will be participating in a task about how people get to know others. This is a follow-up study to a previous impression formation experiment assessing reaction times and social cognition.” These participants were unaware of a future memory task. Participants given memorization instructions were told, “We are interested in learning about how accurate people are at assessing others. These pictures and sentences were taken out of a job database with biographies written about each person. Try and memorize your impressions of the people based on their behaviors, because later on, you will complete a task where you try and predict the jobs of the people in the current task.” While participants in the memorization group believed that they would see the individuals again, they were not aware of the true nature of the surprise memory task.

Procedure

Encoding task After providing informed consent, participants practiced the encoding task. Stimuli were presented via E-Prime (Psychology Software Tools, Pittsburgh, PA). To implicitly activate preexisting stereotypes based on facial characteristics, participants viewed each face alone for 2 s. Participants were told to press the number 1 after looking at the face. Then participants saw the same face paired with a behavioral sentence implying dominance (e.g., “When the compass broke, he led the group north”) or submissiveness (e.g., “He asked everyone which movie they wanted to see”) for 5 s (Fig. 1a). Participants were told to press the number 2 after reading the sentence. They were told that their buttonpresses would not abort the presentation of the face–behavior pairs and were instructed to quickly enter their responses. Buttonpresses served to maintain attentional vigilance during encoding and provided a check that participants processed both the faces and the sentences. They also provided credibility to the

Table 1 Comparison of mean babyfacedness and attractiveness ratings of faces classified as babyfaced or mature-faced split by age–gender group

	Babyfaced (<i>M, SD</i>)	Mature-faced (<i>M, SD</i>)	<i>t</i> statistic	<i>p</i> -value
A. Babyfacedness ratings				
Older male	3.89 (0.36)	2.17 (0.27)	10.85	< .001
Younger male	4.33 (0.31)	2.93 (0.41)	7.71	< .001
Older female	3.60 (0.16)	1.98 (0.24)	16.04	< .001
Younger female	4.41 (0.52)	2.48 (0.45)	7.92	< .001
B. Attractiveness ratings				
Older male	3.95 (0.53)	3.35 (0.99)	1.40	.14
Younger male	3.73 (0.94)	3.55 (0.71)	0.44	.71
Older female	3.89 (0.49)	3.16 (0.94)	1.94	.09
Younger female	4.08 (0.47)	3.56 (0.76)	1.64	.27

impression formation instructions that indicated that this task assessed reaction times.

Faces were presented in four fixed age–gender blocks of 16 trials each, with 6 s of fixation between each block. In the eight versions of the task, the order of the age–gender blocks was counterbalanced (i.e., one fourth of participants saw older female faces first, one fourth saw older male faces first, etc.). To improve performance, each face–behavior pair appeared twice, with the same behavioral sentence, once per run in a random order, for a total of two runs. After finishing the encoding task, participants completed a digit comparison measure (Hedden et al., 2002) to reduce recency effects.

Retrieval task Participants completed a self-paced retrieval task that assessed memory for face–behavior associations. Participants were told that they would be viewing all the faces they saw in the previous task, without presentation of any new faces, and that they should base their responses on their memory of the behaviors. The task instructions, read to all participants, were the following: “You will see a question below two people from the previous task. Based on the question, you will decide either which person is more

dominant or which person is more submissive. If you are not sure, go with your gut instinct. Decide based on your memory for their behaviors.” Faces were presented in one block, two at a time, in a random order. One face was the target, and the other a lure, matched in age and sex. Half of the lure faces also matched the target in facial characteristics (e.g., babyfaced–babyfaced), and half had facial characteristics mismatching the target (e.g., babyfaced–mature-faced). Lures with facial characteristics matching or differing from those of the targets were evenly distributed among younger and older stereotype-congruent and -incongruent targets. All lure faces had been seen previously at encoding. Each face was used once as a target and once as a lure during retrieval, but no two faces appeared together twice during the task. Below the faces, a question implicitly referenced the target’s encoding behavior by asking, “Who is more submissive?” or “Who is more dominant?” Participants were instructed to answer the question on the basis of what they had learned in the previous task. Targets performing dominant behaviors were always paired with the dominant question, and those performing submissive behaviors were always paired with the submissive question (Fig. 1b). Lure faces had been originally paired with an encoding behavior mismatching the question,

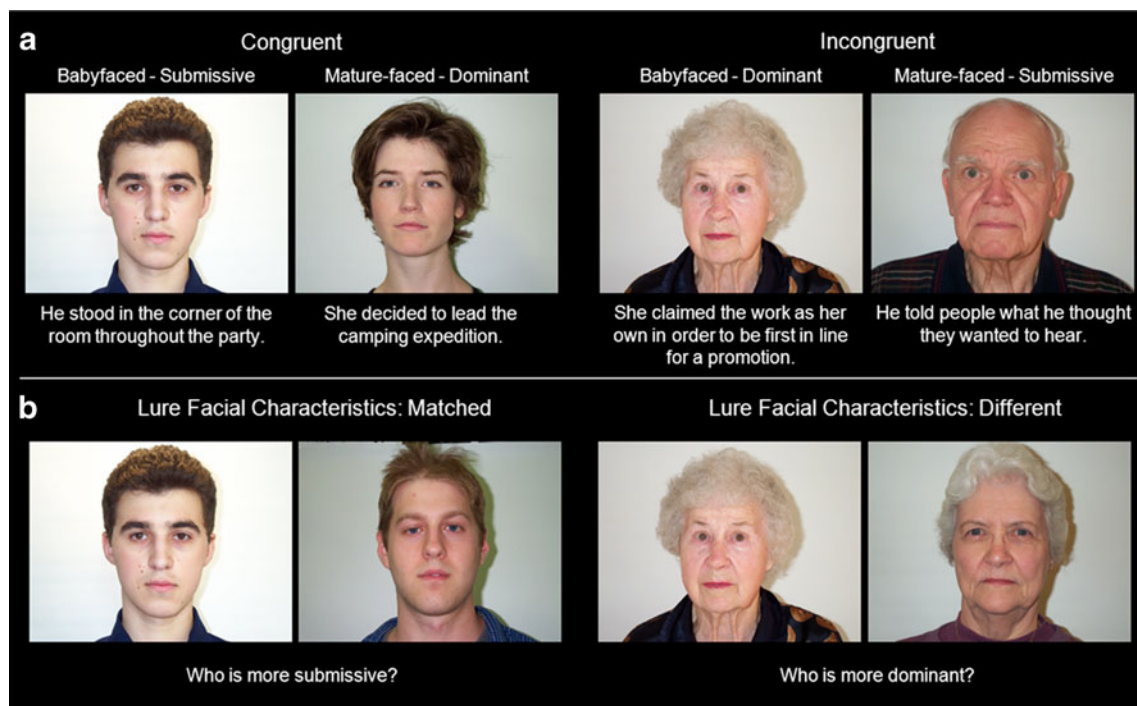


Fig. 1 **a** Example encoding stimuli: Participants saw each face alone, followed by the face paired with a behavioral sentence. Behaviors were congruent or incongruent with facial characteristics, but participants did not see this information. **b** Example retrieval stimuli: Participants viewed two faces from the previous task on the screen along with a prompt. All previously viewed and no new faces were viewed during

retrieval, and each face was used once as a target and once as a lure. Lure facial characteristics either matched or were different from the target on babyfacedness, but participants did not see this information. The target’s encoding behavior always matched the question, while the lure’s encoding behavior never matched the question

to ensure that the target was the correct response. Participants then completed additional cognitive measures.

Study design

Within the task, there was one between-group variable of interest: instructions. Half of the participants received memorization instructions, and half received impression formation instructions. There were three within-group variables of interest, all evenly counterbalanced among the face–behavior pairs. For face–behavior congruence, half of the faces were paired with behaviors congruent with appearance, while half were paired with behaviors incongruent with appearance. For lure facial characteristics, half of the lures had facial characteristics (e.g., babyfaced or mature-faced) differing from those of targets, while half had facial characteristics matching those of the targets. For age of face, half of the faces were older adult faces, and half were younger adult faces.

Results

Retrieval accuracy

We analyzed participants' accuracy (proportion of correct responses in identifying dominant and submissive sources) in the retrieval task using a $2 \times 2 \times 2 \times 2$ mixed ANOVA with instructions (memorization, impression formation) as a between-groups factor and face–behavior congruence (congruent, incongruent), lure facial characteristics (different, matched), and age of face (younger, older) as within-group factors.

As was predicted, participants receiving impression formation instructions ($M = 79.23\%$, $SD = 12.98\%$) had greater accuracy than did participants receiving memorization instructions ($M = 70.70\%$, $SD = 12.99\%$), $F(1, 46) = 5.18$, $p = .03$, $\eta_p^2 = .10$. Also as predicted, our young participants accurately identified more younger ($M = 79.70\%$, $SD = 10.29\%$) than older ($M = 70.20\%$, $SD = 9.80\%$) target faces, $F(1, 46) = 28.81$, $p < .001$, $\eta_p^2 = .39$.¹ There were no main effects of face–behavior congruence or lure facial characteristics. However, we did find the predicted face–behavior congruence \times lure facial characteristics interaction, $F(1, 46) = 6.19$, $p = .02$, $\eta_p^2 = .12$ (Fig. 2a). Contrasts showed that when the lure's facial characteristics were different from the target, participants showed less

accurate identification of targets whose behaviors were incongruent with facial characteristics ($M = 72.00\%$, $SD = 11.76\%$) versus congruent ($M = 77.80\%$, $SD = 12.25\%$), $F(1, 46) = 4.66$, $p = .04$, $\eta_p^2 = .09$. When the lure's facial characteristics matched the target, performance did not differ between incongruent ($M = 75.80\%$, $SD = 10.78\%$) and congruent ($M = 74.30\%$, $SD = 10.78\%$) face–behavior pairs, $F(1, 46) = 0.89$, $p = .35$.²

There was a marginal three-way interaction between face–behavior congruence, lure facial characteristics, and age of face, $F(1, 46) = 2.92$, $p = .09$, $\eta_p^2 = .06$ (Fig. 2b). Consistent with our prediction, the two-way interaction described above was seen for older adult face–behavior pairs, but not for younger adult ones. For older pairs in which lure and target babyfacedness differed, participants showed less accurate identification of targets whose behavior was incongruent with facial characteristics ($M = 64.84\%$, $SD = 23.30\%$) than when it was congruent ($M = 74.20\%$, $SD = 21.17\%$), $F(1, 46) = 6.25$, $p = .02$, $\eta_p^2 = .12$. For older pairs in which the lure and target were similar in babyfacedness, accuracy did not differ between incongruent ($M = 72.40\%$, $SD = 18.58\%$) and congruent ($M = 69.53\%$, $SD = 20.45\%$) face–behavior pairs, $F(1, 46) = 1.09$, $p = .30$. For younger face–behavior pairs in which lure and target babyfacedness differed, there was no difference in accuracy between incongruent ($M = 79.17\%$, $SD = 21.32\%$) and congruent ($M = 81.00\%$, $SD = 21.40\%$) face–behavior pairs, $F(1, 46) = 0.30$, $p = .58$. There also was no difference in accuracy between incongruent ($M = 79.17\%$, $SD = 17.74\%$) and congruent ($M = 79.00\%$, $SD = 18.10\%$) younger face–behavior pairs when lure and target babyfacedness were similar, $F(1, 46) = 0$, $p = .99$. No other interactions approached significance.

Discussion

Appearance-based inferences biased source memory. Memory for sources who had behaved dominantly or submissively was influenced by whether their behaviors were congruent or incongruent with their babyfacedness and whether their babyfacedness differed from the lure's. When

¹ We added age order (younger first, older first) into the previously described ANOVA to assess whether viewing older before younger faces would impact the main effect of age of face. There was no performance difference between participants viewing older ($M = 76.20\%$, $SD = 12.74\%$) or younger ($M = 73.70\%$, $SD = 12.74\%$) faces first, $F(1, 44) = 0.46$, $p = .50$. All other effects maintained direction and significance.

² The tests in the present study are based on the congruity of information previously associated with *target* faces. However, because the lure was also previously encoded, it is also possible that *lure* face–behavior congruence influenced source memory. We tested for this possibility in a 2×2 ANOVA using target face–behavior congruence (congruent, incongruent) and lure face–behavior congruence (congruent, incongruent) as factors. Participants had enhanced source memory when lures were congruent ($M = 76.80\%$, $SD = 14.55\%$) over incongruent ($M = 73.20\%$, $SD = 14.55\%$), $F(1, 47) = 6.10$, $p = .02$, $\eta_p^2 = .12$. No other effects approached significance. Because a lure's behavior would never match the prompt at retrieval, a congruent lure might make the target easier to endorse over an incongruent lure. This provides further evidence that appearance-based inferences bias source memory.

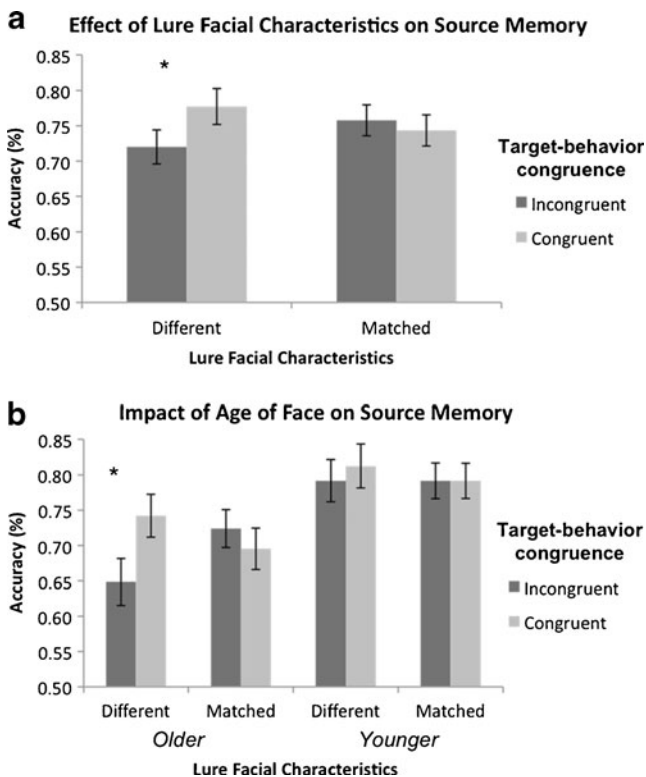


Fig. 2 Experiment 1 results. **a** Participants identified fewer sources whose behaviors were incongruent, relative to congruent, with facial characteristics when the lure had facial characteristics differing from the target. When the lure's facial characteristics matched the target's face, this bias was not apparent. **b** This interaction was marginally influenced by actor age. Participants remembered fewer older than younger sources, and older face-behavior incongruent targets paired with lures having differing facial characteristics were misattributed more than congruent targets. There is a nonsignificant visual trend in this direction for younger faces. * $p < .05$

target-behavior pairs were incongruent (e.g., mature-faced individual performing a submissive behavior) and lure faces mismatched the targets (and thus were congruent with the behavior), participants were less likely to accurately identify the target when recalling dominance or submissiveness. This occurred even though lure faces had been previously paired with behaviors mismatching the source attribution question. This indicates that the strength of appearance-based inferences overrode previous exposure to the actual target and lure behaviors. There was no similar appearance-based source memory bias when congruent and incongruent target-behavior pairs were coupled with lures whose facial characteristics matched the targets. Although face-behavior congruity did not influence source memory overall, the data demonstrate that appearance affected source attribution in some conditions, extending previous research using facial trustworthiness (Nash et al., 2010; Suzuki & Suga, 2010) to the babyface stereotype.

More younger than older sources were correctly identified, consistent with other work showing that young adults

have an own-age bias in face recognition (Anastasi & Rhodes, 2006). More novel was our finding that target age also influenced whether lure face congruence coupled with target-behavior congruence qualified source memory. For older targets, when lure facial characteristics mismatched the targets, participants identified fewer incongruent than congruent sources, whereas this bias was not evident for younger targets. This may relate to a visual effect of self-reference (Symons & Johnson, 1997) in memory, referring to the tendency to remember more information that relates to oneself over material with less personal relevance. In the present experiment, viewing younger relative to older faces might have enhanced self-relevance and resulted in a deeper level of encoding for younger adults. However, it is unclear whether relatively high overall memory performance in the present experiment limited the effect of congruity when younger sources were identified or whether younger faces were resistant to this effect due to their in-group status.

Participants given impression formation instructions had more accurate source memory than did those given memorization instructions, extending impression formation work showing this distinction (Chartrand & Bargh, 1996; Hamilton et al., 1980) to the appearance-based congruity effects present in source memory (Nash et al., 2010; Suzuki & Suga, 2010). Participants with an impression formation goal may have spontaneously formed more impressions than those with a memorization goal, thereby enhancing memory. Better performance when given impression formation than when given memorization instructions might also reflect transfer-appropriate learning (Morris, Bransford, & Franks, 1977), such that individuals given impression formation instructions during encoding would be predicted to perform better when making impression-based source inferences at retrieval, in contrast to individuals given a more general memorization goal. For example, reading with a memorization goal for a future job prediction task may have caused participants to attend to details unrelated to dominance or submissiveness, potentially impairing performance at retrieval. Interestingly, encoding material with an impression formation goal engages a neural network associated with impression formation and social cognition, relative to encoding material with a memorization goal (Mitchell et al., 2004). Our results suggest that engagement of this network may enhance accurate memory about targets' behaviors. Additional research using different instruction types in both behavioral and neuroimaging environments would be useful for testing this hypothesis. Notably, better memory under impression formation over memorization instructions did not moderate the effect of behavioral congruence and lure similarity. This suggests that the tendency to use appearance-based inferences as source decision clues may be prevalent across many contexts.

Experiment 2

An unresolved question from Experiment 1 had to do with why participants tended to misidentify more incongruent older sources when lure facial characteristics mismatched the targets, while this effect was not shown for younger sources. One reason for this difference could be an in-group bias for encoding and retrieving source information about younger faces that is resistant to congruity effects. However, relatively high overall performance could also limit the presence of these effects when younger sources are identified. Because younger adults have less accurate source memory for older faces in general, source memory for older individuals may be more vulnerable to error and may be more prone to bias and congruity effects. In contrast, more accurate overall memory for younger faces may limit the emergence of congruity effects in source memory.

In Experiment 2, we investigated this question by decreasing encoding time and increasing the retention interval in order to reduce overall source memory performance. If younger faces were resistant to congruity effects, these encoding and retention interval time changes should not affect memory for young sources but would leave older sources particularly vulnerable to congruity effects, especially in the case of older incongruent face–behavior pairs. On the other hand, if the presence of congruity effects in source memory had been limited by high performance when younger sources were identified, making the task more difficult not only might proportionately decrease source memory performance when younger and older sources were identified, but also might exacerbate congruity effects such that memory for incongruent, relative to congruent, sources would be reduced regardless of the source’s age or the lure’s facial characteristics.

Method

Participants

Thirty-two young adults (18 to 31 years old, 9 males; $M = 19.81$, $SD = 2.38$) recruited from Brandeis University, who had not participated in Experiment 1 or provided face and sentence ratings, participated.

Procedure

Procedures were modified from Experiment 1, with the following differences. All participants were given the impression formation instructions as described in Experiment 1. Consistent with Experiment 1, participants viewed each face alone for 2 s and were told to press the number 1 after viewing the face. The time participants viewed the same face paired with a behavioral sentence was reduced from 5

to 3 s. Participants were told to press the number 2 after reading the sentence. The retention interval was increased to 8 min (vs. 3.5 min in Experiment 1), during which participants completed cognitive measures. All other aspects of the encoding and retrieval tasks were equivalent to those in Experiment 1.

Results

Retrieval accuracy

We analyzed participants’ accuracy (proportion of correct responses in identifying dominant and submissive sources) in the retrieval task, using a $2 \times 2 \times 2$ ANOVA with face–behavior congruence (congruent, incongruent), lure facial characteristics (different, matched), and age of face (younger, older) as factors.

Overall, retrieval accuracy was reduced to 68.16 % using the modified encoding task (as compared with 74.97 % in Experiment 1). All retrieval accuracy scores were above chance. Participants again accurately identified more younger ($M = 72.60$ %, $SD = 12.46$ %) than older ($M = 63.80$ %, $SD = 11.88$ %) target faces, $F(1, 31) = 15.73$, $p < .001$, $\eta_p^2 = .34$ (Fig. 3a). However, unlike in Experiment 1, participants also accurately identified fewer targets whose behaviors were incongruent ($M = 65.20$ %, $SD = 13.58$ %) versus congruent ($M = 71.10$ %, $SD = 11.31$ %) with their facial characteristics, $F(1, 31) = 5.74$, $p = .02$, $\eta_p^2 = .16$. Also replicating Experiment 1, there was an interaction between face–behavior congruence and lure facial characteristics, $F(1, 31) = 4.49$, $p = .04$, $\eta_p^2 = .13$ (Fig. 3b). When the lure’s facial characteristics were different from those of the target, participants showed less accurate identification of targets whose behaviors were incongruent with facial characteristics ($M = 62.30$ %, $SD = 16.97$ %) versus congruent ($M = 73.40$ %, $SD = 15.84$ %), $F(1, 31) = 6.73$, $p = .01$, $\eta_p^2 = .18$. When the lure’s facial characteristics matched those of the target, performance did not differ between incongruent ($M = 68.20$ %, $SD = 14.71$ %) and congruent ($M = 68.80$ %, $SD = 13.58$ %) face–behavior pairs, $F(1, 31) = 0.81$, $p = .81$, $\eta_p^2 < .01$. There was no three-way interaction between face–behavior congruence, lure facial characteristics, and age of face, $F(1, 31) = 0.05$, $p = .83$, $\eta_p^2 < .01$. Thus, as in Experiment 1, the tendency for face–behavior incongruence to reduce source memory was moderated by lure appearance. Unlike in Experiment 1, however, this tendency was equally strong for younger and older targets. No other effects approached significance.

Discussion

By increasing task difficulty, we sought to reduce performance overall to better determine how congruity effects

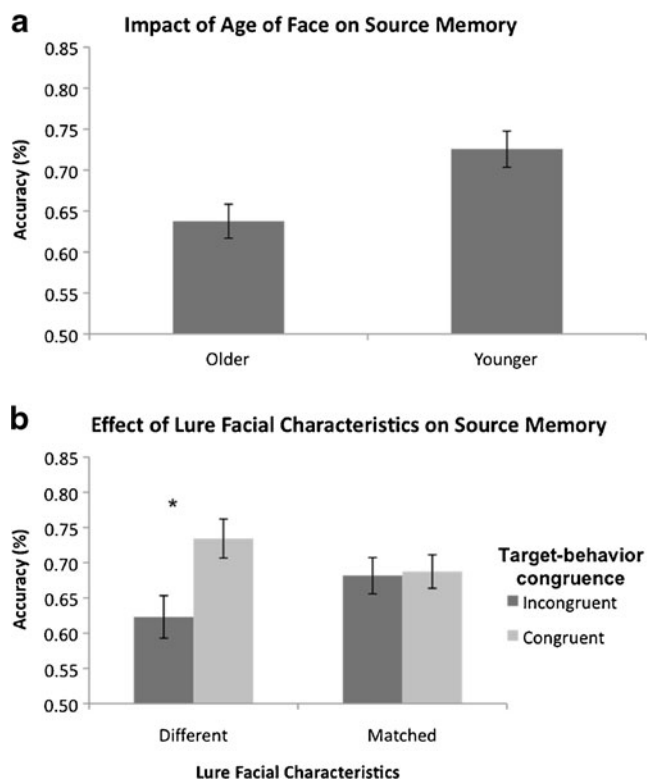


Fig. 3 Experiment 2 results. **a** Participants better identified younger over older sources. **b** Replicating Experiment 1, participants identified fewer sources whose behaviors were incongruent, relative to congruent, with facial characteristics when the lure had facial characteristics differing from, but not matching, the target. $*p < .05$

impact source memory and predicted that source memory could be affected in one of two different ways. More specifically, we wanted to examine whether the tendency to misidentify more incongruent older over younger sources when lure facial characteristics mismatched the targets was due to an in-group bias for younger faces that was resistant to congruity effects or whether high overall performance when younger sources were identified limited the emergence of congruity effects across conditions, such that they emerged only for targets most vulnerable to these effects. Replicating the findings of Experiment 1, greater accuracy in identifying younger over older sources remained significant in Experiment 2. In addition, the tendency to identify more targets whose behavior was congruent, rather than incongruent, with their appearance was moderated by lure appearance. However, unlike in Experiment 1, this tendency was not further moderated by target age. This demonstrates that as increasing task difficulty worsened source memory performance, younger faces became vulnerable to the congruity effects that were shown only for older faces in Experiment 1, even though overall accuracy differences in identifying younger versus older sources persisted. Thus, it seems that younger faces are not generally more resistant to congruity effects than are older faces when younger

perceivers are remembering sources. Rather, younger faces are equally vulnerable given a more difficult task that results in less accurate source memory overall.

Unlike in Experiment 1, participants accurately identified fewer sources whose facial characteristics were incongruent, relative to congruent, with their behaviors across conditions, perhaps because less stereotype-inconsistent material is memorable under increased task demands when compared with stereotype-consistent information (Macrae, Hewstone, & Griffiths, 2006). In a source memory context, people tend to rely on schemas at retrieval (Hicks & Cockman, 2003; Marsh et al., 2006; Nash et al., 2010). Due to the increased encoding demands of this task, relative to Experiment 1, the initial encoding of incongruent face–behavior information may have been particularly difficult, relative to the encoding of stereotype-congruent face–behavior information. Although some work shows that people remember more stereotype-inconsistent than -consistent information (Stangor & McMillan, 1992), this pattern is reversed as task demands increase and encoding becomes more difficult (Macrae et al., 2006). Even though source memory performance varied by whether target and lure facial characteristics matched or differed, the overall tendency to remember fewer incongruent than congruent sources suggests that the initial encoding of incongruent face–behavior pairs may have been impeded by the task’s increased difficulty, relative to Experiment 1. Therefore, the present experiment shows that although congruity effects may persist in source memory across many conditions, the severity of these effects may depend on the difficulty of the task at hand.

General discussion

In two experiments, we investigated the contributions of intuitive appearance-based impressions to source memory. Previous work (Kleider et al., 2012; Nash et al., 2010; Suzuki & Suga, 2010) has shown that facial characteristics can bias source monitoring, particularly when facial characteristics mismatch actor behaviors (e.g., misattributing an implausible newspaper headline to an untrustworthy-looking source). The present experiments extended this literature by using the babyface stereotype and revealed that if a person *looks* dominant, that person will be remembered as dominant, over a person whose facial characteristics convey submissiveness, even if previous behaviors say otherwise. Although this effect was limited to other-age older faces in Experiment 1, it was present for both younger and older faces in Experiment 2, where decreasing the time to encode person information also resulted in less accurate identification of incongruent, relative to congruent, sources in general. The fact that limiting encoding time exacerbates the effects of face–behavior congruity on memory suggests that

the basis of the congruity effects demonstrated in these experiments began at encoding, rather than purely at retrieval. At the same time, the appearance of the lure moderated the congruity effects in both experiments, demonstrating that congruity effects can also reflect processes operating at retrieval.

In Experiment 1, memory for sources whose behaviors were congruent or incongruent with their babyfacedness varied depending on whether their babyfacedness differed from the lures at retrieval. More specifically, participants made more source misattributions when target-behavior pairs were incongruent, relative to congruent, and lure faces mismatched the targets such that the target's behavior was congruent with the lure's face. Moreover, this pattern was shown only for older targets, consistent with previous work showing that people better remember their same-age peers (Anastasi & Rhodes, 2006). When the encoding task became more difficult in Experiment 2, participants' memory for incongruent, but not congruent, targets continued to be moderated by the facial characteristics of the lure, but this relationship persisted regardless of the target's age. Moreover, unlike in Experiment 1, there was decreased memory for incongruent over congruent sources overall. The higher overall performance in Experiment 1 may have limited the effects of face-behavior congruity on source memory, such that decreased performance appeared in arguably the most difficult condition at retrieval, where an incongruent target was paired with a lure of differing facial characteristics and the target and lure were not of the same age as participants. In this context, the lure provided a tempting schema-consistent source decision cue. This is consistent with work showing that people tend to use schema-consistent information when making source-monitoring decisions (Hicks & Cockman, 2003; Marsh et al., 2006; Nash et al., 2010). The findings also demonstrate that congruity effects are apparent when sources of various ages are identified but may not manifest among members of one's in-group until conditions reach a particular level of difficulty.

The source misattributions demonstrated in these studies both reflect the use of stereotypes to help process our surroundings (Macrae, Bodenhausen, & Milne, 1994) and also highlight bias as a "sin" of memory (Schacter, 2001). It may be easier to rely on stereotypical generalizations (e.g., a babyfaced person is submissive) when we encode and retrieve person information than to comprehensively process all incoming information, especially in difficult task conditions. The present work extends evidence that people implicitly use schematic, stereotyped information to make source-monitoring decisions (Bayen et al., 2000; Hicks & Cockman, 2003; Marsh et al., 2006; Sherman & Bessenoff, 1999) to the overgeneralized inferences shown in the babyface stereotype (Montepare & Zebrowitz, 1998). Interestingly, the effects demonstrated in this work may also, in part, represent implicit biases, such that when individuals were unable to

explicitly retrieve person information, they may have used implicit information gleaned from facial appearance to make decisions during the memory task. Thus, the present findings may be related to both source memory errors of source and implicit bias. Future work may clarify this distinction through comparing variations of retrieval tasks. Regardless, this work provides initial evidence that reactions to childlike and more adultlike facial appearances may serve as cues for an inferential bias in source memory, with appearance overriding the effects of learned behavioral information in certain conditions.

These findings may have intriguing implications for the legal system, particularly the powerful influence of eyewitness testimony on jury behavior (Leippe, 1995), since alleged perpetrators and eyewitnesses may not always be of the same age group. For example, if facial similarity and age, along with height and build, are equated among individuals selected for police lineups, witnesses may better identify suspects. Interestingly, some work shows that older adults are not as susceptible to own-age biases in face memory as are younger adults (Harvard & Memon, 2009; Wiese, Schweinberger, & Hansen, 2008), perhaps due to greater perceptual experience with faces from both age groups. Future aging research can clarify whether self-relevance or perceptual experience (i.e., experience with both age groups) drives the current effects.

This work provides the first evidence that appearance-based reactions to babyfaced and mature-faced individuals influence source memory, showing that despite learned behavioral information, these appearance-based inferences persist. This work also suggests important potential implications in the legal system. Appearance-based inferences about plaintiffs and defendants can influence jury behavior and the accurate attribution of source information by eyewitnesses, both of which may undermine legal proceedings. Appearances critically impact our impressions of others. The potential ramifications of facial characteristics for memory and decision-making processes should garner consideration in future research.

Acknowledgments We thank Avi Aizenman and Victoria Jonas for research assistance. The National Institute on Aging Grant R21 AG032382 (to A.H.G.), an NSF graduate fellowship (to B.S.C.), and a Brandeis University Graduate Student Association research grant (to B.S.C) supported this work.

References

- Anastasi, J., & Rhodes, M. G. (2006). Evidence for an own-age bias in face recognition. *North American Journal of Psychology*, 8(2), 237–252.
- Bayen, U., Nakamura, G., Dupuis, S., & Yang, C. (2000). The use of schematic knowledge about sources in source monitoring. *Memory and Cognition*, 28(3), 480–500.

- Berry, D., & McArthur, L. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin*, *100*, 3–18.
- Berry, D., & Zebrowitz-McArthur, L. (1988). What's in a face? Facial maturity and the attribution of legal responsibility. *Personality and Social Psychology Bulletin*, *14*(1), 23–33.
- Brownlow, S., & Zebrowitz, L. (1990). Facial appearance, gender, and credibility in television commercials. *Journal of Nonverbal Behavior*, *14*(51–60).
- Cassidy, B., & Gutchess, A. (2012). Social relevance enhances memory for impressions in older adults. *Memory*, *20*(4), 332–345.
- Chartrand, T. L., & Bargh, J. A. (1996). Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions. *Journal of Personality and Social Psychology*, *71*(3), 464–478.
- Collins, M., & Zebrowitz, L. (1995). The contributions of appearance to occupational outcomes in civilian and military settings. *Journal of Applied Social Psychology*, *25*, 129–163.
- Fiske, S., Neuberg, S. L., Beattle, A., & Milberg, S. (1987). Category-based and attribute-based reactions to others: some informational conditions of stereotyping and individuating processes. *Journal of Experimental Social Psychology*, *23*(5), 399–427.
- Hamilton, D., Katz, L., & Leirer, V. (1980). Cognitive representation of personality impressions: organizational processes in first impression formation. *Journal of Personality and Social Psychology*, *39*(6), 1050–1063.
- Havard, C., & Memon, A. (2009). The influence of face age on identification from a video line-up: A comparison between older and younger adults. *Memory*, *17*(8), 847–859.
- Hedden, T., Park, D., Nisbett, R., Ji, L., Jing, Q., & Jiao, S. (2002). Cultural variation in verbal versus spatial neuropsychological function across the lifespan. *Neuropsychology*, *16*, 65–73.
- Heider, J., Scherer, C., Skowronski, J. J., Wood, S., Edlund, J., & Hartnett, J. (2007). Trait expectancies and stereotype expectancies have the same effect on person memory. *Journal of Experimental Social Psychology*, *43*(2), 265–272.
- Hicks, J., & Cockman, D. (2003). The effect of general knowledge on source memory and decision processes. *Journal of Memory and Language*, *48*(3), 489–501.
- Johnson, M., Hashtroudi, S., & Lindsay, D. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28.
- Kleider, H., Cavrak, S., & Knuycky, L. (2012). Looking like a criminal: stereotypical black facial features promote face source memory error. *Memory and Cognition*. doi:10.3758/s13421-012-0229-x
- Leippe, M. (1995). The case for expert testimony about eyewitness memory. *Psychology, Public Policy, and Law*, *1*(4), 909–959.
- Macrae, C. N., Bodenhausen, G., & Milne, A. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, *66*, 34–47.
- Macrae, C. N., Hewstone, M., & Griffiths, J. (2006). Processing load and memory for stereotype-based information. *European Journal of Social Psychology*, *23*(1), 77–87.
- Marsh, R., Cook, G., & Hicks, J. (2006). Gender and orientation stereotypes bias source-monitoring attributions. *Memory*, *14*(2), 148–160.
- Mather, M., Johnson, M., & DeLeonardis, D. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology*, *16*(3/4/5), 437–458.
- Minear, M., & Park, D. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 630–633.
- Mitchell, J., Macrae, C., & Banaji, M. (2004). Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience*, *24*(21), 4912–4917.
- Montepare, J., & Zebrowitz, L. (1998). *Person perception comes of age: The salience and significance of age in social judgments* (Vol. 30). San Diego, CA: Academic Press.
- Morris, C., Bransford, J., & Franks, J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533.
- Nash, R., Bryer, O., & Schlagh, F. (2010). Look who's talking! Facial appearance can bias source monitoring. *Memory*, *18*(4), 451–457.
- Olivola, C., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*, 83–110.
- Schacter, D. L. (2001). *The seven sins of memory*. Boston: Houghton Mifflin Company.
- Sharot, T., & Phelps, E. (2004). How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(3), 294–306.
- Sherman, J., & Bessenoff, G. (1999). Stereotypes as source-monitoring cues: On the interaction between episodic and semantic memory. *Psychological Science*, *10*(2), 106–110.
- Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, *111*(1), 42–61.
- Suzuki, A., & Suga, S. (2010). Enhanced memory for the wolf in sheep's clothing: Facial trustworthiness modulates face-trait associative memory. *Cognition*, *117*, 224–229.
- Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*(3), 371–394.
- Todorov, A., & Uleman, J. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*(6), 549–562.
- Uleman, J. (1988). Trait and gist inference norms for over 300 potential trait-implying sentences. *Unpublished raw data*.
- Wiese, H., Schweinberger, S., & Hansen, K. (2008). The age of the beholder: ERP evidence of an own-age bias in face memory. *Neuropsychologia*, *46*(12), 2973–2985.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-Ms exposure to a face. *Psychological Science*, *17*(7), 592–598.
- Zebrowitz, L. (1997). *Reading faces: Window to the soul?* Boulder, CO: Westview Press.
- Zebrowitz, L., Kendall-Tackett, K., & Fafel, J. (1991). The influence of children's facial maturity on parental expectations and punishments. *Journal of Experimental Child Psychology*, *52*(2), 221–238.
- Zebrowitz, L., & McDonald, S. (1991). The impact of litigants' babyfacedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior*, *15*(6), 603–623.
- Zebrowitz, L., & Montepare, J. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517.
- Zebrowitz, L., Tenenbaum, D., & Goldstein, L. (1991). The impact of job applicants' facial maturity, gender, and academic achievement on hiring recommendations. *Journal of Applied Social Psychology*, *21*(7), 525–548.