Psychological Bulletin
2012, Vol. 138, No. 1, 146–174

0033-2909/11/$12.00 DOI: 10.1037/a0025750

The Own-Age Bias in Face Recognition: A Meta-Analytic and Theoretical Review

Matthew G. Rhodes
Colorado State University

Jeffrey S. Anastasi
Sam Houston State University

A large number of studies have examined the finding that recognition memory for faces of one's own age group is often superior to memory for faces of another age group. We examined this *own-age bias* (OAB) in the meta-analyses reported. These data showed that hits were reliably greater for same-age relative to other-age faces ($g = 0.23$) and that false alarms were reliably less likely for same-age compared with other-age faces ($g = -0.23$). Further meta-analyses of measures of signal detection demonstrated that, although no difference in response criterion was evident ($g = -0.01$), discriminability was reliably better for same-age compared with other-age faces ($g = 0.37$). As well, children, younger adults, and older adults exhibited superior discriminability for same-age compared with other-age age faces. Thus, the OAB appears to be a robust effect that influences the accuracy of face recognition. Theoretical accounts of the OAB have generally suggested that it reflects more extensive, recent experiences with one's own age group relative to other-age groups. Additional analyses were supportive of this account as the OAB was present even for groups (e.g., older adults) that had prior experiences as members of another age group. However, the most comprehensive account of the OAB will also likely invoke mechanisms suggested by social–cognitive theories.

Keywords: own-age bias, face recognition, recognition, ingroup/outgroup bias

Faces are perhaps the most conspicuous feature of daily social interactions. The face reveals a host of information, including details of age, gender, and other identity-relevant information specific to the individual. In addition to information about physical appearance, the face reveals information about the individual's mood, his or her attentiveness, and potentially his or her intentions (Bruce & Young, 1986). The face also likely provides the most distinctive information available to identify an individual as someone who is known.

Although human beings have extraordinary abilities to recognize previously seen faces, a number of factors influence the accuracy of face recognition (Shapiro & Penrod, 1986). This is perhaps most important in the context of the justice system, which is still largely reliant on eyewitnesses as a significant source of information in investigating crimes (see Wells, Memon, & Penrod, 2006, for a review). However, the perils of relying on eyewitnesses have been well documented. For example, the Innocence Project (Scheck, Neufeld, & Dwyer, 2000; see also http://www.innocenceproject.org/) comprises a group of lawyers who seek to assist prisoners in cases where innocence may be proven through DNA testing. As of August 2011, the group had assisted 273 prisoners in obtaining their exoneration and release. The most common cause of these wrongful convictions has been mistaken eyewitness identifications. Given such stakes, it is imperative that science develop a complete understanding of those factors that influence face recognition.

Of greatest relevance to the current article is research suggesting that recognition memory is superior for ingroup relative to outgroup faces. Perhaps the most widely investigated example of this phenomenon is the *own-race bias*, the finding that memory for faces of one's own race or ethnicity is frequently superior to memory for faces of another race or ethnicity (Malpass & Kravitz, 1969; see Meissner & Brigham, 2001, for a review and meta-analysis). Similar results have been reported indicating that memory is better for faces of one's own gender (e.g., D. B. Wright & Sladden, 2003), species (Diamond & Carey, 1986; see also Pascalis et al., 2005), or sexual orientation (Rule, Ambady, Adams, & Macrae, 2007). Moreover, several recent studies (e.g., Bernstein, Young, & Hugenberg, 2007; Hehman, Mania, & Gaertner, 2010; Hugenberg & Corneille, 2009; cf. MacLin & Malpass, 2001) have indicated that such ingroup memory biases may also be manifest for perceptually ambiguous faces (i.e., faces that provide no perceptual information regarding group membership). For example, Bernstein et al. (2007) had participants study faces presented on a red or green background, with the red background denoting individuals putatively attending the same university as the participants and the green background denoting individuals attending a rival university. Results showed that recognition was superior for individuals depicted as being from the same university than from a rival university despite the fact that, across participants, there were no differences in target faces. Thus, face recognition appears to be superior for ingroup faces across a variety of contexts for both perceptually identifiable and perceptually ambiguous faces.

This article was published Online First November 7, 2011.

Matthew G. Rhodes, Department of Psychology, Colorado State University; Jeffrey S. Anastasi, Department of Psychology, Sam Houston State University.

Correspondence concerning this article should be addressed to Matthew G. Rhodes, Department of Psychology, Colorado State University, Fort Collins, CO 80523-1876. E-mail: matthew.rhodes@colostate.edu

In the current article, we focus on another source of ingroup bias in face recognition, namely, the finding that memory is often superior for faces of one's own age group relative to faces of another age group (i.e., the *own-age bias*; *OAB*). The face provides a number of cues that disclose the age of the individual (for reviews, see Berry & McArthur, 1986; Enlow, 1982; M. G. Rhodes, 2009). For example, a newborn's head looks wide and vertically short with a relatively small, puglike nose and jaw bones that have yet to fully develop. Additionally, although the eyes of a young child appear large relative to the face, the appearance of the eyes becomes proportionally smaller as the nasal and jaw regions grow. Likewise, the forehead of a child looks large and high because the facial features below it have not fully developed. The forehead continues to grow (and becomes more sloping) but appears more proportionate as the remainder of the face develops. Development is also characterized by the growth of the nose and nasal bridge (necessary to accommodate a progressively larger lung capacity) into a more angular shape. Although growth ceases at approximately 20 years of age, the face continues to change. Nose and ear cartilage grow, enlarging the size of those features. The skin also undergoes an array of changes in texture and appearance, initially being soft and firm during childhood and drooping, sagging, and wrinkling by middle age. In addition to such wrinkling, hair becomes gray and of a lesser quantity; eyebrows become thicker; the relative size of the eyes, now more sunken, changes; and the shape of the lips changes as fatty tissue dissipates.

Perhaps because a rich set of facial cues indicative of age are available, several lines of evidence suggest that age information is rapidly extracted. For example, on the basis of event-related potential (ERP) data, Mouchetant-Rostaing and Giard (2003; see also Ebner, He, Fichtenholtz, McCarthy, & Johnson, 2011; Wiese, Schweinberger, & Neumann, 2008) concluded that age processing occurs within 145–185 ms of the onset of a face. Other behavioral data indicate that individuals rapidly categorize the age of a face (e.g., Johnston, Kanazawa, Kato, & Oda, 1997; see also M. G. Rhodes, 2009, for a review) and can make such age categorizations accurately following presentations as brief as 200 ms (e.g., Bruyer, Lafalize, & Distefano, 1991; Bruyer, Mejias, & Doublet, 2007). As well, age appears to be one of several essential sources of information that are extracted in obligatory fashion from familiar and unfamiliar faces (cf. Bruce & Young, 1986; G. Rhodes, 1988; Valentine, 1991; Wright et al., 2008). The impact of this obligatory processing on memory for faces may be dependent on both the age of the rememberer and the age of the target face.

Studies of the interaction of the age of the rememberer and the age of target faces typically involve presenting participants with a succession of faces, some from the same age group and some from a different age group, one at a time. Memory for such faces can then be tested by re-presenting the studied faces, most often in addition to faces that were not studied, for a test of recognition. The earliest work on the OAB (e.g., Bäckman, 1991; Bartlett & Leslie, 1986; Fulton & Bartlett, 1991) was motivated by the finding that children and older adults often exhibit poorer recognition performance than do younger adults (e.g., Adams-Price, 1992; Chance & Goldstein, 1984; see also Hildebrandt, Sommer, Herzmann, & Wilhelm, 2010). However, most conclusions about age-related differences in face recognition were drawn using college-aged stimuli. Thus, there was some question as to whether age-related differences in face recognition reflected a developmental change in memory for faces or an interaction between the age of the rememberer and the age of the face (i.e., an OAB).

Early work provided equivocal answers with regard to the prevalence of the OAB, foreshadowing patterns that would emerge in more recent studies. For example, Fulton and Bartlett (1991) had younger and older adults study faces of younger and older adults. Results from a subsequent recognition test showed that younger adults exhibited superior recognition for young compared with older faces, whereas older adults showed no difference in recognition on the basis of face age. Bäckman (1991) also observed that younger adults exhibited superior memory for faces of younger individuals compared with faces of older individuals. However, whereas a group of younger older adults (63–70 years of age) showed better memory for older compared with younger faces, two groups of older older adults (with mean ages of 76 years and 85 years, respectively) showed no such OAB. An early study with children (Goldstein & Chance, 1964) likewise provided mixed data regarding the prevalence of an OAB, but subsequent studies have suggested that children may exhibit an OAB (e.g., Anastasi & Rhodes, 2005; Hills & Lewis, 2011).

More recently, there has been a significant surge in interest in the interaction of participant age and target age in memory. For example, prior to 2000, approximately 15 studies examined memory for faces or memory for the characteristics of studied individuals (e.g., List, 1986) as a function of the age of the target and age of the rememberer. However, since 2000, at least 43 additional studies have been conducted, including 36 since 2005. Although a number of studies have revealed a robust OAB (e.g., Anastasi & Rhodes, 2005; He, Ebner, & Johnson, 2011; , Cassia, Picozzi, & Bricolo, 2008; Perfect & Moon, 2005), other studies have failed to observe an OAB (e.g., Ebner & Johnson, 2009; Memon, Bartlett, Rose, & Gray, 2003; Mondloch, Maurer, & Ahola, 2006; Rose, Bull, & Vrij, 2005; Wilcock, Bull, & Vrij, 2007). Such inconsistencies in part motivate the meta-analytic review reported in this article, as meta-analysis is ideally suited for determining whether the OAB is a reliable effect. However, more important, a meta-analysis can address theoretical perspectives on the OAB.

Theories of the OAB

The predominant account of the OAB (e.g., Kuefner et al., 2008; but see the General Discussion section for an alternative account) suggests that individuals have more extensive experience or contact with members of their own age group relative to other age groups. Such contact facilitates the development of perceptual expertise that supports face recognition for same-age compared with other-age faces (cf. Valentine, 1991). A parallel to this idea may be found in theories suggesting that the own-race bias reflects more extensive contact and thus greater perceptual expertise for own-race relative to other-race individuals (see Meissner & Brigham, 2001, for a review). Within the own-race bias literature, this has been investigated by examining self-reported contact with individuals of another race (e.g., Slone, Brigham, & Meissner, 2000) or by testing individuals who vary in the amount of contact with individuals of another race (e.g., Chiroro, Tredoux, Radaelli, & Meissner, 2008; Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005; D. B. Wright, Boyd, & Tredoux, 2003). For example, Sangrigoli et al. (2005) tested Korean adults who had been adopted as children (prior to the age of 9 years) by Caucasian families in France and Korean adults who had briefly resided in France. Adoptees exhibited better recognition of Caucasian compared with Asian faces, whereas Korean adults living briefly in France demonstrated superior memory for Asian compared with Caucasian faces (but see de Heering, de Liedekerke, Deboni, & Rossion, 2010). Thus, these data suggest a face processing system that is amenable to experience.

Similar strategies have been used within studies reported in the OAB literature to understand the role of experience and contact in recognition of same-age and other-age faces. For example, several studies have examined recognition of same- and other-age faces as a function of self-reported contact with these groups (Chance, Goldstein, & Andersen, 1986; Ebner & Johnson, 2009; He et al., 2011). Chance et al. (1986) had young adults study faces of adults and infants and collected data on the amount of experience each participant had providing care for infants. Although all participants exhibited superior recognition of adults compared with infants (i.e., an OAB was evident), participants with prior experience with infants demonstrated better memory for faces of infants than participants with little experience with infants. He et al. (2011) had younger and older adults study same-age and other-age faces and then take a recognition test. They also administered a measure of contact to participants, inquiring about personal exposure to other age groups, exposure through the media, and any other forms of exposure to other age groups. Both older and younger adults reported more contact with same-age relative to other-age individuals, and both groups exhibited an OAB in recognition memory. In addition, the difference in the amount of exposure to same- and other-age faces was reliably related ($\beta = .30$) to the difference in recognition for same-age versus other-age faces. Ebner and Johnson (2009) likewise administered a measure of contact to older and younger adults, and both groups reported more contact with same-age than other-age individuals. Participants also studied younger and older faces but did not exhibit an OAB on a subsequent test of recognition. For younger adults, the amount of contact with older adults was positively related to recognition of older faces ($\beta = .43$), whereas, for older adults, no reliable relationship was apparent between contact with younger adults and recognition of younger faces. Thus, on the basis of the few studies available, the amount of contact measured via questionnaires generally appears to be related to recognition of other-age faces, in contrast to evidence from the own-race bias literature suggesting little role for measured contact (Meissner & Brigham, 2001).

Another approach has been to test individuals with differing degrees of actual contact with another age group (Cassia, Kuefner, Picozzi, & Vescovo, 2009; Cassia, Picozzi, et al., 2009; Harrison & Hole, 2009; Kuefner et al., 2008; Kuefner, Cassia, Vescovo, & Picozzi, 2010). For example, Harrison and Hole (2009) had trainee teachers, with an average of 1.5 years of experience with children, and age-matched controls with no extensive experience with children, study the faces of children and same-aged adults. Control participants exhibited an OAB, with better memory for faces of adults than children. However, trainee teachers demonstrated no OAB and exhibited a nonsignificant trend to have superior memory for faces of children ($d' = 2.35$) relative to adults ($d' = 2.17$). Thus, extensive experience with another age group was associated with enhanced recognition of faces from that age group.

Although experience appears to play a role in the OAB, the nature of the experiences that drive the OAB remains open to debate. In particular, most previously documented ingroup biases in face recognition have tested individuals who will always belong to a specific group. For example, studies of the own-gender bias (e.g., Wright & Sladden, 2003), with only the rarest exceptions, involve individuals who will not become members of another gender during their lifetime. Likewise, studies of the own-race bias (e.g., Malpass & Kravitz, 1969) only include individuals who are or always will be a member of a particular race or ethnicity in their lifetime.[1] Thus, most studies of ingroup biases in face recognition include individuals whose ingroup membership will never be altered. The OAB represents a key exception to this regularity. That is, aging and development necessitate that individuals who were once members of a particular group (e.g., children) become members of other age groups (i.e., young adults, middle-aged adults, older adults). This progression of membership in multiple age groups may indicate that sufficient experience develops with prior age groups to facilitate recognition beyond one's current age group.

Cassia, Kuefner, et al. (2009) reported data indicating that more distant prior experiences may influence the OAB. In particular, they observed that mothers of infants only exhibited a marker of the OAB (i.e., superior memory for upright relative to inverted faces of adults but no inversion effect for infants) if they had not had a younger sibling as a child. In contrast, mothers who had a younger sibling as a child did not exhibit this marker of the OAB. Thus, for mothers who had a younger sibling as a child, a prior experience from childhood modulated perceptual expertise for other-age faces in adulthood. By extension, this might suggest that individuals with opportunities to develop expertise with other-age faces by virtue of prior membership in such groups (e.g., older adults' prior experience as young adults) would be less likely to exhibit the OAB.

Such a premise is consistent with several reports of younger adults but not older adults demonstrating the OAB (e.g., Bartlett & Leslie, 1986; Fulton & Bartlett, 1991; Havard & Memon, 2009; M. G. Rhodes, Castel, & Jacoby, 2008; Wiese, Schweinberger, & Hansen, 2008). For example, Wiese, Schweinberger, and Hansen (2008) had older and younger adults study old and young faces. Whereas younger adults exhibited a strong OAB, older adults showed no OAB. Wiese, Schweinberger, and Hansen speculated that young adults had primarily been exposed to younger adult faces during their lifetime and thus exhibited enhanced recognition of those faces. In contrast, older adults, who had formerly been young, had sufficient experience with both age groups to support recognition performance for both young and old faces (but see Freund, Kourilova, & Kuhl, 2011). As such, prior experiences may contribute to recognition of same-age and other-age faces and thus may eliminate or diminish the OAB.

An alternative experience-based account suggests that only the most recent experiences contribute to the OAB (Hills & Lewis, 2011; cf. Harrison & Hole, 2009; O'Neil & Webster, 2011). For example, Hills and Lewis (2011) had young adults and 4- to 6-year-old, 7- to 9-year-old, and 9- to 12-year-old children study the faces of young adults, older adults, and 8-year-old children.

[1] We note that race cannot be precisely defined (e.g., Malpass, 1993) and any designation of an individual as belonging to a particular race is purely a social construction.

Whereas younger adults demonstrated superior memory for faces of young adults relative to other age groups, an OAB in face recognition was only apparent for children 7–9 years old. However, children from other age groups exhibited no difference in recognition for faces of children compared with other age groups. There are important limitations of these data. First, as Hills and Lewis noted, a longitudinal design would be necessary to determine whether face recognition actually changes with age. Second, the age range of the child stimuli used consisted of only faces of 8-year-olds rather than faces from all age groups, leaving open the possibility that the effects reported may reflect the particular stimulus materials used rather than a general principle of face recognition. Nevertheless, Hills and Lewis's (2011) data are suggestive of a highly flexible face processing system that rapidly acquires expertise with frequently experienced faces but rapidly loses this expertise as the ability to distinguish faces of a particular age group becomes less important. From this perspective, the OAB should only be prevalent for those age groups with which individuals have the most frequent, recent experiences, likely those of the same age (but see Cassia, 2011, for a different version of this account).

Thus, we tested two variants of an experience-based account in the meta-analyses reported. One account suggests that only the most recent experiences contribute to the OAB (Hills & Lewis, 2011). Support for this would come from finding that the OAB is similar across age groups. An alternative account would suggest that prior experiences as a member of another age group mitigate the tendency for memory to be superior for same-age faces (cf. Lindholm, 2005). Support for this would come from finding variations in the OAB across age groups. This is best tested by examining effect sizes for older adults. That is, older adults have been a member of other age groups, in contrast to children and younger adults who have little or no experience as a member of another age group. Thus, finding that older adults do not exhibit an OAB would be consistent with the idea that experiences across a lifetime support recognition of same-age and other-age faces. In all, the OAB represents a unique opportunity, generally unavailable when examining other ingroup biases, to determine whether prior experience in an outgroup moderates subsequent ingroup biases in face recognition.

Overview of the Meta-Analyses

In the current study, we report a meta-analytic and theoretical review of the OAB literature. A meta-analysis is a quantitative summary of a research literature that describes and tests the magnitude of various effect sizes across studies. One can think of an effect size as a standardized score, analogous to a *z* score, for a dependent variable or variables of interest. Using such effect sizes, one can examine the influence of possible moderator variables and determine where effect sizes reliably converge or differ (Hedges & Olkin, 1985).

The meta-analyses reported can serve several purposes. In particular, prior work has provided mixed evidence regarding the existence of an OAB in face recognition, and meta-analysis is ideally suited for determining whether the OAB is a reliable effect. In addition, a number of methodological variations have been used in studies of the OAB, such as different types of criterion tests, different instructions at encoding, and variations in the time allotted to study faces. A meta-analysis can determine the degree to which these factors moderate the OAB. Most important, a meta-analytic review can address and clarify the role of experience in OAB. As noted previously, finding that the OAB is similar across age groups would be consistent with accounts suggesting that the OAB is driven by more recent experiences with one's own age group. In contrast, an asymmetric OAB, particularly any null effects for older adults, would suggest that experiences accrued over a lifetime drive the OAB.

We investigated the OAB by focusing on four key dependent measures. In particular, we examined the OAB with regard to hits (i.e., correctly identifying a previously studied face as "old") and false alarms (i.e., mistakenly identifying a face that was not previously studied as "old"). However, the levels of hits and false alarms alone do not provide a complete account of recognition performance. That is, optimal recognition performance is reflected by a high level of hits and a correspondingly low level of false alarms. Thus, for those studies that provided such data, we examined signal detection estimates (Macmillan & Creelman, 2005) of performance to examine estimates of discriminability and response criterion. *Discriminability* (i.e., sensitivity) refers to the ability to distinguish between previously studied and new items, typically measured as the standardized difference between hits and false alarms, respectively (i.e., d'). *Response criterion* refers to the amount of evidence the rememberer requires to deem an item "old." It is typically calculated on the basis of the distance from the intersection of the old and new distributions (i.e., $d'/2$) and is measured in standardized units: $C = z_{\mathrm{FA}} - d'/2$. Neutral responding is indicated by a value of 0, whereas values above 0 are indicative of conservative responding and values below 0 are indicative of liberal responding.[2]

Of key interest are patterns of data for discriminability. That is, a robust OAB would be evident in positive effect sizes for discriminability, whereas null effects (or negative effect sizes) for discriminability would argue against a general OAB. As well, reliable effect sizes for discriminability across age groups, particularly for older adults, would be consistent with a narrow, experienced-based account. In contrast, asymmetries in effect sizes for discriminability across age groups (e.g., no OAB for older adults) would be consistent with the notion that more distant prior experiences moderate the OAB. In all, we tested the robustness and basis for the OAB in the meta-analyses reported.

[2] We note that we could not calculate signal detection measures for studies using a forced-choice test. For a forced-choice test, participants are presented with a studied face and one or more lures simultaneously and must decide which stimulus was presented during the study phase. Hits are measured as the proportion of correct responses and false alarms as the proportion of incorrect responses, which is the complement (inverse) of the correct identification rate. Given that a participant must always deem one face *studied*, such measures are thought to be criterion neutral. Although it is possible to calculate a measure of discriminability for forced-choice tests (see Macmillan & Creelman, 2005), none of the studies reported provided sufficient data to do so. Likewise, signal detection measures are not possible for studies using a lineup at test. These are typically a variant on the forced-choice procedure and include only one or two observations per participant.

Method

Studies for the meta-analyses were obtained from a computerized search of PsycINFO, Web of Science, Google Scholar, and Dissertation Abstracts using the keywords *recognition*, *face recognition*, *other-age effect*, and *own-age bias*. Additional searches were conducted through manual inspection of published articles and through reverse citation searches of published articles. Finally, an attempt was made to document the unpublished literature by requesting unpublished data from researchers active in the face recognition literature. All searches were concluded by May 2011.

For an article to be included in the meta-analyses, the experiment(s) reported must have (a) manipulated the age of the face studied, comprising stimuli that were congruent with the age of the participants and at least one group of stimuli that was not congruent with the age of the participants; (b) used photorealistic, upright faces that were unfamiliar to the participants; (c) administered some form of a recognition memory test for the faces studied; and (d) reported means and standard deviations or enough statistical information (e.g., F, t, or p values) to permit the estimation of effect sizes.[3] When studies did not include enough information, we contacted the authors to ask them to provide the missing information. If this was unsuccessful, the study was either excluded from the meta-analyses or, if partial data were present, the effect size was conservatively estimated on the basis of the information reported in the text.

A total of 123 effect sizes from 43 studies met the inclusion criteria for the meta-analysis based on hits. These articles were published between 1964 and 2011 and included a total of 4,101 participants.

A total of 127 effect sizes from 42 studies met the inclusion criteria for the meta-analysis based on false alarms. These articles were published between 1964 and 2011 and included a total of 4,088 participants.

A total of 87 effect sizes from 27 studies met the inclusion criteria for the meta-analysis based on discriminability. These articles were published between 1986 and 2011 and included a total of 2,879 participants.

A total of 67 effect sizes from 22 studies met the inclusion criteria for the meta-analysis based on measures of response criterion. These articles were published between 1986 and 2011 and included a total of 2,237 participants.

Listing and Discussion of Coded Variables

The studies included in the meta-analyses are listed in Table 1. Each study was coded with respect to several variables that may influence the strength and direction of effect sizes. The majority of these variables comprised methodological characteristics of the experiments (e.g., study duration), although other classes of variables, such as the age of the participants tested, have been manipulated. We first discuss the categorical and then the continuous variables that were coded.

Age group. Although the majority of studies have tested college-aged students (i.e., younger adults), other studies have examined memory for same- and other-age faces in children (e.g., Anastasi & Rhodes, 2005), middle-aged adults (e.g., Anastasi & Rhodes, 2006; D. B. Wright & Stroud, 2002), and older adults (e.g., Bartlett & Leslie, 1986; Firestone, Turk-Browne, & Ryan, 2007). Thus, *age group* was coded as a categorical variable with levels of *children*, *younger adults*, *middle-aged adults*, and *older adults*. Children were coded as individuals under the age of 13 years, younger adults as individuals between the ages of 18 and 35 years, middle-aged adults as individuals between the ages of 40 and 55 years, and older adults as individuals over the age of 55 years.

Type of encoding. The majority of studies have not provided specific instructions at encoding, allowing participants to study faces in any manner chosen. Such *self-directed encoding* can be contrasted with other studies that have directed participants to attend to item-specific characteristics of the faces, such as pleasantness (e.g., Bartlett & Fulton, 1991; Bartlett & Leslie, 1986), kindness (Crookes & McKone, 2009), emotion (e.g., Ebner & Johnson, 2009), or the age of the individual pictured (e.g., Anastasi & Rhodes, 2006), among other tasks. An additional study instructed participants to engage in elaborative encoding by forming an association between two faces (M. G. Rhodes et al., 2008). Thus, *type of encoding* was coded as a categorical variable with levels of *self-directed encoding*, *item-specific encoding*, or *associative encoding*.[4]

Intention. In prior studies, with nearly equal frequency, researchers have either instructed participants to encode faces in preparation for an upcoming memory test (i.e., *intentional encoding*) or tested participants with no forewarning that they would be tested (i.e., *incidental encoding*). Thus, *intention* was coded as a categorical variable with levels of *intentional encoding*, *incidental encoding*, or *unspecified*.

Type of test. A variety of different tests have been used in previous studies. For tests of single items, this includes *item recognition* tests, which require the participant to judge a single

[3] Several studies were excluded on these grounds. For example, Slessor, Laird, Phillips, Bull, and Filippou (2010) and Ebner, He, Fichtenholtz, et al. (2011) examined gaze following in younger and older adults for young and old faces but did not collect recognition memory data. Likewise, Melinder, Gredebäck, Westerlund, and Nelson (2010) examined ERPs for young and old faces but did not collect recognition memory data. Other studies have investigated memory for same- and other-age faces but have used dependent measures that fall outside the scope of the meta-analyses reported (e.g., Bartlett, Strater, & Fulton, 1991; Lindholm, 2005; List, 1986; McGillivray & Castel, 2010; Yarmey, 1993). In addition, several studies have tested participants' memory for upright and inverted same- and other-age faces (e.g., Cassia, Kuefner, et al., 2009; Cassia, Picozzi, et al., 2009; Crookes & McKone, 2009; Kuefner, Cassia, Picozzi, & Bricolo, 2008; Perfect & Moon, 2005). Only data for upright faces were included in the meta-analyses reported. An additional study (Kuefner et al., 2010) only presented participants with the upper halves of same-age and other-age faces and was excluded. Finally, one study was excluded because the stimuli used were line drawings (McKelvie, Standing, St. Jean, & Law, 1993). We excluded data for inverted faces, halves of faces, or line drawings so as to base conclusions solely on data that reflected a natural presentation of the face.

[4] As noted previously, several studies (e.g., Anastasi & Rhodes, 2005, 2006; Firestone et al., 2007) had participants make judgments of the age of the individual during the encoding phase. One could argue that encoding tasks that specifically focus on the age of the face may be qualitatively different from other encoding tasks. However, analyses indicated that attending to age during encoding resulted in the same effect size as other forms of item-specific encoding. Thus, these tasks are grouped together for present purposes.

Table 1
Characteristics of the Studies Included in the Meta-Analyses

Study	Meta-analyses	Age group	Encoding	Intention	Type of test	Type of photo at test	Publication status	Encoding duration	No. of items at encoding	Retention interval	No. of test items	Test time
Anastasi & Rhodes (2005)	H, F, D, R											
		C	Item	Incidental	IR	Different	Published	10	32	5	64	10
		O	Item	Incidental	IR	Different	Published	10	32	5	64	10
Anastasi & Rhodes (2006)	H, F, D, R											
Experiment 1		Y	Item	Incidental	IR	Same	Published	7	24	18	48	7
Experiment 1		M	Item	Incidental	IR	Same	Published	7	24	18	48	7
Experiment 1		O	Item	Incidental	IR	Same	Published	7	24	18	48	7
Experiment 2		Y	Item	Incidental	IR	Different	Published	7	24	2,880	48	7
Experiment 2		O	Item	Incidental	IR	Different	Published	7	24	2,880	48	7
Bäckman (1991)	D											
Experiment 2		Y	Self	Unspecified	IR	Same	Published	5	60	20	120	SP
Experiment 2		O	Self	Unspecified	IR	Same	Published	5	60	20	120	SP
Experiment 2		O	Self	Unspecified	IR	Same	Published	5	60	20	120	SP
Experiment 2		O	Self	Unspecified	IR	Same	Published	5	60	20	120	SP
Bartlett & Fulton (1991)	H, F, D, R											
Experiment 2		O	Item	Incidental	IR	Different	Published	10	52	0	72	10
Experiment 2		Y	Item	Incidental	IR	Different	Published	10	52	0	72	10
Bartlett & Leslie (1986)	H, F, D, R											
Experiment 1		Y	Item	Incidental	IR	Combined	Published	12	48	0	72	10
Experiment 1		O	Item	Incidental	IR	Combined	Published	12	48	0	72	10
Experiment 1		Y	Item	Incidental	IR	Combined	Published	12	48	0	72	10
Experiment 1		O	Item	Incidental	IR	Combined	Published	12	48	0	72	10
Experiment 2		Y	Item	Incidental	IR	Combined	Published	12	48	0	72	10
Experiment 2		O	Item	Incidental	IR	Combined	Published	12	48	0	72	10
Cassia, Kuefner, Picozzi, & Vescovo (2009)	H, F											
Experiment 1		C	Self	Intentional	FC	Same	Published	5	1[a]	0	1[a]	SP
Experiment 1		C	Self	Intentional	FC	Same	Published	5	1[a]	0	1[a]	SP
Experiment 2		Y	Self	Intentional	FC	Same	Published	1	1[a]	0	1[a]	SP
Experiment 3		Y	Self	Intentional	FC	Same	Published	1	1[a]	0	1[a]	SP
Unpublished		Y	Self	Intentional	FC	Same	Unpublished	1	1[a]	0	1[a]	SP
Cassia, Picozzi, Kuefner, & Casati (2009)	H, F											
		Y	Self	Intentional	FC	Different	Published	1	1[a]	0	1[a]	SP
		Y	Self	Intentional	FC	Different	Published	1	1[a]	0	1[a]	SP
Chance, Goldstein, & Andersen (1986)	D											
Experiment 1		Y	Self	Intentional	IR	Same	Published	3	10	U	32	5
Chung (1997)	H, F, D											
		C	Self	Intentional	IR	Same	Published	3	32	0	64	SP
		C	Self	Intentional	IR	Same	Published	3	32	0	64	SP
		C	Self	Intentional	IR	Same	Published	3	32	0	64	SP
		C	Self	Intentional	IR	Same	Published	3	32	0	64	SP
		C	Self	Intentional	IR	Same	Published	3	32	0	64	SP
		Y	Self	Intentional	IR	Same	Published	3	32	0	64	SP

(table continues)

Table 1 (*continued*)

Study	Meta-analyses	Age group	Encoding	Intention	Type of test	Type of photo at test	Publication status	Encoding duration	No. of items at encoding	Retention interval	No. of test items	Test time
Crookes & McKone (2009)	H, F, D, R											
Experiment 3		C	Item	Intentional	IR	Same	Published	5	15[b]	4	30[b]	SP
Experiment 3		C	Item	Intentional	IR	Same	Published	5	15[b]	4	30[b]	SP
Experiment 3		Y	Item	Intentional	IR	Same	Published	5	15[b]	4	30[b]	SP
Ebner & Johnson (2009)	H, F, D, R											
		Y	Item	Incidental	IR	Same	Published	5	48	5	72	SP
		O	Item	Incidental	IR	Same	Published	5	48	5	72	SP
Ebner, Riediger, & Lindenberger (2009)	H, F, D, R											
Experiment 2		Y	Item	Incidental	IR	Same	Published	30	32	10,080	64	SP
Experiment 2		O	Item	Incidental	IR	Same	Published	30	32	10,080	64	SP
Firestone, Turk-Browne, & Ryan (2007)	H, F, D, R											
		Y	Item	Incidental	IR	Unspecified	Published	7	24	5	48	SP
		O	Item	Incidental	IR	Unspecified	Published	7	24	5	48	SP
Fulton & Bartlett (1991)	H, F, D, R											
		Y	Item	Incidental	IR	Different	Published	10	48	0	72	10
		O	Item	Incidental	IR	Different	Published	10	48	0	72	10
Gilchrist & McKone (2003)	H, F											
		Y	Item	Intentional	FC	Same	Published	5	30	5	30	SP
		C	Item	Intentional	FC	Same	Published	5	30	0	30	SP
Goldstein & Chance (1964)	H, F											
Experiment 1		C	Self	Intentional	FC	Same	Published	5	8[c]	0	8[c]	SP
Experiment 2		C	Self	Intentional	FC	Same	Published	5	8[c]	0	8[c]	SP
Goodsell, Neuschatz, & Gronlund (2009)	H, F											
Experiment 1		Y	Self	Intentional	L	Different	Published	25	1	10,080	1	SP
Experiment 1		Y	Self	Intentional	L	Different	Published	25	1	10,080	1	SP
Experiment 1		O	Self	Intentional	L	Different	Published	25	1	10,080	1	SP
Experiment 1		O	Self	Intentional	L	Different	Published	25	1	10,080	1	SP
Gross (2009)	H, F, D, R											
		Y	Self	Intentional	IR	Different	Published	7	32	0	64	SP
		Y	Self	Intentional	IR	Different	Published	7	32	0	64	SP
		Y	Self	Intentional	IR	Different	Published	7	32	0	64	SP
		Y	Self	Intentional	IR	Different	Published	7	32	0	64	SP
Harrison & Hole (2009)	H, F, D, R											
		Y	Self	Intentional	IR	Different	Published	3	32	3	64	2.5
		Y	Self	Intentional	IR	Different	Published	3	32	3	64	2.5
Havard & Memon (2009)	H, F											
		Y	Self	Intentional	L	Different	Published	90	2	50[d]	2	SP
		O	Self	Intentional	L	Different	Published	90	2	50[d]	2	SP
He, Ebner, & Johnson (2011)	H, F, D, R											
		Y	Self	Incidental	IR	Same	Published	4	48	10	96	3
		O	Self	Incidental	IR	Same	Published	4	48	10	96	3

Table 1 (*continued*)

Study	Meta-analyses	Age group	Encoding	Intention	Type of test	Type of photo at test	Publication status	Encoding duration	No. of items at encoding	Retention interval	No. of test items	Test time
Hills & Lewis (2011)	H, F, D, R											
		C	Item	Incidental	IR	Different	Published	Self-paced	24	2.5	48	SP
		C	Item	Incidental	IR	Different	Published	Self-paced	24	2.5	48	SP
		C	Item	Incidental	IR	Different	Published	Self-paced	24	2.5	48	SP
		Y	Item	Incidental	IR	Different	Published	Self-paced	24	2.5	48	SP
Hourihan, Benjamin, & Gronlund (2010)	H, F, D, R											
Experiment 1		Y	Self	Intentional	IR	Different	Unpublished	5	40	0	80	SP
Experiment 1		O	Self	Intentional	IR	Different	Unpublished	5	40	0	80	SP
Experiment 2		Y	Self	Intentional	IR	Different	Unpublished	3	60	0	120	SP
Experiment 2		O	Self	Intentional	IR	Different	Unpublished	3	60	0	120	SP
Kuefner, Cassia, Picozzi, & Bricolo (2008)	H, F											
Experiment 1		Y	Self	Intentional	FC	Unspecified	Published	1	1[a]	0	1[a]	SP
Experiment 2		Y	Self	Intentional	FC	Unspecified	Published	1	1[a]	0	1[a]	SP
Experiment 3		Y	Self	Intentional	FC	Unspecified	Published	1	1[a]	0	1[a]	SP
Unpublished		Y	Self	Intentional	FC	Unspecified	Unpublished	1	1[a]	0	1[a]	SP
Lamont, Stewart-Williams, & Podd (2005)	H, F, D, R											
		Y	Self	Unspecified	IR	Same	Published	5	20 or 40[e]	0	40 or 80[e]	8
		O	Self	Unspecified	IR	Same	Published	5	20 or 40[e]	0	40 or 80[e]	8
		O	Self	Unspecified	IR	Same	Published	5	20 or 40[e]	0	40 or 80[e]	8
Mason (1986)	H, F, D, R											
		Y	Self	Unspecified	IR	Same	Published	10	40	0	80	10
		O	Self	Unspecified	IR	Same	Published	10	40	0	80	10
Memon, Bartlett, Rose, & Gray (2003)	H, F											
		Y	Self	Incidental	L	Different	Published	43	2	35	2	SP
		Y	Self	Incidental	L	Different	Published	43	2	10,080	2	SP
		Y	Self	Incidental	L	Different	Published	43	2	35	2	SP
		Y	Self	Incidental	L	Different	Published	43	2	10,080	2	SP
		O	Self	Incidental	L	Different	Published	43	2	30	2	SP
		O	Self	Incidental	L	Different	Published	43	2	10,080	2	SP
		O	Self	Incidental	L	Different	Published	43	2	30	2	SP
		O	Self	Incidental	L	Different	Published	43	2	10,080	2	SP
Mondloch, Maurer, & Ahola (2006)	H, F, D, R											
Experiment 2		Y	Self	Intentional	IR	Same	Published	0.2	1[a]	0	1[a]	SP
Experiment 2		C	Self	Intentional	IR	Same	Published	0.2	1[a]	0	1[a]	SP
Perfect & Harris (2003)	H, F											
Experiment 3		Y	Self	Intentional	L	Same	Published	20	4	22.5	4	SP
Experiment 3		O	Self	Intentional	L	Same	Published	20	4	22.5	4	SP
Perfect & Moon (2005)	H, F, D, R											
		Y	Self	Intentional	IR	Different	Published	5	32	5	64	SP
		O	Self	Intentional	IR	Different	Published	5	32	5	64	SP
Pozzulo & Dempsey (2009)	H, F											
		Y	Self	Incidental	L	Different	Published	60	1	20	1	SP
		Y	Self	Incidental	L	Different	Published	60	1	20	1	SP

(table continues)

Table 1 (*continued*)

Study	Meta-analyses	Age group	Encoding	Intention	Type of test	Type of photo at test	Publication status	Encoding duration	No. of items at encoding	Retention interval	No. of test items	Test time
Randall et al. (2011)	H, F, D, R											
Unpublished		Y	Item	Incidental	IR	Different	Unpublished	10	32	10	64	10
Unpublished		Y	Item	Incidental	IR	Different	Unpublished	10	32	10	64	10
Unpublished		Y	Item	Incidental	IR	Different	Unpublished	10	32	10	64	10
Unpublished		Y	Item	Incidental	IR	Different	Unpublished	10	32	10	64	10
Rehnman & Herlitz (2006)	D											
		C	Self	Intentional	IR	Same	Published	3	20[f]	7	40[f]	10
		C	Self	Intentional	IR	Same	Published	3	20[f]	7	40[f]	10
		C	Self	Intentional	IR	Same	Published	3	20[f]	7	40[f]	10
		C	Self	Intentional	IR	Same	Published	3	20[f]	7	40[f]	10
Rehnman & Herlitz (2007)	D											
		Y	Self	Intentional	IR	Same	Published	3	30[g]	8	60[g]	5
		Y	Self	Intentional	IR	Same	Published	3	30[g]	8	60[g]	5
		Y	Self	Intentional	IR	Same	Published	3	30[g]	8	60[g]	5
		Y	Self	Intentional	IR	Same	Published	3	30[g]	8	60[g]	5
Rhodes, Castel, & Jacoby (2008)	H, F, D, R											
Experiment 1		Y	Self	Intentional	AR	Same	Published	4	20	0	30	SP
Experiment 1		O	Self	Intentional	AR	Same	Published	4	20	0	30	SP
Experiment 2		Y	Self	Intentional	AR	Same	Published	8,414[h]	20	0	30	SP
Experiment 2		O	Self	Intentional	AR	Same	Published	11,597[h]	20	0	30	SP
Experiment 3		Y	Self	Intentional	AR	Same	Published	4	16	0	48	SP
Experiment 3		O	Self	Intentional	AR	Same	Published	4	16	0	48	SP
Unpublished		Y	Elaborate	Intentional	AR	Same	Unpublished	4	20	0	30	SP
Unpublished		O	Elaborate	Intentional	AR	Same	Unpublished	4	20	0	30	SP
Unpublished		Y	Elaborate	Intentional	AR	Same	Unpublished	4	20	0	30	SP
Unpublished		O	Elaborate	Intentional	AR	Same	Unpublished	4	20	0	30	SP
Unpublished		Y	Self	Intentional	IR	Same	Unpublished	4	20	0	40	SP
Rodin (1987)	H											
Experiment 2		Y	Self	Incidental	IR	Same	Published	3[i]	30	25	60	SP
Experiment 2		O	Self	Incidental	IR	Same	Published	3[i]	30	25	60	SP
Experiment 3		Y	Self	Incidental	L	Different	Published	U	1	2	1	SP
Rose, Bull, & Vrij (2003)	H, F											
		Y	Self	Incidental	L	Different	Published	11	2	30	2	SP
		O	Self	Incidental	L	Different	Published	11	2	30	2	SP
Rose, Bull, & Vrij (2005)	H, F											
		Y	Self	Incidental	L	Different	Published	11	2	30	2	SP
		O	Self	Incidental	L	Different	Published	11	2	30	2	SP
Wiese, Schweinberger, & Hansen (2008)	H, F, D, R											
		Y	Self	Intentional	IR	Same	Published	5	20[j]	0.5	40[j]	2
		O	Self	Intentional	IR	Same	Published	5	20[j]	0.5	40[j]	2
Wilcock & Bull (2010)	H, F											
Experiment 2		O	Self	Incidental	L	Different	Published	6.5[k]	2	30	2	SP
Experiment 2		O	Self	Incidental	L	Different	Published	6.5[k]	2	30	2	SP

Table 1 (*continued*)

Study	Meta-analyses	Age group	Encoding	Intention	Type of test	Type of photo at test	Publication status	Encoding duration	No. of items at encoding	Retention interval	No. of test items	Test time
Wilcock, Bull, & Vrij (2005)	H, F											
		Y	Self	Incidental	L	Different	Published	6.5[k]	2	30	2	SP
		O	Self	Incidental	L	Different	Published	6.5[k]	2	30	2	SP
Wilcock, Bull, & Vrij (2007)	H, F											
		Y	Self	Incidental	L	Different	Published	6.5[k]	2	30	2	SP
		Y	Self	Incidental	L	Different	Published	6.5[k]	2	30	2	SP
		O	Self	Incidental	L	Different	Published	6.5[k]	2	30	2	SP
		O	Self	Incidental	L	Different	Published	6.5[k]	2	30	2	SP
C. I. Wright et al. (2008)	H, F											
		Y	Self	Unspecified	IR	Same	Published	0.5	16	4	40	SP
		O	Self	Unspecified	IR	Same	Published	0.5	16	4	40	SP
D. B. Wright & Stroud (2002)	H, F											
Experiment 1		Y	Self	Incidental	L	Different	Published	7[l]	4	1,440	4	SP
Experiment 1		Y	Self	Incidental	L	Different	Published	7[l]	4	10,080	4	SP
Experiment 1		M	Self	Incidental	L	Different	Published	7[l]	4	1,440	4	SP
Experiment 1		M	Self	Incidental	L	Different	Published	7[l]	4	10,080	4	SP
Experiment 2		Y	Self	Incidental	L	Different	Published	7[l]	4	1,440	4	SP
Experiment 2		M	Self	Incidental	L	Different	Published	7[l]	4	1,440	4	SP

Note. H = provided data for the meta-analysis of hits; F = provided data for the meta-analysis of false alarms; D = provided data for the meta-analysis of discriminability; R = provided data for the meta-analysis of response criterion; C = child; Y = young; M = middle; O = older adult; Self = self-directed; IR = item recognition; FC = forced-choice recognition; L = lineup; SP = self-paced; U = unspecified.

[a] Used a continuous task such that participants studied one face and were then tested on that face across a number of faces in succession (see footnote 4). [b] Participants studied and were tested in two blocks consisting of 15 study items and 30 test items (15 targets, 15 lures). [c] Participants studied and were tested in three blocks consisting of eight studied faces followed by eight test items. [d] The retention interval was 40–60 min with the value listed comprising the midpoint. [e] The data reported were collapsed across participants who studied 20 or 40 items and were tested on 40 or 80 items—data were coded at the midpoint of these ranges. [f] Participants studied and were tested in two blocks consisting of 20 study items and 40 test items (20 targets, 20 lures). [g] Participants studied and were tested in two blocks consisting of 30 study items and 60 test items (30 targets, 30 lures). [h] The encoding duration was self-paced and thus the values represent average encoding time. [i] Participants were given 90 s to study an array of 30 faces with the value listed represented the average time available per face. [j] Participants studied and were tested in six blocks consisting of 20 study items and 40 test items (20 targets, 20 lures). [k] Two individuals were seen in the encoding phase for 6 or 8 s—data are coded at the midpoint of this range. [l] Two individuals were seen in the encoding phase for 6 or 8 s—data are coded at the midpoint of this range.

face as "old" (previously studied) or "new" (not previously studied) and *forced-choice recognition* tests, which require the participant to identify a studied face amid one or more foils.[5] Studies that used a lineup as the criterion test were coded separately. Such studies typically involve observing a single individual in a long encoding phase (e.g., in a video simulating a crime), followed by a test in which the target may or not be present amid five to six lures (e.g., Rose, Bull, & Vrji, 2003; Rose et al., 2005). Given these distinct methodological differences, we separated performance on lineups from other types of tests. Finally, M. G. Rhodes et al. (2008) had participants study pairs of faces consisting of opposite sex couples of the same age or a different age. Participants were administered an *associative recognition* test, which required them to endorse pairs of faces that were studied and to reject pairs of faces that were not studied. Thus, the *type of test* was coded as a categorical variable with levels of *item recognition*, *forced choice*, *lineup*, and *associative recognition*.

Type of photograph at test. The target face presented at test may be either an alternate photograph of the same individual or the same photograph presented in the study phase. Most commonly, the alternate photograph involves the same individual in a different pose (e.g., Anastasi & Rhodes, 2005; Harrison & Hole, 2009; D. B. Wright & Stroud, 2002). Studies that use a different photograph of the same individual may reflect the purest test of the OAB because participants must recognize the individual studied rather than an idiosyncratic aspect of a previously studied picture (cf. Bruce & Young, 1986; Vokey & Read, 1992). Thus, *type of photograph at test* was coded as a categorical variable with levels of *same photograph*, *different photograph*, *combined*, or *unspecified*. The designation *combined* reflects one study (Bartlett & Leslie, 1986) that used a within-subjects manipulation of the same and different photographs of studied individuals.

Time at test. Most studies allowed participants to make recognition decisions at their own pace. Such *self-paced* studies can be contrasted with other studies (e.g., Anastasi & Rhodes, 2005; He et al., 2011) that gave participants a fixed amount of time, ranging from 2–10 s, to make recognition decisions. Because the majority of studies used a self-paced procedure, we coded *time at test* as a categorical rather than continuous variable with levels of *self-paced*, *5–10 s*, and *less than 5 s*.

Publication status. Each study was coded for its status as *published* or *unpublished*. Effect sizes derived from unpublished studies permit an examination of whether those experiments accepted for publication provide a different estimate of the true effect size than do unpublished studies.

Continuous variables. Several continuous variables pertaining to methodological characteristics of each study were coded. This included the presentation rate for each face (*encoding duration*), the number of faces presented prior to the test (*number of items at encoding*), the amount of time between the end of the encoding phase and the beginning of the test phase (*retention interval*), and the number of items presented at test (*number of test items*).

Effect-Size Calculation

All calculations of effect sizes and analyses were carried out using Comprehensive Meta-Analysis (Version 2.0; Borenstein, Hedges, Higgins, & Rothstein, 2005). In the meta-analyses reported, effect sizes were calculated as the difference between the mean value of the same-age and other-age conditions (for the dependent measure of interest) divided by the pooled standard deviations. Equation 1 can be used to calculate Cohen's d:

$$d = \frac{M_{SA} - M_{OA}}{\sqrt{\dfrac{(n_{SA} - 1)s_{SA}^2 + (n_{OA} - 1)s_{OA}^2}{n_{SA} + n_{OA} - 2}}} \tag{1}$$

where M_{SA} is the mean of the same-age condition (for the dependent measure of interest), M_{OA} is the mean of the other-age condition (for the dependent measure of interest), n_{SA} is the number of participants in same-age group, n_{OA} is the number of participants in the other-age group, s_{SA} is the standard deviation of the same-age group (for the dependent measure of interest), and s_{OA} is the standard deviation of the other-age group (for the dependent measure of interest). Cohen's d introduces a small bias that can lead to overestimates of effect size. This bias is eliminated by using the following correction factor, J:

$$J = 1 - (3/4df - 1), \tag{2}$$

where the degree of freedom is $n_{SA} + n_{OA} - 2$. The product of J and Cohen's d thus provides an unbiased estimate of the effect size that is often referred to as Hedges's g (Hedges, 1981):

$$g = J \times d. \tag{3}$$

For purposes of the present meta-analyses, all effect sizes were examined in terms of Hedges's g. Positive values of g indicate larger effect sizes for the same-age condition on the dependent measure of interest, and negative values of g indicate larger effect sizes for the other-age condition on the dependent measure of interest. When means and standard deviations (or standard errors) were not available, effect sizes were calculated using formulas that give d when provided with t, F, or a p value (see, e.g., Borenstein, 2009, for formulas).[6]

[5] Several studies used *continuous* (i.e., delayed-match-to-sample) item or forced-choice tests (e.g., Kuefner et al., 2008, 2010; Mondloch et al., 2006). In a continuous recognition test, participants typically study a face for a brief duration and are then immediately presented with a test face, requiring the participant to identify the face as studied or not (*item recognition*) or presented with two faces and asked to choose which face was studied (*forced-choice recognition*). This sequence of a study trial followed by a test trial continues until the stimulus set is exhausted. Given the paucity of such studies, we have collapsed continuous recognition tests into the general categories of *item recognition* and *forced-choice recognition*. As well, given that the retention interval, number of items at encoding, and number of items at test were examined as continuous variables, any variability in effect sizes evident when using a continuous procedure should be captured by these variables.

[6] Test statistics reported by researchers conducting studies using within-subjects manipulations of face age may inflate the resulting effect size (Dunlap, Cortina, Vaslow, & Burke, 1996) unless one corrects for the correlation between means. This was only necessary to calculate four effect sizes for discriminability from Bäckman (1991) and three effect sizes for hits from Rodin (1987). Although correlations were not reported for those studies, we obtained correlations between same-age and other-age faces for hits and discriminability for 14 studies comprising 49 different effect sizes and 1,630 participants. We then imputed the mean weighted correlations across these studies (r = .38 for hits and r = .27 for discriminability) when calculating effect sizes (see Borenstein, 2009, for formulas). Analyses that excluded these effect sizes resulted in no departures from the data reported.

All effect sizes were weighted as a function of the sample size. Specifically, weights (w_i) were calculated as the inverse of the conditional variance (v_i)—that is, $1/v_i$, which in turn is inversely proportional to the study sample size. The conditional variance is calculated as follows:

$$v_i = [(n_{SA} + n_{OA})/(n_{SA} \ldots n_{OA} + [g^2/2(n_{SA} + n_{OA})], \quad (4)$$

where n_{SA} is the sample size of the same-age condition, n_{OA} is the sample size of the other-age condition, and g is the effect size. Because weights are calculated as the inverse proportion of the conditional variance (i.e., $w_i = 1/v_i$), larger weights will be derived from studies with larger sample sizes. Thus, studies with larger sample sizes, which presumably provide a more accurate estimate of the population effect size, are given more weight than studies with smaller sample sizes (see Shadish & Haddock, 2009, for a more detailed discussion of these issues).

Random-effects models were used in the analyses reported (see the Method of Analysis sections for additional information on the logic of such models). Whereas fixed-effects models only account for variability reflected by sampling error (v_i), random-effects models also include between-studies variation (τ^2; see Hedges & Vevea, 1998, for details of this calculation). As such, the conditional variance for the random-effects models used, denoted v_i^*, is given by Equation 5 (adapted from Hedges & Vevea, 1998):

$$v_i^* = v_i + \tau^2, \quad (5)$$

where v_i is the within-studies sampling variation and τ^2 is the between-studies variation.

It should be noted that 30 of the effect sizes for the meta-analysis of hits and 36 of the effect sizes for the meta-analysis of false alarms were derived from studies that used lineups as the criterion test. Such studies produced dichotomous outcomes, as participants made a single choice as to whether a previously studied individual was present in a lineup (although several studies also used target-absent lineups whereby a previously studied individual was not present in the lineup). For example, consider a hypothetical study in which 100 participants see a video of an individual committing a crime, with 50 participants exposed to an individual of the same age in the video and 50 participants exposed to an individual of another age in the video. Suppose participants were then provided a lineup in which the individual in the video was presented along with five lures. One could measure hits by examining the number of participants who made a correct identification as well as the number of individuals making incorrect identifications. Table 2 provides hypothetical data for such a study. Columns, denoted X, refer to the number of participants making a correct compared with an incorrect identification. The rows, denoted Y, specify the age of the target face. Effect sizes for such dichotomous outcomes can be derived by calculating an odds ratio (OR), which is given by Equation 6 (adapted from Fleiss & Berlin, 2009):

$$OR = \frac{X_1Y_1 \times X_2Y_2}{X_1Y_2 \times X_2Y_1}. \quad (6)$$

Thus, for the hypothetical data provided in Table 2, the odds ratio would be calculated as $(28 \times 29)/(21 \times 22) = 1.76$. As such, this would indicate that the odds of identifying an individual from the same age group were 1.76 times greater than the odds of identifying an individual from another age group. The odds ratio is indeterminate when one of the cells is equal to 0 and thus, following prior recommendations (e.g., Gart & Zweifel, 1967), 0.5 was added to each cell before calculating an odds ratio (the adjusted odds ratio for the hypothetical data from Table 1 is 1.74). In turn, the odds ratio can be converted to Cohen's d by Equation 7 (see Borenstein, 2009):

$$d = \frac{ln(\text{o})\sqrt{3}}{\pi}, \quad (7)$$

where $ln(o)$ is the natural logarithm of the (adjusted) odds ratio and π is the mathematical constant (approximately 3.14159). For the hypothetical data provided in Table 2, d would be .305 (see Sánchez-Meca, Marín-Martínez, & Chacón Moscoso, 2003, for details on calculating the variance). Thus, each study using a lineup as the criterion test was converted to a d-based effect size and entered into the analyses reported.

Table 2
Hypothetical Data for a Lineup Experiment

	X		
Y	Correct identification	No correction identification	Total
Same-age face	28	22	50
Other-age face	21	29	50
Total	49	51	100

Method of Analysis

Following Hedges and Vevea (1998; see also Raudenbush, 2009), we used a random-effects model. Random-effects models assume that all treatments have been randomly sampled from a population of possible treatments, with the corollary that sampling error is calculated as variance relating to both random effects and estimation variance (Hedges & Vevea, 1998; see Equation 5). Effect-size estimates include both sources of variance and thus will be more conservative, with larger confidence intervals, than estimates from fixed-effects models, which only account for estimation variance. However, random-effects models have the advantage of permitting the meta-analyst to make generalizations to the universe of possible studies rather than inferences confined to the particular set of studies examined.

Analyses began by determining whether all effect sizes were homogeneous (Q_T). The Q_T statistic has an approximate chi-square distribution with $k - 1$ degrees of freedom, where k is the number of studies. A significant Q_T statistic indicates that there is significant unexplained heterogeneity among the effect sizes examined and likewise serves as a test of whether the between-studies variance component (τ^2) exceeds zero (Shadish & Haddock, 2009). The proportion of total variation in Q_T that is attributable to heterogeneity rather than chance is reported via the I^2 statistic (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003). Variation between levels of a categorical variable was tested using the between-groups fit statistic (Q_B). This test is analogous to the one-way analysis of variance omnibus F test for variation of group means, and a significant Q_B indicates that the average weighted effect size differed across groups.

The effect of continuous variables (e.g., study duration) was estimated by the weighted methods of moments, with weighting done in the same manner used to calculate effect sizes. The unstandardized regression coefficient derived (b_j) can be tested by dividing by the standard error of the estimated coefficient. In addition to this test, we also calculated the 95% confidence interval (CI). A statistically reliable coefficient indicates that there was a reliable relationship between the variable of interest and effect sizes.

Results and Discussion

Descriptive statistics for each measure (hits, false alarms, discriminability, and response criterion) from the studies examined are presented in Table 3. Given that different criterion tests have been used, these data are broken down by the type of test. In addition, although these data are unweighted, means weighted by sample size are essentially identical to those depicted in the table. As can be seen, there appears to be a mild advantage for recognition of same-age faces, evident in higher levels of hits, lower levels of false alarms, and superior discriminability, relative to other-age faces. In the following sections, we report meta-analyses of each of these dependent measures. Unless otherwise noted, the alpha level for all statistical tests was set to .05.

Meta-Analysis of Hits

Figure 1 displays a stem-and-leaf plot of mean weighted effect sizes for hits. Thirty-one (25.20%) of the effect sizes were negative, 86 (69.92%) were positive, and six (4.88%) were at zero. An examination of the figure indicates that the distribution is approximately normal with a slight positive skew (skewness = .19). We analyzed these data for outliers by classifying any effect size that was 3 standard deviations above or below the mean as an outlier. On the basis of this criterion, one outlier was present (g = 1.55). This effect size was included in the analyses to be reported; however, additional analyses were also conducted that excluded this outlier. Those analyses resulted in no substantive departures from the data reported.

The average weighted effect size for hits was significant (g = 0.23, 95% CI [0.17, 0.30]), indicating that hits were reliably higher for same-age compared with other-age faces, consistent with an OAB. In addition, effect sizes were clearly not homogeneous, as there was substantial heterogeneity (Q_T = 318.22, $p < .001$). The I^2 value (61.66%) indicates that well over half of the variation in effect sizes was due to heterogeneity.

Analysis of categorical variables. Table 4 presents a summary of main effect analyses of categorical variables for the analysis of effect sizes for hits, including the number of studies for each level of a moderator variable (N), the number of effect sizes examined (k), the mean weighted effect size (g), and the 95% CI. The test of Q_B (with degrees of freedom equivalent to the number of groups for a variable minus 1) was used to determine whether the levels of a categorically coded variable differed significantly from each other.

Analysis of the effect of the age of the participants or age group (children, younger adults, middle-aged adults, older adults), did not yield reliable differences between groups (Q_B = 0.92, p = .82). Whereas effect sizes for young and older adults reliably differed from zero, those for children and middle-aged adults did not.

The effect of type of encoding (self-directed, item specific, elaborative) did not yield reliable differences between groups (Q_B = 0.26, p = .88). Only effect sizes for elaborative encoding did not differ from zero.

The effect of intention (intentional encoding, incidental encoding, unspecified) did not yield reliable differences between groups (Q_B = 1.40, p = .50). Only effect sizes for studies using intentional and incidental encoding differed from zero.

Analysis of the effect of the type of test (item recognition, forced choice, lineup, associative recognition) indicated that effect sizes were marginally, but not reliably, different from each other (Q_B = 6.37, p = .095). This marginal difference may reflect the fact that effect sizes were smallest for lineups, which did not differ from zero. Specifically, if data for lineups are removed from the analysis, the Q_B statistic is markedly reduced (Q_B = 1.01, p = .65).

Analysis of the effect of the type of photograph at test (same photograph, different photograph, combined, not specified) did not yield reliable differences between groups (Q_B = 2.58, p = .46). All effect sizes reliably differed from zero with the exception of the combined condition (g = 0.17).

The effect of time at test (self-paced, 5—10 s, less than 5 s) did not yield reliable differences between groups (Q_B = 1.84, p = .40). All effect sizes were reliably greater than zero.

Finally, analysis of the effect of publication status (published, unpublished) indicated that effect sizes did not reliably differ for

Table 3
Descriptive Statistics for the Studies Examined by Measure and the Type of Criterion Test

	Hits				False alarms				d'				C'			
	SA face		OA face		SA face		OA face		SA face		OA face		SA face		OA face	
Type of test	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Item recognition	0.77	0.10	0.73	0.10	0.23	0.13	0.27	0.12	1.92	0.75	1.56	0.67	0.003	0.26	−0.001	0.38
Associative recognition	0.68	0.14	0.61	0.15	0.13	0.10	0.14	0.04	2.32	0.86	1.92	0.69	0.46	0.29	0.54	0.30
Forced-choice recognition	0.80	0.20	0.79	0.19	0.18	0.20	0.21	0.20								
Lineup	0.41	0.23	0.40	0.19	0.34	0.21	0.40	0.21								

Note. SA = same-age; OA = other-age; d' = discriminability; C' = response criterion.

Stem	Leaf
−1.0	
−0.9	4
−0.8	
−0.7	4
−0.6	
−0.5	7
−0.4	4, 6, 8
−0.3	0, 4, 6, 8
−0.2	1, 2, 3, 4, 6, 6, 7, 7, 8
−0.1	3, 7, 9, 9, 9
−0.0	1, 3, 7, 7, 8, 8, 9
0.0	0, 0, 0, 0, 0, 0, 1, 2, 3, 4, 6, 8, 8, 9, 9
0.1	1, 1, 1, 2, 2, 3, 5, 5, 9, 9
0.2	1, 2, 2, 2, 2, 2, 3, 4, 4, 6, 6, 7, 8, 8, 9, 9, 9
0.3	1, 3, 4, 5, 5, 6, 7, 7
0.4	0, 0, 0, 2, 5, 5, 6, 6, 7, 8, 8
0.5	1, 2, 3, 3, 3, 4, 4, 4, 4, 5, 7, 9
0.6	4, 4, 5, 8
0.7	2, 3, 5, 6, 6, 8, 9
0.8	3, 5, 5, 6
0.9	0
1.0	
1.1	2
1.2	
1.3	
1.4	2
1.5	5

Figure 1. Stem-and-leaf plot of effect sizes for hits. The first two digits of each effect size are listed in the stem column, with the third digit listed in the leaf column. Thus, the bottom entry of the figure reads as 1.55.

unpublished compared with published studies ($Q_B = 2.32$, $p = .13$).

Analysis of continuous variables. Table 5 presents a summary of the analyses of continuous variables for effect sizes for hits, including the coefficient, the standard error of this coefficient, the 95% CI of the coefficient, and the corresponding *Z* test of the coefficient. *Z* values that exceed 1.96 indicate that the predictor variable of interest is reliably related to effect sizes. As can be seen, none of the variables examined were reliably related to the magnitude of effect sizes.

Summary of the meta-analysis of hits. Overall, same-age faces were associated with a significantly greater level of hits than other-age faces ($g = 0.23$). There was a marginal difference in effect sizes as a function of the type of test, with lineups yielding the smallest effect sizes. No other variables appeared to reliably moderate effect sizes.

Meta-Analysis of False Alarms

Figure 2 displays a stem-and-leaf plot of mean weighted effect sizes for false alarms. Eighty-one (63.78%) of the effect sizes were negative, 39 (30.71%) were positive, and seven (5.51%) were at zero. An examination of the figure indicates that the distribution was generally normal with a slight negative skew (skewness = −.23). As well, no outliers were present.

The average weighted effect size for false alarms was significant ($g = -0.23$, 95% CI [−0.31, −0.16]). The direction of the effect size indicates that the level of false alarms was reliably greater for other-age compared with same-age faces, consistent with an OAB. In addition, effect sizes were not homogeneous ($Q_T = 463.54$, $p < .001$), with almost three quarters of the variation in effect sizes due to heterogeneity ($I^2 = 72.82\%$).

Analysis of categorical variables. Table 6 presents a summary of main effect analyses of categorical variables for the analysis of effect sizes for false alarms. The effect of age group (children, younger adults, middle-aged adults, older adults) was reliable ($Q_B = 26.93$), as the average weighted effect size differed on the basis of the age group tested. Only the mean weighted effect size for younger adults reliably differed from zero ($g = -0.41$), suggesting that only younger adults had reliably fewer false alarms to same-age relative to other-age faces. If the young adult data are removed, the remaining effect sizes no longer reliably differ ($Q_B = 0.57$, $p = .75$).

The effect of type of encoding (self-directed, item specific, elaborative) did not yield reliable differences between groups ($Q_B = 0.27$, $p = .88$). In contrast to the self-directed and item-specific conditions, effect sizes for the elaborative condition did not differ from zero ($g = -0.14$).

The effect of intention (intentional encoding, incidental encoding, unspecified) did not yield reliable differences between groups ($Q_B = 0.35$, $p = .84$). All effect sizes reliably differed from zero.

Analysis of the effect of the type of test (item recognition, forced choice, lineup, associative recognition) indicated that effect sizes did not reliably differ from each other ($Q_B = 2.92$, $p = .40$). Only effect sizes for lineups did not reliably differ from zero ($g = -0.16$, $p = .07$).

Analysis of the effect of the type of photograph at test (same photograph, different photograph, combined, not specified) indicated that effect sizes were not reliably different from each other ($Q_B = 3.51$, $p = .32$). All effect sizes reliably differed from zero with the

Table 4
Summary of Main Effect Analyses for Hits

Variable	N	k	g	95% confidence interval Lower	Upper	Q_{between}
Age group						
Children	8	17	0.15	−0.02	0.33	0.92
Younger adults	40	64	0.24	0.15	0.33	
Middle-aged adults	2	4	0.23	−0.10	0.57	
Older adults	26	38	0.25	0.13	0.37	
Type of encoding						
Self-directed	31	83	0.24	0.16	0.32	0.26
Item-specific	12	36	0.21	0.10	0.32	
Elaborative	1	4	0.25	−0.13	0.63	
Intention						
Intentional encoding	20	59	0.27	0.18	0.37	1.40
Incidental encoding	20	57	0.19	0.10	0.29	
Not specified	3	7	0.24	−0.01	0.50	
Type of test						
Item recognition	24	68	0.25	0.17	0.33	6.37[a]
Forced choice	7	15	0.36	0.17	0.55	
Line-up	12	30	0.07	−0.08	0.22	
Associative recognition	2	10	0.28	0.04	0.51	
Type of photograph at test						
Same photograph	19	53	0.24	0.14	0.34	2.58
Different photograph	22	58	0.21	0.11	0.30	
Combined	1	6	0.17	−0.10	0.45	
Not specified	3	6	0.44	0.16	0.72	
Time at test						
Self-paced	32	91	0.21	0.13	0.28	1.84
5–10 s	8	26	0.27	0.14	0.40	
Less than 5 s	3	6	0.38	0.11	0.65	
Publication status						
Published	38	108	0.21	0.14	0.28	2.32
Unpublished	5	15	0.36	0.18	0.54	

Note. N = number of studies; k = number of effect sizes; g = mean weighted effect size; Q_{between} = heterogeneity between conditions.
[a] Differences among effect sizes were marginally significant ($p < .10$).

exception of the combined condition ($g = -0.19$). Given that the majority of effect sizes were calculated for the same photograph and different photograph conditions, we also compared those two conditions separately. That analysis indicated that effect sizes did not differ when a different photograph compared with the same photograph was used at test ($Q_B = 1.47$, $p = .23$). Thus, the OAB did not vary on the basis of the type of photograph used at test.

The effect of time at test (self-paced, 5—10 s, less than 5 s) did not yield reliable differences between groups ($Q_B = 2.73$, $p = .26$). Only effect sizes for studies allotting less than 5 s at test did not differ from zero.

Finally, analysis of publication status (published, unpublished) indicated that effect sizes were reliably greater for unpublished compared with published studies ($Q_B = 8.55$).

Analysis of continuous variables. Table 7 presents a summary of the analyses of continuous variables for effect sizes for false alarms. These data show only a marginal effect of retention interval ($p = .06$). In particular, there was a positive relationship between effect sizes and retention interval, such that longer retention intervals were associated with positive effect sizes (i.e., a smaller or nonexistent OAB for false alarms). This may in part reflect the fact that studies using lineups as the criterion test

Table 5
Regression Coefficients for the Analysis of Hits

Predictor	Coefficient	SE	95% confidence interval Lower	Upper	Z
Encoding duration	.000001	.00003	−.00005	.00005	0.03
Number of items at encoding	.005	.002	−.003	.004	0.28
Retention interval	−.00002	.00001	−.00004	.00001	1.28
Number of test items	.008	.001	−.001	.002	0.78

Note. Z = test of the relationship between the predictor and effect sizes (values exceeding 1.96 indicate a reliable relationship).

Stem	Leaf
- 1.5	5
- 1.4	2, 5
- 1.3	
- 1.2	
- 1.1	0, 3, 3, 9
- 1.0	
- 0.9	0, 1, 2, 2
- 0.8	0, 2, 4, 4, 8
- 0.7	0, 0, 1, 2, 3, 4, 6, 6, 8
- 0.6	1, 3, 3, 6, 7, 8
- 0.5	0, 0, 2, 2, 2, 2, 3, 3, 3, 4, 7, 7, 7,
- 0.4	4, 4, 7, 7, 8
- 0.3	0, 0, 1, 3, 4, 4, 5, 6
- 0.2	0, 2, 3, 4, 4, 5, 6, 7, 9
- 0.1	0, 0, 2, 4, 4, 8
- 0.0	1, 2, 2, 2, 4, 6, 6, 8, 9
0.0	0, 0, 0, 0, 0, 0, 0, 3, 4, 6, 6, 8, 8, 9,
0.1	0, 0, 0, 1, 2, 2, 3, 4, 4, 7, 7, 8, 9
0.2	0, 3, 4, 6
0.3	0, 3, 5, 5, 7, 8
0.4	2, 2, 6, 6
0.5	4
0.6	6
0.7	4
0.8	2, 4
0.9	
1.0	

Figure 2. Stem-and-leaf plot of effect sizes for false alarms. The first two digits of each effect size are listed in the stem column, with the third digit listed in the leaf column. Thus, the bottom entry of the figure reads as 0.84.

yielded the smallest effect sizes for false alarms ($g = -0.09$) while also using longer retention intervals than many studies, including intervals of 1 week (e.g., Goodsell, Neuschatz, & Gronlund, 2009; Memon et al., 2003; D. B. Wright & Stroud, 2002). Indeed, when these data were reanalyzed after excluding those studies that used a lineup as the final criterion test, the length of the retention interval was unrelated to the magnitude of effect sizes ($Z = 0.97$, $p = .33$).

Summary of the meta-analysis of false alarms. Overall, false alarms were more likely for other-age compared with same-age faces ($g = -0.23$). Two variables reliably moderated the magnitude of effect sizes for false alarms. Specifically, younger adults exhibited lower levels of false alarms for same-age relative to other-age faces, whereas children and older adults showed no difference in false alarms on the basis of face age. As well, the published literature yielded smaller effect sizes than the unpublished literature and, among the continuous variables examined, only retention interval was marginally related to the magnitude of effect sizes. However, when studies using lineups were excluded from the analyses, the length of the retention interval was no longer related to the magnitude of effect sizes.

Meta-Analysis of Discriminability

Figure 3 displays a stem-and-leaf plot of mean weighted effect sizes for discriminability. Seventeen (19.54%) of the effect sizes were negative, 68 (78.16%) were positive, and two (2.30%) were at zero. An examination of the figure indicates that the distribution was generally normal with a moderate positive skew (skewness = .58). No outliers were present.

The average weighted effect size for discriminability was significant ($g = 0.37$, 95% CI [0.28, 0.45]). That is, participants exhibited reliably better discriminability for same-age compared with other-age faces, consistent with an OAB. In addition, effect sizes were not homogeneous ($Q_T = 297.84$, $p < .001$), and nearly three quarters of the variation in effect sizes was due to heterogeneity ($I^2 = 71.13\%$).

Analysis of categorical variables. Table 8 presents a summary of main effect analyses of categorical variables for the analysis of effect sizes for discriminability. As noted previously, experienced-based theories are best tested via measures of discriminability. Support for a narrow version of an experienced-based theory would come from finding that the OAB is evident across age groups. In contrast, asymmetry in the OAB across age groups (e.g., no OAB for older adults) would be consistent with the idea that more distant experiences drive the effect. Analysis of the effect of age group (children, younger adults, middle-aged adults, older adults) yielded marginal differences among effect sizes ($Q_B = 7.15$, $p = .07$). All effect sizes reliably differed from zero with the exception of those of middle-aged adults ($g = 0.81$, $p = .06$), for which there were few effect sizes. Thus, children, younger adults, and older adults exhibited effect sizes that differed from zero (i.e., an OAB). Such data, particularly the finding that older adults exhibited a reliable effect size despite a lifetime of prior experiences in other age groups, are consistent with the notion that recent experiences rather than distant experiences drive the OAB.

The effect of type of encoding (self-directed, item specific, elaborative) was not reliable ($Q_B = 2.69$, $p = .26$). Only effect sizes for the elaborative condition did not reliably differ from zero ($g = .25$, $p = .23$).

Table 6
Summary of Main Effect Analyses for False Alarms

Variable	N	k	g	95% confidence interval Lower	Upper	Q_{between}
Age group						
Children	8	17	−0.04	−0.22	0.15	26.93[a]
Younger adults	39	66	−0.41	−0.51	−0.31	
Middle-aged adults	2	4	−0.19	−0.58	0.20	
Older adults	26	40	−0.03	−0.16	0.10	
Type of encoding						
Self-directed	30	87	−0.24	−0.34	−0.15	0.27
Item specific	12	36	−0.22	−0.36	−0.08	
Elaborative	1	4	−0.14	−0.57	0.30	
Intention						
Intentional encoding	20	59	−0.24	−0.35	−0.13	0.35
Incidental encoding	19	61	−0.22	−0.33	−0.10	
Not specified	3	7	−0.31	−0.62	−0.003	
Type of test						
Item recognition	23	66	−0.25	−0.35	−0.15	2.92
Forced choice	7	15	−0.37	−0.60	−0.15	
Line-up	11	36	−0.16	−0.32	0.01	
Associative recognition	2	10	−0.13	−0.41	0.15	
Type of photograph at test						
Same photograph	18	51	−0.17	−0.29	−0.05	3.51
Different photograph	21	64	−0.27	−0.38	−0.16	
Combined	1	6	−0.19	−0.53	0.16	
Not specified	3	6	−0.46	−0.79	−0.14	
Time at test						
Self-paced	31	95	−0.24	−0.33	−0.15	2.73
5–10 s	8	26	−0.27	−0.43	−0.10	
Less than 5 s	3	6	0.03	−0.30	0.36	
Publication status						
Published	37	112	−0.19	−0.27	−0.11	8.55[a]
Unpublished	5	15	−0.53	−0.74	−0.32	

Note. N = number of studies; k = number of effect sizes; g = mean weighted effect size; Q_{between} = heterogeneity between conditions.
[a] Differences among effect sizes were statistically significant ($p < .05$).

The effect of intention (intentional encoding, incidental encoding, unspecified) did not yield reliable differences between groups ($Q_B = 3.75$, $p = .15$). However, when discriminability was analyzed comparing only incidental ($g = 0.46$) with intentional ($g = 0.29$) encoding instructions, a reliable difference was evident ($Q_B = 4.01$). All effect sizes reliably differed from zero.

Analysis of the effect of the type of test (item recognition, associative recognition) did not yield reliable differences between groups ($Q_B = 0.45$, $p = .50$). Each effect size was reliably greater than zero.

Analysis of the effect of the type of photograph at test (same photograph, different photograph, combined, not specified) indicated that effect sizes were marginally different from each other ($Q_B = 6.63$, $p = .08$). All effect sizes reliably differed from zero with the exception of the unspecified condition that did not identify the nature of the photographs used ($g = 0.26$). Follow-up analyses, limited to studies using the same or a different photograph, indicated that effect sizes were reliably larger when the test used a different photograph compared with the same photograph ($Q_B = 6.48$).

Table 7
Regression Coefficients for the Analysis of False Alarms

Predictor	Coefficient	SE	95% confidence interval Lower	Upper	Z
Encoding duration	.000001	.00003	−.00006	.00007	0.18
Number of items at encoding	.0002	.002	−.004	.005	0.08
Retention interval	.00003	.00001	−.000002	.00006	1.85[†]
Number of test items	−.0003	.001	−.002	.002	0.25

Note. Z = test of the relationship between the predictor and effect sizes (values exceeding 1.96 indicate a reliable relationship).
[†] $p < .10$.

Stem	Leaf
- 0.5	
- 0.4	
- 0.3	7
- 0.2	0, 0, 0, 5, 6, 7
- 0.1	1, 1, 2, 2, 5
- 0.0	2, 4, 4, 5, 8
0.0	0, 0, 3, 4, 6, 8, 9
0.1	0, 1, 1, 2, 2, 8, 8
0.2	2, 3, 4, 4, 4, 5, 6, 6, 7, 8, 9
0.3	0, 2, 3, 4, 4, 7, 8
0.4	4, 4, 6, 6, 7, 9
0.5	0, 1, 2, 4, 6, 6, 9
0.6	0, 4, 5, 7, 7, 7
0.7	1, 2, 4
0.8	1, 5, 8, 9
0.9	2, 4, 8
1.0	1, 2, 2
1.1	5, 9
1.2	7
1.3	7, 8
1.4	
1.5	9

Figure 3. Stem-and-leaf plot of effect sizes for discriminability (*d'*). The first two digits of each effect size are listed in the stem column, with the third digit listed in the leaf column. Thus, the bottom entry of the figure reads as 1.59.

The effect of time at test (self-paced, 5–10 s, less than 5 s) did not yield reliable differences between groups ($Q_B = .21$, $p = .90$). Only effect sizes for studies allotting less than 5 s at test were not reliably greater than zero.

Finally, analysis of publication status (published, unpublished) indicated that effect sizes were reliably greater for unpublished compared with published studies ($Q_B = 6.16$).

Analysis of continuous variables. Table 9 presents a summary of the analyses of continuous variables for effect sizes for discriminability. These data show that none of the continuous variables were reliably related to the magnitude of effect sizes.

Summary of the meta-analysis of discriminability. Overall, discriminability was reliably better for same-age compared with other-age faces ($g = 0.37$). Several moderator variables had an impact on the magnitude of effect sizes. Specifically, effect sizes marginally differed across age groups, incidental encoding instructions yielded larger effect sizes than intentional encoding instructions, and effect sizes were reliably larger when the final criterion test used different photographs. In addition, unpublished studies yielded reliably larger effect sizes than published studies. Most important, the OAB was evident in children, younger adults, and older adults. Such data indicating an OAB across age groups, particularly data for older adults, suggests that more recent experiences drive the OAB rather than more distant prior experiences. We return to this finding in the General Discussion section.

Meta-Analysis of Response Criterion

Figure 4 displays a stem-and-leaf plot of mean weighted effect sizes for response criterion. Thirty-five (51.47%) of the effect sizes were negative, 32 (47.06%) were positive, and one (1.47%) was at zero. An examination of the figure indicates that the distribution was generally normal with a moderate negative skew (skewness = −.32). No outliers were present.

As noted previously, criterion was examined via the measure C'. Given that higher values of C' denote a more conservative criterion, positive effect sizes are indicative of a more conservative criterion for same-age relative to other-age faces and negative effect sizes are indicative of a more liberal criterion for same-age faces relative to other-age faces. The average weighted effect size for response criterion did not differ from zero ($g = -0.01$, 95% CI [−0.07, 0.05]). Thus, measured response criterion did not differ for same- compared with other-age faces. In addition, effect sizes were not homogeneous ($Q_T = 237.27$, $p < .001$) and almost three quarters of the variation in effect sizes was due to heterogeneity ($I^2 = 71.76\%$).

Analysis of categorical variables. Table 10 presents a summary of main effect analyses of categorical variables for the analysis of effect sizes for response criterion. Analysis of the effect of age group (children, younger adults, middle-aged adults, older adults) was reliable ($Q_B = 26.54$). In particular, whereas older adults exhibited more liberal responding for same-age faces, younger adults exhibited more conservative responding for same-age faces. Effect sizes reliably differed from zero for those groups but did not differ from zero for children ($g = -0.12$, $p = .28$) or middle-aged adults ($g = 0.53$, $p = .12$), for which there were little data.

The effect of type of encoding (self-directed, item specific, elaborative) did not yield reliable differences between groups ($Q_B = 0.29$, $p = .87$). As well, effect sizes did not reliably differ from zero.

The effect of intention (intentional encoding, incidental encoding, unspecified) did not yield reliable differences between groups ($Q_B = 0.58$, $p = .75$). No effect size reliably differed from zero.

Analysis of the effect of the type of test (item recognition, associative recognition) indicated that effect sizes did not differ

Table 8
Summary of Main Effect Analyses for Discriminability

Variable	N	k	g	95% confidence interval Lower	Upper	$Q_{between}$
Age group						
Children	6	16	0.24	0.05	0.43	7.15[a]
Younger adults	25	43	0.46	0.35	0.58	
Middle-aged adults	1	1	0.81	−0.05	1.68	
Older adults	17	27	0.26	0.12	0.41	
Type of encoding						
Self-directed	15	49	0.32	0.20	0.43	2.69
Item specific	11	34	0.45	0.32	0.59	
Elaborative	1	4	0.25	−0.15	0.65	
Intention						
Intentional encoding	13	45	0.29	0.17	0.40	3.75
Incidental encoding	11	33	0.46	0.33	0.60	
Not specified	3	9	0.38	0.12	0.64	
Type of test						
Item recognition	26	77	0.38	0.29	0.46	0.45
Associative recognition	2	10	0.28	0.02	0.54	
Type of photograph at test						
Same photograph	16	51	0.28	0.17	0.39	6.63[a]
Different photograph	10	28	0.51	0.37	0.65	
Combined	1	6	0.39	0.04	0.74	
Not specified	1	2	0.26	−0.38	0.89	
Time at test						
Self-paced	13	46	0.36	0.24	0.48	0.21
5–10 s	11	35	0.38	0.25	0.51	
Less than 5 s	3	6	0.30	−0.02	0.61	
Publication status						
Published	24	74	0.32	0.23	0.41	6.16[b]
Unpublished	3	13	0.62	0.40	0.84	

Note. N = number of studies; k = number of effect sizes; g = mean weighted effect size; $Q_{between}$ = heterogeneity between conditions.
[a] Differences among effect sizes were marginally significant ($p < .10$). [b] Differences among effect sizes were statistically significant ($p < .05$).

($Q_B = 0.80$, $p = .37$). Effect sizes reliably differed from zero for both types of tests.

Analysis of the effect of the type of photograph at test (same photograph, different photograph, combined, not specified) indicated that effect sizes were not reliably different from each other ($Q_B = 1.24$, $p = .74$). None of the effect sizes reliably differed from zero.

The effect of time at test (self-paced, 5—10 s, less than 5 s) did not yield reliable differences between groups ($Q_B = .25$, $p = .88$). None of the effect sizes were reliably greater than zero.

Finally, analysis of publication status (published, unpublished) indicated that effect sizes did not differ for unpublished compared with published studies ($Q_B = 0.37$, $p = .54$).

Analysis of continuous variables. Table 11 presents a summary of the analyses of continuous variables for effect sizes for response criterion. These data show that none of the continuous variables were reliably related to the magnitude of effect sizes.

Summary of the meta-analysis of response criterion. Overall, effect sizes for response criterion did not differ for same-age compared with other-age faces ($g = -0.01$). However, given that there was significant heterogeneity among effect sizes, we analyzed potential moderators. Those analyses showed that only age group moderated the magnitude of effect sizes. Specifically, whereas younger adults exhibited more conservative responding to same-age faces, older adults exhibited more liberal responding to same-age faces. All other variables had no impact on effect sizes.

General Discussion

The meta-analyses reported examined the finding that memory is superior for individuals of one's own age group compared with individuals of another age group (i.e., the OAB). Results showed that the OAB is indeed a robust effect. That is, although there was no OAB for response criterion ($g = -0.01$), mean weighted effect sizes were reliably greater than zero for hits ($g = 0.23$), false alarms ($g = -0.23$), and discriminability ($g = 0.37$). Translated to an odds ratio, hits were approximately 1.55 times more likely for same-age than other-age faces and false alarms were approximately 1.55 times less likely for same-age than other faces. As well, discriminability was approximately 1.99 times greater for same-age compared with other-age faces. Within Cohen's (1988) nomenclature, these effect sizes can be characterized as small (hits, false alarms) to moderate (discriminability) in magnitude. Thus, the interaction of the age of the rememberer and the age of the target influences recognition memory for faces.

Moderators of the OAB

Several variables reliably moderated the magnitude of effect sizes, whereas others had no impact on effect sizes. The type of test resulted in marginal differences in effect sizes for hits, likely because studies with lineups yielded the smallest effect sizes. The reason for this is not readily apparent, but the vast methodological

Table 9
Regression Coefficients for the Analysis of Discriminability

Predictor	Coefficient	*SE*	95% confidence interval Lower	95% confidence interval Upper	*Z*
Encoding duration	−.00002	.00003	−.00008	.00004	0.65
Number of items at encoding	.002	.003	−.004	.009	0.65
Retention interval	−.00001	.00003	−.00006	.00004	0.30
Number of test items	.002	.002	−.002	.006	1.02

Note. Z = test of the relationship between the predictor and effect sizes (values exceeding 1.96 indicate a reliable relationship).

differences between studies using lineups and studies that present a succession of faces may account for the differences obtained. The amount of time allotted for a decision at test had no impact on the magnitude of effect sizes; neither did intention, with the exception of effect sizes for discriminability. The type of encoding also had no impact on effect sizes, suggesting that the particular manner that participants study faces did not impact the magnitude of effect sizes. This may indicate that the OAB is relatively automatic and therefore largely insensitive to the manner of encoding. However, the range of encoding manipulations has not been exhausted (see the Future Directions section) so we regard these conclusions as tentative.

Effect sizes were reliably larger for discriminability when a different photograph from the encoding phase was used at test. This may suggest that using a different photograph represents a purer test of the OAB. That is, when identical photographs are used at study and test, recognition may be driven by memory for the identity of the individual or by idiosyncratic features of the photograph (e.g., lighting, shadowing) that are independent of identity. Thus, using a different photograph necessitates that the rememberer rely on identity-specific information rather than information specific to the photograph. Finally, among the continuous variables examined, only the duration of the retention interval was related to effect sizes for false alarms, with longer intervals associated with a positive effect size. Given that a negative effect size (i.e., more false alarms for other-age than same-age faces) is indicative of an OAB, such data suggest that the OAB was weaker or nonexistent at longer interval, although a follow-up analysis indicated that this was largely confined to studies using lineups. However, of greatest importance are those data that bear on theoretical accounts of the OAB, discussed in the sections that follow.

An Experience-Based Account of the OAB

As noted previously, the predominant theoretical accounts of the OAB suggest that it reflects more extensive experience with members of the same age group relative to other age groups. Such experience facilitates perceptual expertise for same-age relative to other-age faces, enhancing recognition memory for same-age faces. Evidence for an experienced-based account comes from several sources. For example, self-reported contact with other age groups diminishes the magnitude of the OAB and predicts recog-

Stem	Leaf
- 1.0	
- 0.9	2, 7
- 0.8	
- 0.7	0, 1, 4
- 0.6	
- 0.5	2, 4, 5
- 0.4	1, 4, 7
- 0.3	1, 7, 8, 9
- 0.2	0, 1, 2, 7, 8, 9
- 0.1	1, 1, 1, 6, 6, 8, 8, 9
- 0.0	1, 2, 3, 5, 5, 9
0.0	0, 2, 3, 6, 7, 9
0.1	1, 2, 4
0.2	0, 5, 5, 5, 7, 7
0.3	3, 4, 4, 4, 6, 7, 9
0.4	1, 3, 3, 3, 5, 9, 9
0.5	3, 7
0.6	6
0.7	8
0.8	
0.9	
1.0	

Figure 4. Stem-and-leaf plot of effect sizes for response criterion. The first two digits of each effect size are listed in the stem column, with the third digit listed in the leaf column. Thus, the bottom entry of the figure reads as 0.78.

Table 10
Summary of Main Effect Analyses for Response Criterion

Variable	*N*	*k*	*g*	95% confidence interval Lower	Upper	$Q_{between}$
Age group						
Children	4	7	−0.12	−0.34	0.10	26.54[a]
Younger Adults	21	36	0.16	0.06	0.26	
Middle-aged adults	1	1	0.53	−0.14	1.19	
Older adults	16	24	−0.24	−0.37	−0.11	
Type of encoding						
Self-directed	11	30	−0.02	−0.15	0.11	0.29
Item specific	11	34	0.01	−0.11	0.14	
Elaborative	1	4	−0.09	−0.49	0.31	
Intention						
Intentional encoding	9	30	−0.05	−0.18	0.09	0.58
Incidental encoding	11	33	0.02	−0.11	0.15	
Not specified	2	5	0.04	−0.27	0.35	
Type of test						
Item recognition	21	58	0.01	−0.09	0.10	0.80
Associative recognition	2	10	−0.11	−0.35	0.13	
Type of photograph at test						
Same photograph	11	32	−0.06	−0.19	0.07	1.24
Different photograph	10	28	0.03	−0.10	0.16	
Combined	1	6	0.02	−0.31	0.35	
Not specified	1	2	0.13	−0.38	0.64	
Time at test						
Self-paced	11	36	−0.01	−0.13	0.11	0.25
5–10 s	8	26	0.01	−0.13	0.16	
Less than 5 s	3	6	−0.07	−0.36	0.22	
Publication status						
Published	19	55	−0.02	−0.12	0.08	0.37
Unpublished	3	13	0.05	−0.15	0.26	

Note. *N* = number of studies; *k* = number of effect sizes; *g* = mean weighted effect size; $Q_{between}$ = heterogeneity between conditions.
[a] Differences among effect sizes were statistically significant ($p < .05$).

nition of other age faces (Chance et al., 1986; Ebner & Johnson, 2009; He et al., 2011). As well, individuals who have had extensive contact with another age group exhibit enhanced memory for that age group relative to individuals who have not had such extensive contact (Cassia, Kuefner, et al., 2009; Cassia, Picozzi, et al., 2009; Harrison & Hole, 2009; Kuefner et al., 2008, 2010). Thus, experience appears to play a significant role in memory for same-age compared with other-age faces.

We examined two variants of an experience-based account. One account suggests that prior experience as a member of multiple age groups may provide sufficient expertise to drive recognition throughout a lifetime (cf. Cassia, Kuefner, et al., 2009). For example, an older adult was once a child, younger adult, and middle-aged adult and may have acquired sufficient expertise with these groups to facilitate recognition of other-age faces. Thus, age groups with opportunities to develop expertise with other-age faces may be less likely to exhibit the OAB, consistent with several reports of younger adults but not older adults demonstrating the OAB (e.g., Bartlett & Leslie, 1986; Fulton & Bartlett, 1991; Havard & Memon, 2009; M. G. Rhodes et al., 2008; Wiese, Schweinberger, & Hansen, 2008). An alternative experience-based account suggests that only the most recent experiences contribute to the OAB (Hills & Lewis, 2011; cf. O'Neil & Webster, 2011). Based on the premise that the most frequent, recent experiences will be with members from one's own age group, this would suggest that the OAB would be prevalent for all age groups (but

Table 11
Regression Coefficients for the Analysis of Response Criterion

Predictor	Coefficient	*SE*	95% confidence interval Lower	Upper	*Z*
Encoding duration	−.00002	.00003	−.00007	.00004	0.61
Number of items at encoding	.001	.004	−.006	.008	0.30
Retention interval	−.00002	.00002	−.00006	.00003	0.70
Number of test items	.008	.002	−.003	.005	0.40

Note. *Z* = test of the relationship between the predictor and effect sizes (values exceeding 1.96 indicate a reliable relationship).

see Cassia, in press, for a different version of this account). We tested this via a meta-analysis of discriminability with those data showing that the OAB was present for children, younger adults, and older adults. Data for older adults are most illuminating because, unlike children and younger adults, they have had a lifetime of experience in other age groups. Given that older adults still exhibited an OAB, we conclude that more recent experiences drive the OAB.

What role would experience have in influencing face recognition? One possibility is that experience might change the manner in which faces are processed during encoding (but see Valentine, 1991, for a different account of the influence of experience). For example, prior experience may increase the probability that one focuses on configural information (e.g., Gauthier & Tarr, 1997; see Maurer, Le Grand, & Mondloch, 2002, for a review of the nature of configural information in face recognition) during encoding, such as the spatial relations between the features of the face (e.g., the distance between the nose and mouth) or the way that features of the face combine to create the face as a whole. In contrast, less experience may lead to a focus on component features of the face that are not diagnostic at retrieval. A prediction from this perspective is that manipulations that hinder configural encoding should disproportionately disrupt memory for ingroup faces, which should be associated with more extensive experience, than outgroup faces (e.g., Tanaka, Kiefer, & Bukach, 2004).

One common method of disrupting configural encoding is to invert faces. Prior work suggests that inversion disrupts memory for faces more than it does for other objects (Diamond & Carey, 1986; Yin, 1969), likely because inversion impedes access to configural information in the face. As well, several studies have shown that inversion disrupts memory for own-race faces more than it does for memory for other-race faces (e.g., G. Rhodes et al., 1989) and that the disproportionate inversion effect for own-race faces may be related to the amount of contact with another race (e.g., Hancock & Rhodes, 2008). A similar logic has been applied to the OAB. That is, if faces from one's own age group are more likely to promote configural encoding, then the decrement in performance wrought by inversion should be stronger for same-age compared with other-age faces. Initial investigations were inconsistent with this prediction (Mondloch, Maurer, & Ahola, 2006; Perfect & Moon, 2005). For example, Perfect and Moon (2005) tested older and younger adults' memory for upright and inverted faces of same-age and other-age individuals. They reported a robust OAB but no impact of inversion on the OAB.

However, several recent studies have examined the inversion effect for same-age and other-age faces and its interaction with experience. For example, Kuefner et al. (2008) tested adults' memory for upright and inverted faces of adults and children. An inversion effect was apparent for both groups of faces but was of a smaller magnitude for faces of children, providing partial support for the idea that same-age faces are more likely than other-age faces to engender configural encoding. In another experiment, Kuefner et al. used the same procedure but tested preschool teachers with an average of approximately 16 years of experience with children. Results showed that preschool teachers exhibited a robust inversion effect for same-age and other-age faces, consistent with the notion that experience with children's faces was associated with stronger configural processing of such faces and no OAB. Similarly, Cassia, Picozzi, et al. (2009) tested maternity ward nurses (mean experience of 16 years) and age-matched controls' memory for pictures of upright and inverted faces of newborns and adults. Whereas controls exhibited an inversion effect for adult but not infant faces, maternity ward nurses exhibited an inversion effect for both groups of faces and a smaller OAB.

Another method of disrupting configural processing of faces is reflected in work on the *face composite effect* (Young, Hellawell, & Hay, 1987). The face composite effect is the finding that participants perceive identical top halves of a face as different when they are coupled with different bottom halves. The illusion disappears if the bottom halves are misaligned (e.g., offset to the right) with the top half of the face. As such, the face composite effect is thought to reflect the fact that individual features are "glued" together into a gestalt that represents the whole face (i.e., holistic processing). Prior work suggests that the face composite effect may be stronger for own-race relative to other-race faces (e.g., Michel, Rossion, Han, Chung, & Caldara, 2006) or for any face deemed part of an ingroup (Hugenberg & Corneille, 2009). Two studies have reported that the composite illusion is stronger for same-age compared with other-age faces (de Heering & Rossion, 2008; Kuefner et al., 2010). However, the magnitude of the effect is moderated by experience. For example, de Heering and Rossion (2008) reported that, relative to age-matched controls, preschool teachers with at least one year of experience exhibited a similar composite effect for faces of children and adults. Thus, experience may facilitate holistic processing of faces.

In sum, several lines of data are compatible with an experienced-based account suggesting that perceptual learning of ingroup faces drives the OAB. Specifically, self-reported contact is related to the magnitude of the OAB (e.g., Chance et al., 1986; He et al., 2011), such that individuals with extensive experience with another age group exhibit superior memory for that age group relative to controls (e.g., Harrison & Hole, 2009) and are more susceptible to manipulations that disrupt configural or holistic processing (e.g., Cassia, Kuefner, et al., 2009; Cassia, Picozzi, et al., 2009; de Heering & Rossion, 2008; Kuefner et al., 2008, 2010). Further, data from the meta-analysis of discriminability, particularly for older adults, suggests that the OAB reflects more recent experiences with one's current age group. However, there are alternative theoretical perspectives on the processes that drive the OAB, which we turn to next.

Social–Cognitive Theories of the OAB

Social–cognitive perspectives on ingroup/outgroup face recognition (e.g., Hugenberg, Young, Bernstein, & Sacco, 2010; Sporer, 2001) posit that superior recognition of ingroup faces results from an initial categorization of a face as belonging to an ingroup. A product of categorization as an ingroup face is that the perceiver engages in individuating encoding that facilitates recognition. In contrast, outgroup faces may be rapidly categorized on the dimension that denotes the outgroup (e.g., age, race, gender; see, e.g., Levin, 1996, 2000) with subsequent encoding focusing on this category-relevant information at the expense of individuating information that supports recognition. In the context of the OAB, participants might attend to other-age faces only in regard to information that identifies the face as being of a particular age, whereas same-age faces might engender greater focus on information that makes the face unique within that category. As a result, same-age faces would be better remembered than other-age faces.

Several lines of evidence indicate that perceivers attend to age. For example, ERP data suggest that age information might comprise one aspect of precognitive processing of the face (Mouchetant-Rostaing & Giard, 2003; see also Wiese, Schweinberger, & Neuman, 2008). In addition, individuals rapidly and accurately categorize age (e.g., Johnston et al., 1997) and can make accurate judgments of age (see M. G. Rhodes, 2009, for a review) even after very brief presentations (e.g., Bruyer et al., 1991, 2007). Thus, it appears that individuals attend to age in a manner that might be largely obligatory.

Other research has examined explicit and implicit attitudes in younger and older adults toward ingroup and outgroup individuals. For example, Chasteen (2005) reported that younger adults perceived less similarity between their own age group and older adults than did a group of older adults tested (see also Ebner, Gluth, et al., 2011). As well, C. I. Wright et al. (2008) had older and younger adults make valence ratings, assessing the degree to which older and younger faces were rated as positive or negative. Younger adults rated older faces as significantly more negative than young faces, whereas older adults exhibited no difference in valence ratings for young and old faces (see also Ebner, He, Fichtenholtz, et al., 2011). These asymmetries in ratings differ from some other reports that both older and younger adults view older adults more negatively than they do younger adults on some explicit (Ebner, 2008; Gluth, Ebner, & Schmiedek, 2010; see Kite, Stockdale, Whitley, & Johnson, 2005, for a review) and implicit (He et al., 2011) measures. Nonetheless, prior work suggests that older and younger adults differ in their attitudes toward same-age and other-age individuals.

What influence do attitudes and ingroup/outgroup categorization have on the OAB? Despite a raft of speculation (e.g., Anastasi & Rhodes, 2005, 2006; Perfect & Moon, 2005; Rehnman & Hertlitz, 2006; Rodin, 1987), direct evidence on the matter is sparse. He et al. (2011) examined the relation between explicit and implicit attitudes toward aging and memory for same-age and other-age faces in a group of older and younger adults. They observed no reliable relationship between such indices of attitudes and the OAB. Likewise, as noted previously, C. I. Wright et al. (2008) collected valence ratings for younger and older faces and reported that younger adults rated old faces as being more negative than younger faces. However, an OAB in recognition memory remained even when controlling for valence. Finally, Lindholm (2005) examined younger and older adults' memory for characteristics (e.g., gender, age, hair color, attire) of same-age and other-age targets. All participants rated how rewarding it would be to meet the target person. Although both groups rated meetings with same-age individuals as more rewarding, this measure of interest was unrelated to memory performance. Such findings echo other unsuccessful attempts in the own-race bias literature to link memory performance to attitudes toward same- and other-race individuals (e.g., Platz & Hosch, 1988; but see Cross, Cross, & Daly, 1971).

A different approach has been to record eye movements during encoding as a measure of interest and attention to same- and other-age targets (e.g., Firestone et al., 2007; He et al., 2011; see also Ebner, He, & Johnson, 2011; Slessor, Laird, Phillips, Bull, & Filippou, 2010). For example, Firestone et al. (2007) tested younger and older adults' memory for same-age and other-age faces and reported different patterns of eye movements. Specifically, older adults spent more time sampling faces of younger adults, whereas younger adults sampled same-age and other-age faces equally. However, only older adults exhibited an OAB. Firestone et al. (2007) suggested that older adults might already encode same-age faces efficiently and thus did not need to sample such older faces as often as younger faces. He et al. (2011) reported that younger and older adults gazed longer at same-age relative to other-age faces during encoding (cf. Goldinger, He, & Papesh, 2009). In addition, gaze time was reliably related to the magnitude of the OAB.

Finally, participants' subjective impressions of age may play a role in the OAB. Consistent with this, Anastasi and Rhodes (2006, Experiment 2) had young and old participants study photographs of younger adults, older adults, and middle-aged individuals. During encoding, participants classified each individual as being a young adult, middle-aged adult, or older adult. Younger adults did not exhibit an OAB on the basis of chronological age but did exhibit an OAB when photographs were classified on the basis of their subjective age estimates. Thus, consistent with a social–cognitive account, the OAB was only present for photographs perceived to be of individuals from another age group. However, we note that this represents only a single experiment and remains to be replicated.

In all, the lack of data and the inconsistent patterns reported thus far make it unclear whether individuals disregard (Rodin, 1987) other-age faces at the expense of same-age faces. As well, theories contingent on ingroup/outgroup categorization are difficult to disentangle from perceptual expertise accounts. For example, extensive prior experience with a particular age group might influence gaze time rather than attitudes per se. Alternatively, categorization may beget differences in strategies that are themselves a function of prior expertise (Hugenberg et al., 2010; Sporer, 2001). From this perspective, a comprehensive theory of the OAB may require an integration of perceptual learning and social–cognitive accounts. We describe such an account in the section that follows.

An Integration of Experience-Based and Social–Cognitive Theories of the OAB

The data reviewed thus far provide compelling evidence for a perceptual learning account of the OAB. For example, prior experience with faces of another age group facilitates recognition of other-age faces (e.g., Harrison & Hole, 2009) and experience with other-age faces predicts sensitivity to manipulations that are diagnostic of configural processing (e.g., Cassia, Kuefner, et al., 2009; Cassia, Picozzi, et al., 2009; Kuefner et al., 2008, 2010). Direct evidence for a social-categorization account appears scant, although some data indicate that participants gaze longer at same-age compared with other-age faces (Ebner, He, Fichtenholtz, et al., 2011; He et al., 2011). However, we suggest that a categorization account has not been fully tested and that the most complete theory of the OAB will likely require elements of perceptual learning theories and those theories that suggest recognition is contingent on categorization. Thus, Hugenberg et al.'s (2010) *categorization–individuation model* (*CIM*), developed as an explanation of the own-race bias, may provide a viable explanation of the OAB (see also Sporer, 2001).

Briefly, the CIM suggests that, when processing a face, individuals engage in categorization or individuation. When engaging in categorization, the rememberer is assumed to encode faces with respect to dimensions indicative of the category, a processing strategy that hinders later discrimination of that individual. In contrast, when engaging in individuation, the rememberer is assumed to attend to "facial characteristics that are identity diagnostic, rather than to characteristics that are category diagnostic" (Hugenberg et al., 2010, p. 1170). Such individuation will facilitate recognition, whereas categorization, which entails attending solely to information emblematic of the category, will hinder recognition. With respect to the OAB, we note two important determinants of processes at encoding proposed by the CIM. First, situational cues can have a substantial impact on whether participants engage in categorization or individuation. For example, although a younger college student might not typically attend to an older adult, he or she might be more likely to attend to such an individual if the older adult was a professor (Hugenberg et al., 2010) or coworker. Thus, faces elicit different levels of motivation to attend to individuating information, leading to differences in recognition. Second, rememberers differ in their expertise in individuating groups of faces, with more extensive experience enhancing the ability to extract individuating information. However, rather than simply suggesting that memory for ingroup faces will be positively related to expertise, Hugenberg et al. (2010) noted that

> individuation experience is only fully employed for faces that perceivers are motivated to individuate. Thus, in many cases, perceivers may not fully deploy the individuation capacities that they have previously accrued. From this perspective, individuation experience does not translate automatically into strong face encoding and memory. Rather, it is only when a target's identity is seen as sufficiently important that perceivers attend to the identity-diagnostic characteristics of those faces. (p. 1171)

Thus, expertise will be most relevant when perceivers are motivated to extract identity-specific information.

Among the strongest lines of evidence for the CIM are differences in memory for perceptually ambiguous faces. For example, Bernstein et al. (2007) had participants study faces presented on a red or green background, with the red background denoting individuals putatively attending the same university and the green background denoting individuals attending a rival university. Results showed that recognition was superior for individuals depicted as being from the same university than from a rival university despite the fact that, across participants, there were no differences in target faces. From the perspective of the CIM, although the faces were perceptually ambiguous, participants were motivated to seek individuating information from own-university faces, facilitating subsequent recognition (see also Hugenberg & Corneille, 2009). Additional work suggests that seeking individuating information may also eliminate the own-race bias (e.g., Hehman et al., 2010).

Testing such a model in regard to the OAB would require that memory for same-age and other-age targets be examined under conditions that encourage individuation or categorization. For example, one might test younger adults' memory for older adults by labeling some faces as those of professors at their institution and labeling other faces as those of older adults from the community. Likewise, memory for children's faces could be tested by designating the faces as those from one's own school or from a nationally representative sample. These examples represent only a modicum of those that could be potentially examined, but such investigations are necessary for a true test of social–cognitive explanations of the OAB. For the present, on the basis of the evidence available and the meta-analyses reported, we conclude that a perceptual learning theory contingent on recent experiences provides the best account of the OAB. However, a complete theory will likely integrate social–cognitive mechanisms, such as those posited by the CIM. This may not only provide a viable account of recognition memory for same-age and other-age faces but also bear on accounts of age estimation (M. G. Rhodes, 2009), emotion recognition (e.g., Ebner & Johnson, 2009), and identification of other features of same-age and other-age faces.

Limitations of the Meta-Analyses Reported

The goal for any meta-analysis is to provide an accurate, unbiased estimate of the true effect size across a range of well-defined studies. It is always possible that a population of unpublished studies with relevant data has gone unreported, perhaps because a manipulation did not produce a result investigators felt was suitable for publication (i.e., the "file drawer problem"; Rosenthal, 1979). This not only leads to an overestimate of effect sizes but reduces the variability in reported effect sizes, yielding small confidence intervals around a biased mean.

Two methods were applied in the meta-analyses reported to address the file-drawer problem. First, significant efforts were made to obtain unpublished data. This yielded 15 effect sizes for hits and false alarms from unpublished sources and 13 effect sizes for discriminability and response criterion from unpublished sources. With the exception of response criterion, effect sizes were larger for unpublished compared with published studies. Thus, the effect sizes reported from published sources may be somewhat more conservative than the true effect size. Second, we calculated Rosenthal's (1979) fail-safe *N*, providing an estimate of the number of unpublished null results that would need to exist before one would conclude that overall results were due to sampling error (see Rosenthal, 1979, for details of the calculation, but see Sutton, 2009, for alternatives). Rosenthal recommended that the number of unpublished studies that may exist that would result in a null effect is unlikely if the fail-safe *N* exceeds $5k + 10$, where k is the number of effect sizes. For the meta-analysis of hits, the fail-safe *N* is 3,952 and far exceeds Rosenthal's tolerance level given 123 effect sizes (i.e., $5k + 10 = 625$). For the meta-analysis of false alarms, the fail-safe *N* is approximately 3,774 and is well above Rosenthal's tolerance level with 127 effect sizes (i.e., $5k + 10 = 645$). For the meta-analysis of discriminability, the fail-safe *N* is approximately 5,228 and is above Rosenthal's tolerance level with 87 effect sizes (i.e., $5k + 10 = 445$). Thus, although relatively small, the effect sizes evident for the meta-analyses of hits, false alarms, and discriminability are unlikely to change as a result of the unpublished literature.

Aside from potential file-drawer problems, the current meta-analysis could only consider the range of variables examined thus far, limiting conclusions about moderator variables drawn from a small number of studies. For example, any conclusions about middle-aged adults and, to some extent, children, are necessarily tentative given that few studies have examined those age groups.

Thus, conclusions drawn from moderators with few effect sizes should be treated with caution, particularly as random-effects models are poorly suited for coping with small numbers of effect sizes (Hedges & Vevea, 1998). We hope that this observation will spur continued investigation of the generality of the OAB across a variety of conditions.

Future Directions

A number of open questions remain with respect to the OAB. For example, as noted previously, a true test of social–cognitive explanations will require manipulations of situational cues that affect perceivers' motivation to individuate same-age and other-age faces. Evidence for a social–cognitive account would come from showing that recognition of other-age faces is facilitated when perceivers are motivated to individuate such faces or to perceive them as part of a larger ingroup (cf. Pauker et al., 2009). Such experiments would serve to link the OAB and other ingroup face recognition advantages such as those related to race (Malpass & Kravitz, 1969) and gender (D. B. Wright & Sladden, 2003). That is, although changes in age also change ingroup membership, making age distinct from other ingroup differences in recognition memory, a social–cognitive account would suggest that age is simply one of a number of dimensions that is used as a basis for classifying an individual as a member of an ingroup or outgroup. At present, such an experiment has yet to be conducted, but it would be a valuable addition to the literature.

As well, much remains to be understood about the impact of experience on recognition of same-age and other-age faces. In particular, few studies (e.g., He et al., 2011) have collected measures of experience, and future work may benefit from refining such measures. For example, older adults living within retirement communities may exhibit different patterns of data than would older adults living in other settings. Thus, further understanding how qualitatively different types of experiences contribute to face recognition will be valuable. In addition, although compelling, all prior work examining the influence of experience (e.g., Harrison & Hole, 2009; Kuefner et al., 2008) has come from testing populations that differ in the level of expertise with same-age and other-age faces. Other work in the own-race bias literature has manipulated the level of contact and/or experience with other-race faces by attempting to train recognition through continuous exposure to other-race faces. For example, Elliot, Wills, and Goldstein (1973) had White participants practice pairing a face of an Asian individual with a number; an associative memory task was then used such that participants were given a face and had to recall the number paired. Performance on a subsequent recognition task showed that participants who practiced learning faces of Asian individuals were better at recognizing novel Asian faces than were members of a control group given no training or training with faces of White individuals. Similar results have been reported by others (Lavrakas, Buri, & Mayzner, 1976; Malpass, Lavigueur, & Weldon, 1973), although Lavrakas et al. (1976) observed that the benefits of training on recognition of other-race faces were not evident after 1 week. Nonetheless, the OAB literature might profit from examining experienced-based accounts using a similar strategy, with participants given training on other-age faces.

Manipulations that attenuate the own-race bias and may also attenuate the OAB have been given little attention. For example, Johnson and Fredrickson (2005) examined the impact of emotion on the own-race bias. In particular, prior to studying photographs of Black and White individuals, a group of White participants viewed videos that induced emotions of joy or fear (a control group watched a neutral video). The results showed that inducing joy entirely eliminated the own-race bias, a finding that was replicated even when emotion induction was done after encoding. Johnson and Fredrickson (2005) suggested that positive emotions led participants to more broadly construe which faces made up an ingroup. Little work has likewise examined the impact of emotion on the OAB and whether such a manipulation might alter which faces are deemed part of an ingroup (but see Ebner & Johnson, 2009, for work with same- and other-age faces displaying different emotions). As well, the metacognitive and phenomenological experience of same-age and other-age faces remains largely unexplored. Some work suggests that memory for own-race faces is characterized by more recollective details (e.g., Marcon, Susa, & Meissner, 2009), and participants are indeed more likely to report retrieving contextual information when remembering own-race faces (Horry, Wright, & Tredoux, 2010; Meissner, Brigham, & Butz, 2005). It is unknown whether similar patterns would be evident for same-age faces. Such information will be crucial as researchers seek to understand those factors that influence memory for same-age and other-age individuals.

Summary and Implications

The reported meta-analyses examined the finding that memory is superior for individuals of one's own age group compared with individuals of another age group. Results showed that, relative to other-age faces, own-age faces engendered higher levels of hits, lower levels of false alarms, and superior discriminability. Theoretical consensus on the mechanisms that govern the OAB remains elusive, but the meta-analysis of discriminability is consistent with a perceptual expertise account, suggesting that the most recent experiences drive the OAB. A social–cognitive explanation, suggesting that categorization of an individual as an ingroup or outgroup member influences encoding, may also be viable, but that possibility remains to be fully tested. Nonetheless, on the basis of recent theory in the own-race bias literature (Hugenberg et al., 2010), we suggest that the most comprehensive explanation of the OAB will likely integrate perceptual experience and social–cognitive perspectives.

The implications of the OAB are broad. Most important, for situations where there is a mismatch between the age of the rememberer and the age of a suspect and/or culprit, the data reported suggest that recognition accuracy will be reduced. This is all the more important given that older adults are frequently the target of scams (Federal Bureau of Investigation, n.d.; see also Jacoby & Rhodes, 2006), most of which are presumably committed by perpetrators of a different age. As well, crimes against children may be most likely to be perpetrated by older individuals. Consequently, accurate identification of the culprit may be less likely when there is a mismatch between the age of the perpetrator and the age of the victim (but see data from the meta-analysis of hits indicating that effect sizes were smaller for studies using lineups). Given that eyewitness misidentification appears to be the leading cause of wrongful convictions (e.g., Scheck et al., 2000), our data suggest that caution should be exercised when the age of a witness differs from the age of a suspect. Although we make no specific arguments for the weight the OAB should be given in legal proceedings, we do suggest that the interaction of the age of the witness and the age of the perpetrator should be given consideration when judging the accuracy of an identification. In all, the OAB joins a catalog of other studies indicating that recognition is superior for ingroup compared with outgroup individuals.

References

References marked with an asterisk indicate studies included in the meta-analyses.

Adams-Price, C. (1992). Eyewitness memory and aging: Predictors of accuracy in recall and person recognition. *Psychology and Aging, 7,* 602–608. doi:10.1037/0882-7974.7.4.602

*Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review, 12,* 1043–1047.

*Anastasi, J. S., & Rhodes, M. G. (2006). Evidence for an own-age bias in face recognition. *North American Journal of Psychology, 8,* 237–253.

*Bäckman, L. (1991). Recognition memory across the adult life span: The role of prior knowledge. *Memory & Cognition, 19,* 63–71. doi:10.3758/BF03198496

*Bartlett, J. C., & Fulton, A. (1991). Familiarity and recognition of faces in old age. *Memory & Cognition, 19,* 229–238.

*Bartlett, J. C., & Leslie, J. E. (1986). Aging and memory for faces versus single views of faces. *Memory & Cognition, 14,* 371–381.

Bartlett, J. C., Strater, L., & Fulton, A. (1991). False recency and false fame of faces in young adulthood and old age. *Memory & Cognition, 19,* 177–188. doi:10.3758/BF03197115

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science, 18,* 706–712. doi:10.1111/j.1467-9280.2007.01964.x

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. doi:10.1037/0033-2909.100.1.3

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis* (pp. 221–235). New York, NY: Russell Sage Foundation.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta-Analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77,* 305–327. doi:10.1111/j.2044-8295.1986.tb02199.x

Bruyer, R., Lafalize, A., & Distefano, M. (1991). Age decisions on familiar and unfamiliar faces. *Behavioural Processes, 24,* 21–35. doi:10.1016/0376-6357(91)90084-D

Bruyer, R., Mejias, S., & Doublet, S. (2007). Effect of face familiarity on age decision. *Acta Psychologica, 124,* 159–176. doi:10.1016/j.actpsy.2006.03.001

Cassia, V. M. (2011). Age biases in face processing: The effects of experience across development. *British Journal of Psychology.* Advance online publication. doi:10.1111/j.2044-8295.2011.02046.x

*Cassia, V. M., Kuefner, D., Picozzi, M., & Vescovo, E. (2009). Early experience predicts later plasticity for face processing: Evidence for the reactivation of dormant effects. *Psychological Science, 20,* 853–859. doi:10.1111/j.1467-9280.2009.02376.x

*Cassia, V. M., Picozzi, M., Kuefner, D., & Casati, M. (2009). Why mix-ups don't happen in the nursery: Evidence for an experience-based interpretation of the other-age effect. *Quarterly Journal of Experimental Psychology, 62,* 1099–1107. doi:10.1080/17470210802617654

Chance, J. E., & Goldstein, A. G. (1984). Face-recognition memory: Implications for children's eyewitness testimony. *Journal of Social Issues, 40,* 69–85. doi:10.1111/j.1540-4560.1984.tb01094.x

*Chance, J. E., Goldstein, A. G., & Andersen, B. (1986). Recognition memory for infant faces: An analog of the other-race effect. *Bulletin of the Psychonomic Society, 24,* 257–260.

Chasteen, A. L. (2005). Seeing eye-to-eye: Do intergroup biases operate similarly for younger and older adults? *The International Journal of Aging and Human Development, 61,* 123–139. doi:10.2190/07Q7-BWYT-NC9E–51FX

Chiroro, P. M., Tredoux, C. G., Radaelli, S., & Meissner, C. A. (2008). Recognizing faces across continents: The effect of within-race variations in the own-race bias in face recognition. *Psychonomic Bulletin & Review, 15,* 1089–1092. doi:10.3758/PBR.15.6.1089

*Chung, M. S. (1997). Face recognition: Effects of age of subjects and age of stimulus faces. *Korean Journal of Developmental Psychology, 10,* 167–176.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York, NY: Academic Press.

*Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face space. *Cognition, 111,* 219–247. doi:10.1016/j.cognition.2009.02.004

Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics, 10,* 393–396. doi:10.3758/BF03210319

de Heering, A., de Liedekerke, C., Deboni, M., & Rossion, B. (2010). The role of experience during childhood in shaping the other-race effect. *Developmental Science, 13,* 181–187. doi:10.1111/j.1467-7687.2009.00876.x

de Heering, A., & Rossion, B. (2008). Prolonged visual experience in adulthood modulates holistic face perception. *PLoS ONE, 3,* Article e2317. doi:10.1371/journal.pone.0002317

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117. doi:10.1037/0096-3445.115.2.107

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177. doi:10.1037/1082-989X.1.2.170

Ebner, N. C. (2008). Age of face matters: Age-group differences in ratings of young and old faces. *Behavior Research Methods, 40,* 130–136. doi:10.3758/BRM.40.1.130

Ebner, N. C., Gluth, S., Johnson, M. R., Raye, C. L., Mitchell, K. J., & Johnson, M. K. (2011). Medial prefrontal cortex activity when thinking about others depends on their age. *Neurocase, 17,* 260–269. doi:10.1080/13554794.2010.536953

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. doi:10.1093/scan/nsq074

Ebner, N. C., He, Y., & Johnson, M. K. (2011). Age and emotion affect how we look at a face: Visual scan patterns differ for own-age versus other-age emotional faces. Advance online publication. *Cognition & Emotion.* doi:10.1080/02699931.2010.540817

*Ebner, N. C., & Johnson, M. K. (2009). Young and older emotional faces: Are there group differences in expression identification and memory? *Emotion, 9,* 329–339. doi:10.1037/a0015179

*Ebner, N. C., Riediger, M., & Lindenberger, U. (2009). Schema reliance for developmental goals increases from early to late adulthood: Improvement for the young, loss prevention for the old. *Psychology and Aging, 24,* 310–323. doi:10.1037/a0015430

Elliott, E. S., Wills, E. J., & Goldstein, A. G. (1973). The effects of discrimination training on the recognition of White and Oriental faces. *Bulletin of the Psychonomic Society, 2,* 71–73.

Enlow, D. (1982). *The handbook of facial growth* (2nd ed.). London, United Kingdom: W. B. Saunders.

Federal Bureau of Investigation. (n.d.). Fraud target: Senior citizens. Retrieved from http://www.fbi.gov/majcases/fraud/seniorsfam.htm

*Firestone, A., Turk-Browne, N., & Ryan, J. (2007). Age-related deficits in face recognition are related to underlying changes in scanning behavior. *Aging, Neuropsychology, and Cognition, 14,* 594–607. doi:10.1080/13825580600899717

Fleiss, J. L., & Berlin, J. A. (2009). Effect sizes for dichotomous outcomes. In H. Cooper, L. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis* (pp. 237–253). New York, NY: Russell Sage Foundation.

Freund, A. M., Kourilova, S., & Kuhl, P. (2011). Stronger evidence for own-age effects in memory for older as compared to younger adults. *Memory, 19,* 429–448. doi:10.1080/09658211.2011.583929

*Fulton, A., & Bartlett, J. C. (1991). Young and old faces in young and old heads: The factor of age in face recognition. *Psychology and Aging, 6,* 623–630. doi:10.1037/0882-7974.6.4.623

Gart, J. J., & Zwiefel, J. R. (1967). On the bias of various estimators of the logit and its variance, with application to quantal bioassay. *Biometrika, 54,* 181–187.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanism for face recognition. *Vision Research, 37,* 1673–1682. doi:10.1016/S0042-6989(96)00286-6

*Gilchrist, A., & McKone, E. (2003). Early maturity of face processing in children: Local and relational distinctiveness effects in 7-year-olds. *Visual Cognition, 10,* 769–793. doi:10.1080/13506280344000022

Gluth, S., Ebner, N. C., & Schmiedek, F. (2010). Attitudes toward younger and older adults: The German Aging Semantic Differential. *International Journal of Behavioral Development, 34,* 147–158. doi:10.1177/0165025409350947

Goldinger, S. D., He, Y., & Papesh, M. H. (2009). Deficits in cross-race face learning: Insights from eye movements and pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1105–1122. doi:10.1037/a0016548

*Goldstein, A. G., & Chance, J. E. (1964). Recognition of children's faces. *Child Development, 35,* 129–136.

*Goodsell, C. A., Neuschatz, J. S., & Gronlund, S. D. (2009). Effects of mugshot commitment on lineup performance in young and older adults. *Applied Cognitive Psychology, 23,* 788–803. doi:10.1002/acp.1512

*Gross, T. F. (2009). Own-ethnicity bias in the recognition of Black, East Asian, Hispanic, and White faces. *Basic and Applied Social Psychology, 31,* 128–135. doi:10.1080/01973530902880381

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology, 99,* 45–56. doi:10.1348/000712607X199981

*Harrison, V., & Hole, G. J. (2009). Evidence for a contact-based explanation of the own-age bias in face recognition. *Psychonomic Bulletin & Review, 16,* 264–269. doi:10.3758/PBR.16.2.264

*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,* 847–859. doi:10.1080/09658210903277318

*He, Y., Ebner, N., & Johnson, M. K. (2011). What predicts the own-age bias in face recognition memory? *Social Cognition, 29,* 97–109.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128. doi:10.2307/1164588

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504. doi:10.1037/1082-989X.3.4.486

Hehman, E., Mania, E. W., & Gaertner, S. L. (2010). Where the division lies: Common ingroup identity moderates the cross-race-facial recognition effect. *Journal of Experimental Social Psychology, 46,* 445–448. doi:10.1016/j.jesp.2009.11.008

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539–1558. doi:10.1002/sim.1186

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557–560. doi:10.1136/bmj.327.7414.557

Hildebrandt, A., Sommer, W., Herzmann, G., & Wilhelm, O. (2010). Structural invariance and age-related performance differences in face cognition. *Psychology and Aging, 25,* 794–810. doi:10.1037/a0019774; 10.1037/a0019774

*Hills, P. J., & Lewis, M. B. (2011). The own-age face recognition bias in children and adults. *Quarterly Journal of Experimental Psychology, 64,* 17–23. doi:10.1080/17470218.2010.537926

Horry, R., Wright, D. B., & Tredoux, C. G. (2010). Recognition and context memory for faces from own and other ethnic groups: A remember–know investigation. *Memory & Cognition, 38,* 134–141. doi:10.3758/MC.38.2.134

*Hourihan, K. L., Benjamin, A. S., & Gronlund, S. D. (2010, November). *An own-group bias in metamnemonic accuracy for faces.* Poster session presented at the 51st Annual Meeting of the Psychonomic Society, St. Louis, MO.

Hugenberg, K., & Corneille, O. (2009). Holistic processing is tuned for in-group faces. *Cognitive Science, 33,* 1173–1181. doi:10.1111/j.1551-6709.2009.01048.x

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review, 117,* 1168–1187. doi:10.1037/a0020463

Jacoby, L. L., & Rhodes, M. G. (2006). False remembering in the aged. *Current Directions in Psychological Science, 15,* 49–53. doi:10.1111/j.0963-7214.2006.00405.x

Johnson, K. J., & Fredrickson, B. L. (2005). "We all look the same to me": Positive emotions eliminate the own-race bias in face recognition. *Psychological Science, 16,* 875–881. doi:10.1111/j.1467-9280.2005.01631.x

Johnston, R. A., Kanazawa, M., Kato, T., & Oda, M. (1997). Exploring the structure of multidimensional face-space: The effects of age and gender. *Visual Cognition, 4,* 39–57. doi:10.1080/713756750

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

*Kuefner, D., Cassia, V. M., Picozzi, M., & Bricolo, E. (2008). Do all kids look alike? Evidence for an other-age effect in adults. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 811–817. doi:10.1037/0096-1523.34.4.811

Kuefner, D., Cassia, V. M., Vescovo, E., & Picozzi, M. (2010). Natural experience acquired in adulthood enhances holistic processing of other-age faces. *Visual Cognition, 18,* 11–25. doi:10.1080/13506280802396507

*Lamont, A. C., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: Effects of target age and memory load. *Memory & Cognition, 33,* 1017–1024.

Lavrakas, P. J., Buri, J. R., & Mayzner, M. S. (1976). A perspective on the recognition of other-race faces. *Perception & Psychophysics, 20,* 475–481. doi:10.3758/BF03208285

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1364–1382. doi:10.1037/0278-7393.22.6.1364

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 129,* 559–574. doi:10.1037/0096-3445.129.4.559

Lindholm, T. (2005). Own-age biases in verbal person memory. *Memory, 13,* 21–30. doi:10.1080/09658210344000549

List, J. A. (1986). Age and schematic differences in the reliability of eyewitness testimony. *Developmental Psychology, 22,* 50–57. doi:10.1037/0012-1649.22.1.50

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

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide.* Mahwah, NJ: Erlbaum.

Malpass, R. S. (1993). "They all look alike to me." In M. Merrens & G. Brannigan (Eds.), *The undaunted psychologist: Adventures in research* (pp. 74–88). New York, NY: McGraw-Hill.

Malpass, R. S., & Kravitz, J. (1969). Recognition of faces of own- and other race. *Journal of Personality and Social Psychology, 13,* 330–334. doi:10.1037/h0028434

Malpass, R. S., Lavigueur, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception & Psychophysics, 14,* 285–292. doi:10.3758/BF03212392

Marcon, J. L., Susa, K. J., & Meissner, C. A. (2009). Assessing the influence of recollection and familiarity in memory for own- versus other-race faces. *Psychonomic Bulletin & Review, 16,* 99–103. doi:10.3758/PBR.16.1.99

*Mason, S. E. (1986). Age and gender as factors in facial recognition and identification. *Experimental Aging Research, 12,* 151–154.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255–260. doi:10.1016/S1364-6613(02)01903-4

McGillivray, S., & Castel, A. D. (2010). Memory for age–face associations in younger and older adults: The role of generation and schematic support. *Psychology and Aging, 25,* 822–832. doi:10.1037/a0021044

McKelvie, S. J., Standing, L., St. Jean, D., & Law, J. (1993). Gender differences in recognition memory for faces and cars: Evidence for the interest hypothesis. *Bulletin of the Psychonomic Society, 31,* 447–448.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7,* 3–35. doi:10.1037/1076-8971.7.1.3

Meissner, C. A., Brigham, J. C., & Butz, D. A. (2005). Memory for own- and other-race faces: A dual process approach. *Applied Cognitive Psychology, 19,* 545–567. doi:10.1002/acp.1097

Melinder, A., Gredebäck, G., Westerlund, A., & Nelson, C. (2010). Brain activation during upright and inverted encoding of own- and other-age faces: ERP evidence for an own-age bias. *Developmental Science, 13,* 588–598. doi:10.1111/j.1467-7687.2009.00910.x

*Memon, A., Bartlett, J., Rose, R., & Gray, C. (2003). The aging eyewitness: Effects of age on face, delay, and source-memory ability. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 58,* P338–P345.

Michel, C., Rossion, B., Han, J., Chung, C., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science, 17,* 608–615. doi:10.1111/j.1467-9280.2006.01752.x

*Mondloch, C. J., Maurer, D., & Ahola, S. (2006). Becoming a face expert. *Psychological Science, 17,* 930–934. doi:10.1111/j.1467-9280.2006.01806.x

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

O'Neil, S. A., & Webster, M. A. (2011). Adaptation and the perception of facial age. *Visual Cognition, 19,* 534–550. doi:10.1080/13506285.2011.561262

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *PNAS: Proceedings of the National Academy of Sciences of the United States of America, 102,* 5297–5300. doi:10.1073/pnas.0406627102

Pauker, K., Weisbuch, M., Ambady, N., Sommers, S. R., Adams, R. B., Jr., & Ivcevic, Z. (2009). Not so black and white: Memory for ambiguous group members. *Journal of Personality and Social Psychology, 96,* 795–810. doi:10.1037/a0013265

*Perfect, T. J., & Harris, L. J. (2003). Adult age differences in unconscious transference: Source confusion or identity blending. *Memory & Cognition, 31,* 570–580.

*Perfect, T. J., & Moon, H. C. (2005). The own-age effect in face recognition. In J. Duncan, P. McLeod, & L. Phillips (Eds.), *Speed, control and age: In honour of Patrick Rabbitt* (pp. 317–340). Oxford, United Kingdom: Oxford University Press.

Platz, S. J., & Hosch, H. M. (1988). Cross-racial/ethnic eyewitness identification: A field study. *Journal of Applied Social Psychology, 18,* 972–984. doi:10.1111/j.1559-1816.1988.tb01187.x

*Pozzulo, J. D., & Dempsey, J. (2009). Could target age explain identification accuracy differences between child and adult eyewitnesses? *Psychiatry, Psychology, and Law, 16,* S137–S144. doi:10.1080/13218710802620414

*Randall, J. L., Tabernik, H. E., Aguilera, A. M., Valk, K. V., & Anastasi, J. S. (2011). *Effects of encoding tasks on the own-age face recognition bias.* Manuscript submitted for publication.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis* (pp. 295–315). New York, NY: Russell Sage Foundation.

*Rehnman, J., & Herlitz, A. (2006). Higher face recognition ability in girls: Magnified by own-sex and own-ethnicity bias. *Memory, 14,* 289–296. doi:10.1080/09658210500233581

*Rehnman, J., & Herlitz, A. (2007). Women remember more faces than men do. *Acta Psychologica, 124,* 344–355. doi:10.1016/j.actpsy.2006.04.004

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 17,* 43–63. doi:10.1068/p170043

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology, 80,* 313–331. doi:10.1111/j.2044-8295.1989.tb02323.x

Rhodes, M. G. (2009). Age estimation of faces: A review. *Applied Cognitive Psychology, 23,* 1–12. doi:10.1002/acp.1442

*Rhodes, M. G., Castel, A. D., & Jacoby, L. L. (2008). Associative recognition of face pairs by younger and older adults: The role of familiarity-based processing. *Psychology and Aging, 23,* 239–249. doi:10.1037/0882-7974.23.2.239

*Rodin, M. J. (1987). Who is memorable to whom: A study of cognitive disregard. *Social Cognition, 5,* 144–165. doi: 10.1521/soco.1987.5.2.144

*Rose, R. A., Bull, R., & Vrij, A. (2003). Enhancing older witnesses' identification performance: Context reinstatement is not the answer. *The Canadian Journal of Police and Security Services, 1,* 173–184.

*Rose, R. A., Bull, R., & Vrij, A. (2005). Non-biased lineup instructions do matter—A problem for older witnesses. *Psychology, Crime & Law, 11,* 147–159. doi:10.1080/10683160512331316307

Rosenthal, R. (1979). The "file-drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638–641. doi:10.1037/0033-2909.86.3.638

Rule, N. O., Ambady, N., Adams, R. B., Jr., & Macrae, C. N. (2007). Us and them: Memory advantages in perceptually ambiguous groups. *Psychonomic Bulletin & Review, 14,* 687–692. doi:10.3758/BF03196822

Sánchez-Meca, J., Marín-Martínez, F., & Chac´on Moscoso, S. (2003).

Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8,* 448–467. doi:10.1037/1082-989X.8.4.448

Sangrigoli, S., Pallier, C., Argenti, A.-M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science, 16,* 440–444.

Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual innocence: When justice goes wrong and how to make it right.* New York, NY: Random House.

Shadish, W. R., & Haddock, C. K. (2009). Combining estimates of effect size. In H. Cooper, L. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis* (pp. 257–277). New York, NY: Russell Sage Foundation.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139–156. doi:10.1037/0033-2909.100.2.139

Slessor, G., Laird, G., Phillips, L. H., Bull, R., & Filippou, D. (2010). Age-related differences in gaze following: Does the age of the face matter? *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 65,* 536–541. doi:10.1093/geronb/gbq038

Slone, A. E., Brigham, J. C., & Meissner, C. A. (2000). Social and cognitive factors affected the own-race bias in Whites. *Basic and Applied Social Psychology, 22,* 71–84. doi:10.1207/15324830051036162

Sporer, S. L. (2001). Recognizing faces of other ethnic groups. *Psychology, Public Policy, and Law, 7,* 36–97. doi:10.1037/1076-8971.7.1.36

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis* (pp. 435–452). New York, NY: Russell Sage Foundation.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93,* B1–B9. doi:10.1016/j.cognition.2003.09.011

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 43*(A), 161–204. doi:10.1080/14640749108400966

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition, 20,* 291–302. doi:10.3758/BF03199666

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest, 7,* 45–75. doi:10.1111/j.1529-1006.2006.00027.x

*Wiese, H., Schweinberger, S. R., & Hansen, K. (2008). The age of the beholder: ERP evidence of an own-age bias in face memory. *Neuropsychologia, 46,* 2973–2985. doi:10.1016/j.neuropsychologia.2008.06.007

Wiese, H., Schweinberger, S. R., & Neuman, M. F. (2008). Perceiving age and gender in unfamiliar faces: Brain potential evidence for implicit and explicit person categorization. *Psychophysiology, 45,* 957–969. doi:10.1111/j.1469-8986.2008.00707.x

*Wilcock, R., & Bull, R. (2010). Novel lineup methods for improving the performance of older eyewitnesses. *Applied Cognitive Psychology, 24,* 718–736. doi:10.1002/acp.1582

*Wilcock, R. A., Bull, R., & Vrij, A. (2005). Aiding the performance of older eyewitnesses: Enhanced non-biased line-up instructions and line-up presentation. *Psychiatry, Psychology, and Law, 12,* 129–140. doi:10.1375/pplt.2005.12.1.129

*Wilcock, R., Bull, R., & Vrij, A. (2007). Are old witnesses always poorer witnesses? Identification accuracy, context reinstatement, own-age bias. *Psychology, Crime & Law, 13,* 305–316. doi:10.1080/10683160600822212

*Wright, C. I., Negreira, A., Gold, A. L., Britton, J. C., Williams, D., & Barrett, L. F. (2008). Neural correlates of novelty and face–age effects in young and elderly adults. *NeuroImage, 42,* 956–968.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Applied Cognitive Psychology, 17,* 365–373. doi:10.1002/acp.898

Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica, 114,* 101–114. doi:10.1016/S0001-6918(03)00052-0

*Wright, D. B., & Stroud, J. S. (2002). Age differences in lineup identification accuracy: People are better with their own age. *Law and Human Behavior, 26,* 641–654. doi:10.1023/A:1020981501383

Yarmey, A. D. (1993). Adult age and gender differences in eyewitness recall in field settings. *Journal of Applied Social Psychology, 23,* 1921–1932. doi:10.1111/j.1559-1816.1993.tb01073.x

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141–145. doi:10.1037/h0027474

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16,* 747–759. doi:10.1068/p160747

Received February 13, 2011
Revision received August 24, 2011
Accepted August 30, 2011 ■