



HHS Public Access

Author manuscript

Psychol Aging. Author manuscript; available in PMC 2015 March 11.

Published in final edited form as:

Psychol Aging. 2014 September ; 29(3): 454–468. doi:10.1037/a0036255.

Older and Younger Adults' Accuracy in Discerning Health and Competence in Older and Younger Faces

Leslie A. Zebrowitz, Robert G. Franklin Jr., Jasmine Boshyan, Victor Luevano, Stefan Agrigoroaei, Bosiljka Milosavljevic, and Margie E. Lachman
Brandeis University

Abstract

We examined older and younger adults' accuracy judging the health and competence of faces. Accuracy differed significantly from chance and varied with face age but not rater age. Health ratings were more accurate for older than younger faces, with the reverse for competence ratings. Accuracy was greater for low attractive younger faces, but not for low attractive older faces. Greater accuracy judging older faces' health was paralleled by greater validity of attractiveness and looking older as predictors of their health. Greater accuracy judging younger faces' competence was paralleled by greater validity of attractiveness and a positive expression as predictors of their competence. Although the ability to recognize variations in health and cognitive ability is preserved in older adulthood, the effects of face age on accuracy and the different effects of attractiveness across face age may alter social interactions across the life span.

Keywords

impression accuracy; rater age; face age; facial appearance cues

Research has documented surprising accuracy in younger adults' ratings of health and intelligence from facial appearance. For example, health ratings of facial photographs of adolescents from a representative sample of individuals were significantly correlated with their actual health scores (Kalick, Zebrowitz, Langlois, & Johnson, 1998). In addition, intelligence ratings based on photographs from the same sample were significantly correlated with IQ scores (Zebrowitz, Hall, Murphy, & Rhodes, 2002). This ability to discern variations in health and intelligence is consistent with an ecological approach to social perception (McArthur & Baron, 1983; Zebrowitz, 2011). According to ecological theory (e.g., Gibson, 1979), "perceiving is for doing" with the result that perceptions of adaptively relevant attributes will be accurate. Certainly the accurate detection of cognitive competence and health may facilitate adaptive behaviors, such as choosing to associate with

© 2014 American Psychological Association

Correspondence concerning this article should be addressed to: Leslie A. Zebrowitz, Department of Psychology, Brandeis University, 415 South Street – MS 062 – PO Box 549110, Waltham, MA 02454-9110. zebrowitz@brandeis.edu.

Leslie A. Zebrowitz, Robert G. Franklin Jr., Jasmine Boshyan, Victor Luevano, Stefan Agrigoroaei, Bosiljka Milosavljevic, and Margie E. Lachman, Department of Psychology, Brandeis University.

Robert G. Franklin Jr. is now in the Department of Psychology at Anderson University; Victor Luevano is now in the Department of Psychology at California State University, Stanislaus; Bosiljka Milosavljevic is now in the Department of Psychology at Goldsmiths College, University of London.

people who do not present a risk of contagion or a caregiving burden and those who can provide useful advice. As discussed below, declines in this ability in older adulthood may be predicted from an older adult “positivity bias” (Mather & Carstensen, 2005), although there also is evidence to suggest that accurate trait impressions from faces will be preserved in older adults (Boshyan, Zebrowitz, Franklin, McCormick, & Carré, 2013). The present study assessed these two possibilities as well as variations in accuracy as a function of face age and the facial cues that contribute to accurate trait impressions.

In addition to finding that younger raters’ impressions of health and intelligence from faces were accurate predictors of actual scores, previous research revealed that accurate impressions were stronger for faces below than above the median in attractiveness, and that attractiveness was a valid cue to health and intelligence for faces ranging from low to average attractiveness, but not for faces ranging from average to high (Zebrowitz & Rhodes, 2004). This result was attributed to the fact that not only do very unattractive faces often signal low health or intelligence—for example, pronounced genetic anomalies like Down’s syndrome and contagious diseases like smallpox—but even moderately unattractive faces in a representative sample of people may do so. For example “minor facial anomalies” and subtle facial characteristics marking fetal alcohol syndrome signal low intelligence (Froud et al., 2012; Cummings, Flynn, & Preus, 1982). Consistent with this argument, research has demonstrated that faces at the low end of the attractiveness continuum in a representative sample of young people are more structurally similar to faces with marked genetic anomalies than are faces high in attractiveness (Zebrowitz, Fellous, Mignault, & Andreoletti, 2003). The evidence that younger faces low in attractiveness contain the most valid cues to low cognitive and physical fitness provides reason to expect rater age and face age to moderate accuracy in judging these traits.

Declines in accuracy perceiving health and competence in older adulthood may be predicted from age-related declines in responsiveness to negative stimuli, an older adult “positivity effect” (Mather & Carstensen, 2005; Murphy & Isaacowitz, 2008; Reed, Chan & Mikels, 2014). Most pertinent to our hypotheses regarding accuracy of trait impressions, recent research has revealed that older adults formed more positive impressions of the health of the faces we used to assess accuracy in the present study, as well as more positive impressions of trustworthiness, although there were no age differences in positivity of competence impressions (Zebrowitz, Franklin, Hillman, & Boc, 2013). In addition, compared with younger raters, older raters’ impressions of health were more responsive to high attractiveness, and their impressions of trustworthiness were less responsive to low attractiveness (Zebrowitz, Franklin, Hillman, & Boc, 2013). These results suggest that older raters may be less attuned to the negatively valenced cues in faces low in attractiveness that facilitate accurate trait impressions by younger adults. If so, then older adults’ trait ratings from faces may be less accurate than those of younger raters, particularly for faces that are relatively low in attractiveness. Whereas the foregoing considerations suggest that younger adults may be more accurate than older adults when assessing competence and health, there is also reason to expect no age differences. Most notably, a recent study revealed that older raters’ impressions of the aggressiveness of young male faces was significantly correlated with the men’s actual aggressiveness, with no difference from the accuracy of younger raters (Boshyan et al., 2013). These results suggest that a positivity effect shown in older adults’

trait impressions does not necessarily imply that those impressions will be less accurate than those of younger adults.

At least two factors may contribute to differences in the accuracy of first impressions of older and younger faces. First, there may be more variability in the health and competence of older individuals, and these qualities may therefore be more evident in their faces. If so, then ratings of health and competence should be more accurate for older than younger faces. In addition, effects of face age on accuracy may be moderated by the level of the attractiveness of the faces. As discussed above, the greater accuracy shown when judging younger faces below than above average in attractiveness has been attributed to the fact that compared with average attractiveness, low attractiveness signals lower fitness, whereas high attractiveness does not signal higher fitness. This may not be true for older faces.

Specifically, factors unrelated to fitness may contribute more to low attractiveness in older faces, such as skin changes that are not associated with pathology (Pariser, 1998). Such factors would reduce the validity of cues to low health or competence provided by low attractive older faces. In addition, in contrast to younger faces, factors that are actually related to fitness may contribute to greater than average attractiveness in older faces. In particular, looking young for one's age is associated with greater attractiveness in older adults (Kotter-Grühn & Hess, 2012), and younger-looking older adults are actually healthier and longer-lived than older looking individuals of the same age (Christensen et al., 2009). Not only may high attractiveness be a more valid cue to good health in older faces than is low attractiveness, it also may be a more valid cue to high competence, because better health in older adults is associated with better cognitive functioning (Alosco et al., 2012; Benedict et al., 2013).

To provide insight into the mechanisms accounting for variations in accuracy, we applied a lens model (Brunswik, 1955) that has been used in previous research (Zebrowitz & Rhodes, 2004; Boshyan et al., 2013). As shown in Figure 1, the paths on the left of the model indicate cue utilization— correlations between facial qualities and perceived traits. The paths on the right of the model indicate cue validity— correlations between facial qualities and measured traits. The path at the bottom of the model indicates accuracy— correlations between measured traits and perceived traits. The cues we assessed included the apparent age of the face, controlling for actual age to get an index of 'looking old for one's age,' as well as facial qualities that had contributed to accurate impressions by younger raters in previous research— attractiveness, symmetry, and facial expression (Halberstadt, Ruffman, Murray, Taumoepeau, & Ryan, 2011; Stanley & Blanchard-Fields, 2008; Zebrowitz & Rhodes, 2004).

To summarize, we predicted the following: a) raters' first impressions of the health and competence of people based only on static facial images would accurately predict indices of actual health and competence; b) older adults would show less accurate impressions than younger adults, consistent with evidence that older adults show less attention to negatively valenced cues that yield greater accuracy by younger adults; c) accurate first impressions of health and competence would be greater for older than younger faces, consistent with age-related increases in diagnostic facial cues; d) younger raters would show more accuracy for faces below than above the median in attractiveness, consistent with previous research, but

older raters would not show this effect due to their lesser attention to negatively valenced cues; e) accurate first impressions of younger faces would be greater for those below than above the median in attractiveness, consistent with previous research, but this will not be true for older faces because of less difference in the diagnostic cues available in low and high attractive older faces; and f) variations in the facial cues utilized when judging health and competence and the cues that are valid indicators of these traits will help to explain variations in the accuracy of trait impressions as a function of rater age and face age.

Method

Participants

Forty-eight younger raters (23 men, $M = 18.8$, $SD = 1.0$) and 48 older raters (24 men, $M = 76.3$, $SD = 6.4$) participated. Younger raters were recruited from a university and completed the study for course credit or payment of \$15. Older raters were recruited from the local community and were paid \$25 for completing the study. Older raters were screened using the Mini-Mental State Examination (Folstein et al., 1975) all scoring above 26 of 30 ($M = 28.9$, $SD = 1.2$). Measures of vision, affect, and cognitive function were also administered to assess participant age differences. Results are consistent with previous studies of community-dwelling older versus younger adults, demonstrating the representativeness of our sample. Specifically, older raters performed worse than younger raters on the following: visual acuity and contrast sensitivity; a speeded pattern comparison task, consistent with decreases in processing speed in older adulthood (Salthouse, 1996); and a letter–number sequencing task and the Wisconsin card sorting task, consistent with decreases in executive function (Daniels, Toth, & Jacoby, 2006). In contrast to poorer performance by older adults on the preceding measures, they performed better than younger adults on a vocabulary task, consistent with their higher education level and the maintenance of crystallized intelligence in older adulthood (Horn & Cattell, 1967) (see Table 1).

Facial Stimuli

Facial stimuli were taken from two databases. One was the Intergenerational Studies (IGS) archive, a longitudinal study that included representative samples of individuals born in Berkeley, California, in 1928–29 (every third birth) or enrolled in 5th or 6th grades in Oakland, California in 1932 (see Block, 1971, for more details). The second was the Boston Study of Management Processes, a Boston-area subsample of participants from the study of Midlife in the United States (MIDUS) who were selected by random digit dialing (Lachman, 1997). The IGS images included two greyscale frontal photographs of 148 individuals (74 men) who were photographed between 17 and 18 years of age (younger adults) and again between 52 and 62 years of age (older adults). The MIDUS images included color frontal facial photographs of 69 individuals (41 men) who were between the ages of 25 and 39 (younger adults, $M = 32.7$, $SD = 4.2$), and 68 different individuals (44 men) who were between the ages of 60 and 74 (older adults, $M = 66.1$, $SD = 4.1$).

Face Ratings

Ratings of the health, competence, and attractiveness of the faces were taken from previous research using the same participants and facial stimuli as the present study to investigate

rater age differences in trait impressions from faces (Zebrowitz et al., 2013) and face stereotypes (Franklin & Zebrowitz, 2013). Ratings of facial symmetry of the IGS faces were taken from another previous study (Zebrowitz, Voinescu, & Collins, 1996) as were ratings of facial symmetry, expression, and age of the Boston MIDUS faces (Luevano, 2007). Raters of symmetry and expression were all younger adults. Raters of apparent age of MIDUS faces included both younger and older adults, with ratings averaged across age groups attributable to a high correlation, $r = .98$. Ratings of the expression and apparent age of IGS faces were collected for the present study from 8 younger raters (5 females).

Ratings of health, competence, and attractiveness used 7-point scales with end points labeled *not at all* and *very*. Apparent age was estimated in an open-ended question with age provided in years, and the other ratings were made on 7-point scales with endpoints labeled *very symmetrical* and *not at all symmetrical, no positive (negative) expression*, and *very positive (negative) expression*. There was high interrater reliability for all measures: symmetry $\alpha = .78$ and $.81$ for IGS and Boston MIDUS faces respectively; positive and negative expression $\alpha = .92$ and $.96$ for IGS faces, and $\alpha = .95$ and $.94$, for Boston MIDUS faces; age $\alpha = .99$ for both IGS and Boston MIDUS faces.

Because positive and negative expression ratings were highly correlated ($r = .93$ for IGS and $.96$ for MIDUS faces), they were standardized and averaged into a single variable with higher values representing a more positive expression. We also created a composite 'attractiveness' variable by standardizing and averaging mean attractiveness and mean symmetry ratings for each face across raters, because these two ratings were positively correlated ($r = .34$ for IGS and $.53$ for MIDUS faces), which is consistent with evidence that symmetry is one of the major components of an attractive facial structure (Rhodes, 2006). We kept apparent age separate from the other cues both on conceptual grounds and also because it was not highly correlated with expression ($r = -.10$ for IGS and $.01$ for MIDUS faces), symmetry ($r = -.15$ for IGS and $-.04$ for MIDUS faces), or attractiveness ($r = -.27$ for IGS and $-.20$ for MIDUS faces). As noted above, apparent age should be understood as 'older-looking' for one's age because actual age was controlled in the analyses.

Criterion Measures

Health—For IGS faces, the younger adult health measure was based on clinical exams and detailed histories from age 11 through 18. Staff physicians examined participants and prepared charts, which were used by the research project medical director to make yearly ratings of each person on a 5-point scale ranging from *no illness* to *severe illness*. Ratings had been previously averaged from ages 11 through 18 to create a doctors' health score (Kalick et al., 1998) that we used in the present study. The alpha reliability for this composite was $.68$. The older adult health measure also used a 5-point scale, and was provided by the participants' personal physicians when they were between ages 58 and 66. Although health ratings in younger adulthood were available for all of the facial stimuli, they were available in older adulthood for only 70 of the faces. Previous research had shown that the younger adult health measure was largely based on the frequency, duration, and severity of infectious conditions and that chronic illnesses had an increased influence on

health ratings in older adulthood (Bayer & Snyder, 1950; Bayer, Whissell-Buechy, & Honzik, 1981).

For Boston MIDUS faces, self-ratings of functional abilities and fitness served as the criteria for assessing the accuracy of health ratings. The measure of functional abilities (Ware & Sherbourne, 1992) was created as a Guttman scale by summing responses to a multipart question from the main MIDUS study questionnaire. The question asked, *How much does your health limit you in doing each of the following?* The tasks were as follows: a) lifting or carrying groceries; b) bathing or dressing yourself; c) climbing several flights of stairs; d) bending, kneeling, or stooping; e) walking more than a mile; f) walking several blocks; g) walking one block. Responses were coded as 1 = *A lot*; 2 = *Some*; 3 = *A little*; 4 = *Not at all*, so that higher scores signified greater functional abilities. This measure has been validated in previous research that demonstrated a relationship to objective indicators of disability (Syddall, Martin, Harwood, Cooper, and Sayer (2009) as well as a sensitivity to various health protective factors (Lachman & Agrigoroaei, 2010). The measure of physical fitness was created by averaging reports of vigorous exercise during the summer and winter, $r = .85$, $p < .001$. The questions were *During the summer, how often do you engage in vigorous physical activity (e.g., running or lifting heavy objects) long enough to work up a sweat?* and *What about during the winter—how often do you engage in vigorous physical activity long enough to work up a sweat?* The response options for each question were identical: 6 = *Several times a week or more*; 5 = *About once a week*; 4 = *Several times a month*; 3 = *About once a month*; 2 = *Less than once a month*; 1 = *Never*.

Competence—For IGS faces, the competence criterion was IQ scores (Stanford-Binet scores at age 17–18, and Wechsler Adult Intelligence full scale scores, WAIS-R, in older adulthood). For Boston MIDUS faces, we used eight measures that were available in this database. To minimize the number of statistical comparisons and to avoid Type I errors, we grouped these measures into four cognitive factors that had been identified in previous research (Miller & Lachman, 2000): *Vocabulary* (vocabulary subscale of Wechsler Adult Intelligence Scale (WAIS); *Reasoning* (Raven’s Advanced Progressive Matrices; Raven, Raven, & Court, 1991; and the Schaie-Thurstone letter series, Schaie, 1985); *Processing Speed* (WAIS digit symbol substitution test and Letter Comparison Salthouse & Babcock, 1991); and *Short-Term Memory, short-term memory (STM)* (WAIS forward and backward digit spans, Wechsler, 1997, and Serial 7s, Folstein, Folstein, & McHugh, 1975).

Procedure

After obtaining informed consent, participants completed a computerized version of the PANAS (Watson, Clark & Tellegen, 1998). Next, MediaLab software (Empirisoft, New York, NY) was used to present the trait rating task. To keep the procedure to a manageable length, the 296 IGS faces were divided into two sets, with each set containing equal numbers of men and women of each age. Participants rated one of the resulting three sets of faces, yielding a total of 64 participants rating IGS faces and 32 participants rating MIDUS faces, with equal numbers of older and younger participants rating each set. Faces were shown in one of four orders, counterbalancing age and sex of face, with one of two rating scale orders that counterbalanced the order of health and competence ratings and attractiveness always

rated last.¹ Participants were asked to rate each face in comparison with the other faces of that age/sex grouping, because we were interested in whether older raters could discern differences in the health and competence of targets within demographic groups. These instructions were also given to raters of symmetry, expression, and apparent age. Faces were shown for either 4 s (MIDUS Boston Study faces) or 3 s (IGS faces), after which the face disappeared and the rating scale appeared on the screen, remaining until the rating was made. Participants rated all faces on one scale before the next one was introduced. An instruction screen appeared before each scale, indicating what the rating would be. Following the face ratings, a demographic and health questionnaire and the remaining control measures were administered.

Results

Overview of Accuracy Analyses

Accuracy analyses were performed at the level of the individual participant. Specifically, following the procedure recommended by Brand and Bradley (2012), we computed accuracy coefficients for health and competence by correlating each participant's ratings of the targets' faces with the targets' actual scores controlling face sex and actual face age. We controlled these variables to ensure that accuracy reflected sensitivity to facial information per se rather than sensitivity to differences in the health or competence of men and women or sensitivity to differences between older versus younger people within each age group. We computed younger and older raters' accuracy coefficients separately for younger and older faces. In addition, we computed a second set of accuracy coefficients for faces below and above the median in attractiveness. The median splits were performed within each face age group to identify high and low attractive faces of each age, and they were calculated using the mean attractiveness ratings of each face by all raters. Faces that fell on the median (6 faces for IGS and 10 faces for MIDUS) were excluded from this analysis. The accuracy coefficients were normalized using a Fisher z transformation for use in inferential statistics (see Franklin & Adams, 2009). Because the mean accuracy coefficients we report are averaged across many independent z transformed correlations computed for each rater, they are akin to mean effect sizes in a meta-analysis rather than a single correlation effect. Nevertheless, the magnitude of the effects they denote can be interpreted like correlation coefficients, because the z transform is equivalent to the r values at the values that we report.

Younger and older raters' accuracy coefficients for all younger faces and all older faces on each criterion measure were submitted to 2 (rater age group) \times 2 (face age group) ANOVAs, with rater age a between-subjects variable and face age a within-subjects variable. Table 2 shows the F and p values for overall accuracy and the main effects of rater age and face age for each accuracy criterion together with the accuracy coefficients and standard errors for younger raters, older raters, younger faces, and older faces. To examine effects of attractiveness, the accuracy coefficients computed separately for faces above and below the median were submitted to 2 (rater age group) \times 2 (face age group) \times 2 (face attractiveness:

¹Participants also rated face trustworthiness, hostility/aggressiveness, and babyfacedness, which are not relevant to the present study focus on accuracy of impressions.

above or below median) ANOVAs, with attractiveness an additional within-subjects variable. Table 3 shows the F and p values for the attractiveness level main effect and interaction effects of attractiveness \times rater age and attractiveness \times face age together with the accuracy coefficients and standard errors for faces above and below the median in attractiveness. Accuracy coefficients and standard deviations corresponding to significant interaction effects are reported in the text. In addition to the ANOVAs, we also performed t tests to determine whether accuracy differed from chance, with these results reported in the text.²

Overall Accuracy

Across all raters, health ratings predicted targets' doctors' health scores, functional abilities, and fitness at greater than chance levels, and competence ratings predicted targets' vocabulary, reasoning ability, processing speed, and STM at greater than chance levels, but not their IQ scores (see the significant overall accuracy effects in Table 2).

Effects of Rater Age on Accuracy

As shown in Table 2, there were no significant effects of rater age on accuracy, contrary to the prediction that older adults would be less accurate.

Effects of Face Age on Accuracy

Health ratings—Consistent with the prediction that traits might be more easily discerned in older than younger adult faces, health ratings from faces predicted doctors' health scores, functional abilities, and fitness more accurately for older than younger targets (see Table 2 and Figure 2, panel a). Furthermore, the descriptive comparisons with chance revealed that the accuracy with which health ratings predicted health indices for older faces was significantly greater than chance for doctors' health scores, $t(63) = 5.38, p < .001$, functional abilities, $t(31) = 13.21, p = .001$, and fitness, $t(31) = 3.50, p = .001$. Accuracy also differed from chance for younger faces' functional abilities, $t(31) = 2.63, p = .012$ and fitness, $t(31) = 3.69, p < .001$, but not for doctors' health scores, $t(63) = 0.90, p = .37$.

Competence ratings—Unlike health ratings, the accuracy of competence ratings was not greater for older than younger faces. Rather, competence ratings predicted IQ scores, reasoning, and STM of younger targets more accurately than older ones (see Table 2 and Figure 2, panel b). The descriptive comparisons to chance further revealed that the accuracy with which competence ratings predicted cognitive indices for younger faces was greater than chance for IQ scores, $t(63) = 3.25, p < .01$, reasoning, $t(31) = 8.32, p < .001$, and STM, $t(31) = 5.96, p < .001$. It was also greater than chance for vocabulary, $t(31) = 8.70$ and processing speed, $t(31) = 5.47$, both $ps < .001$. In the case of older faces, the accuracy with which competence ratings predicted cognitive indices was significantly greater than chance for vocabulary, $t(31) = 9.08, p < .001$, and reasoning, $t(31) = 4.63, p < .001$. However

²To minimize Type I error, we submitted younger and older raters' accuracy coefficients for the four MIDUS competence measures and the two MIDUS health measures to multivariate analyses (MANOVAs). The MANOVAs revealed significant overall accuracy, face age effects, and face age \times attractiveness effects. They did not reveal significant effects of rater age, rater age \times face age, or attractiveness \times rater age, and neither did the univariate ANOVAs on the IGS data. We therefore do not elaborate any of the latter effects in the text.

accuracy did not differ significantly from chance when predicting older targets' processing speed, $t(31) = 1.91, p = .095$, or STM, $t(31) = 1.16, p = .25$, and accuracy predicting their IQ scores was significantly worse than chance, indicating that older targets who looked more competent actually had lower IQs, $t(63) = -2.44, p = .02$.

Effects of Face Attractiveness on Accuracy

Health ratings—As predicted, health ratings more accurately predicted doctors' health scores and fitness when faces were below than above the median in attractiveness (see Table 3). Furthermore, the accuracy coefficient for doctors' health scores differed significantly from zero for faces below the median in attractiveness, $t(63) = 4.61, p < .001$, but not above the median, $t(63) = .24, p = .81$. Similarly, the accuracy coefficient for fitness differed significantly from zero for faces below the median in attractiveness, $t(31) = 3.93, p < .001$, but not above the median, $t(31) = .67, p = .51$. However, contrary to prediction, health ratings more accurately predicted functional abilities when faces were above than below the median in attractiveness, and the accuracy coefficient for functional abilities differed significantly from zero for faces above the median, $t(31) = 8.56, p < .001$, but not below the median, $t(31) = 1.11, p = .27$ (see Table 3).

Competence ratings—Competence ratings were more accurate for faces below than above the median in attractiveness for three of the five competence indices. Specifically, compared with ratings of faces above the median in attractiveness, competence ratings of faces below the median more accurately predicted reasoning ability, processing speed, and STM, with no significant effect for IQ scores or vocabulary (see Table 3). In addition, descriptive comparisons to chance revealed that when faces were below the median in attractiveness, the accuracy coefficients were greater than chance for reasoning ability, $t(15) = 3.79, p < .001$, processing speed $t(31) = 2.95, p < .01$, and STM, $t(15) = 2.72, p = .01$, as well as vocabulary, $t(15) = 2.52, p = .02$. When faces were above the median, accuracy was greater than chance only for vocabulary, $t(15) = 2.42, p = .02$. Accuracy predicting IQ scores did not differ significantly from chance for faces below or above the median in attractiveness, both $t_s < 1$.

Moderation of Attractiveness Effects by Rater Age

As shown in Table 3, there were no significant attractiveness level by rater age effects on accuracy, contrary to the prediction that greater accuracy for faces below than above the median in attractiveness would be weaker for older than younger raters.

Moderation of Attractiveness Effects by Face Age

Health ratings—The prediction that greater accuracy for faces below than above the median in attractiveness would be stronger for younger faces was supported for the fitness scores, which revealed a significant attractiveness level by face age effect (see Table 3). Planned comparisons revealed that the greater overall accuracy of health ratings when predicting fitness for faces below than above the median in attractiveness was significant for younger faces ($M_{\text{below}} = .12, SE = .03; M_{\text{above}} = -.07, SE = .03, p < .001$), but not for older faces ($M_{\text{below}} = .06, SE = .03, M_{\text{above}} = .09, SE = .03, p = .44$). The accuracy with which health ratings predicted fitness was significantly greater than chance when younger faces

were below the median in attractiveness, $t(31) = 3.52, p < .01$, and significantly worse than chance when faces were above the median in attractiveness, $t(31) = 2.61, p = .014$, indicating that for faces varying from average to high attractiveness, those that were judged as healthier actually had lower fitness. For older faces, on the other hand, the accuracy with which health ratings predicted fitness was significantly greater than chance when they were above the median in attractiveness $t(31) = 3.01, p < .01$, but not when they were below the median, $t(31) = 1.91, p = .065$ (Figure 3, panels a and b).

Competence ratings—The prediction that greater accuracy for faces below than above the median in attractiveness would be stronger for younger than older faces was supported by the reasoning and processing speed scores, which revealed significant attractiveness level by face age effects (see Table 3). Planned comparisons revealed that the greater overall accuracy of competence ratings when predicting reasoning for faces below than above the median in attractiveness held true for younger faces, $p < .001$, but not older faces, $p = .75$, as predicted. Similarly, the greater overall accuracy predicting processing speed for faces below than above the median in attractiveness also held true for younger faces, $p < .001$, but not older faces, $p = .20$.³ The moderation by face age was also shown in the descriptive comparisons of accuracy coefficients to chance. Competence ratings predicted reasoning with above chance accuracy only for younger faces below the median in attractiveness, $t(31) = 8.54, p < .001$, with no difference from chance for older faces below the median, $t(31) = -.30, p = .77$, younger faces above the median, $t(31) = .26, p = .80$, or older faces above the median, $t(31) = -.53, p = .60$, as predicted. Competence ratings also predicted processing speed with better than chance accuracy for younger faces below the median, $t(31) = 6.73, p < .001$, with no difference from chance for older faces below the median $t(31) = .11, p = .91$. In contrast, ratings predicted processing speed for older faces above the median with greater than chance accuracy, $t(31) = 2.15, p = .038$, while predicting for younger faces above the median with significantly *less* than chance accuracy $t(31) = -2.62, p = .013$, indicating that among younger faces varying from average to high attractiveness, those that looked more competent actually had slower processing speed (see Table 3 and Figure 3, panels c and d).⁴

Lens Models

The aim of the lens models was to ascertain whether variations in the accuracy of impressions could be explained by variations in the facial cues that are utilized when judging health and competence and/or those that are valid indicators of these traits. We therefore present lens models separately for older and younger faces but not for older and

³The tendency for accuracy when predicting processing speed from competence ratings to be greater for younger faces below than above the median, with no such effect for older faces, was qualified by a triple order interaction between attractiveness level, face age, and rater age, $F(1, 30) = 5.58, p = .025$. Both older and younger raters showed more accuracy for younger faces below than above the median in attractiveness (Older raters: $M_{\text{below}} = .20, SE = .03; M_{\text{above}} = -.11, SE = .05, p < .001$; Younger raters: $M_{\text{below}} = .12, SE = .03, M_{\text{above}} = -.07, SE = .05; t(15) = 3.50, p < .01$). For older faces, on the other hand, older raters were more accurate when rating faces above than below the median ($M_{\text{below}} = -.07, SE = .05; M_{\text{above}} = .06, SE = .05, t(15) = 1.79, p < .05$), whereas younger raters showed no difference in accuracy for older faces below versus above the median ($M_{\text{below}} = .11, SE = .05; M_{\text{above}} = .10, SE = .05, t(15) = .21, p = .83$).

⁴We also investigated whether the similarity of faces' age to perceivers' own age would influence the accuracy of impressions, since previous research has revealed that older raters show better performance for older than younger faces in face recognition (Anastasi & Rhodes, 2005; Fulton & Bartlett, 1991), age recognition (Voelkle, Ebner, Lindenberger, & Riediger, 2012), emotion recognition (Malatesta, Izard, Culver, & Nicolich, 1987), and recognition of criminals in lineups (Wright & Stroud, 2002). There were no significant rater age \times face age effects for the accuracy of health or competence ratings.

younger raters, because there was no variation in accuracy as a function of rater age. Cue utilization coefficients were computed by correlating each participant's ratings of the health or competence of targets' faces with mean ratings of the targets on each facial cue (attractiveness, positive expression, old-looking), controlling face sex and actual face age to parallel the accuracy analyses. We controlled these variables to insure that utilization reflected sensitivity to the facial cues per se rather than sensitivity to differences in the facial cues provided by men and women or older versus younger people within each age group. The resulting cue utilization correlations for each rater were averaged, and the mean coefficients for older and younger faces are shown in Figures 4 and 5. Cue validity correlations were computed by correlating the mean ratings of each face on a particular cue and the mean scores for each face on the health or competence criterion measures. These correlations also controlled face sex and actual face age to ensure that cue validity reflected validity of the facial cues per se rather than validity attributable to covariation of the cue with face sex or actual age variations within each age group. It should be recalled that controlling actual face age when assessing the cue validity and utilization of apparent age reveals whether looking older than one's age is associated with perceived health and competence as well as actual health and cognitive scores. Finally, the 'Accuracy' coefficients shown in Figures 4 and 5 are identical to those presented for older and younger faces in Table 2.

Attractiveness—Attractiveness was not a valid cue to any of the health measures for younger faces. However, it was significantly utilized in health ratings, which may have hindered accuracy. In contrast to younger faces, greater attractiveness was a valid cue to the functional abilities and fitness of older faces, and health ratings utilized this valid cue, which may have contributed to their accuracy (see Figure 4). Whereas attractiveness was not a valid cue when predicting the health of younger faces, it was a valid predictor of their vocabulary, reasoning, and STM, and competence ratings were correlated with this cue, which may have helped raters to achieve accuracy. In the case of older faces, attractiveness was a valid predictor only of vocabulary. Again, competence ratings utilized this cue, which may have contributed to accuracy (see Figure 5).

Positive expression—A more positive expression was not a valid cue to any health measures for younger or older faces, although health ratings utilized it, which may have hindered accuracy (see Figure 3). In contrast to health outcomes, a more positive expression was a valid cue to faster processing speed and better STM in younger faces, although the former effect was only marginally significant. Competence ratings utilized this cue, which may have contributed to their accuracy. In the case of older faces, a more positive expression was a marginally significant valid cue only for reasoning. Competence ratings of older faces utilized this cue, which may have contributed to accuracy (see Figure 5).

Older-looking—Looking older than one's age was not a valid cue to any health measures for younger faces, although it was utilized, which may have hindered accuracy. In contrast, looking older was a valid cue to lower functional abilities and fitness of older faces, and health ratings utilized this valid cue, which may have helped accuracy (see Figure 4). Looking older also was a valid cue to the vocabulary, reasoning, and STM of younger faces

in the Boston MIDUS sample. However, competence ratings of these faces did not utilize apparent age, which may have depressed accuracy even though it was significant. In contrast competence ratings of the younger faces in the IGS sample did utilize apparent age, with older-looking people judged more competent, probably because faces in this group were only 17–18 years old. This may have depressed accuracy in predicting IQ scores, since apparent age was not a valid cue to IQ (see Figure 5). Looking older did not predict any competence criterion scores for older faces, although this cue was consistently utilized, which may have depressed accuracy.

Summary—Greater attractiveness and a younger-looking face each provided valid cues to health measures for older faces. Health ratings were responsive to these cues, consistent with their accuracy predicting these health outcomes. None of the cues we assessed provided valid indicators of the actual health of younger faces, which is consistent with the failure of health ratings to show accuracy for younger faces. Attractiveness and a more positive expression were valid cues to cognitive abilities in both younger and older faces. Attractiveness and a more positive expression were also correlated with competence ratings. Utilization of one or more of these cues was associated with accuracy of several competence ratings of younger faces, but only when predicting vocabulary of older faces. The failures of competence ratings to achieve accuracy for older faces may have been attributable in part to the utilization of an older-looking appearance even though that cue was not a valid predictor.

Discussion

Consistent with an ecological approach to perception (Gibson, 1979; McArthur & Baron, 1983; Zebrowitz, 2011), which holds that perceptions of adaptively relevant attributes will be accurate, ratings of health and competence from facial photographs predicted multiple indices of the targets' actual health and competence with greater than chance accuracy. Although the accuracy effect sizes were modest, this may indeed be a situation in which “small effects are impressive” given the limited amount of information still photographs provided to perceivers (Prentice & Miller, 1992). Ratings by older and younger raters predicted indices of actual health and competence with equal accuracy despite poorer performance by older raters on many control measures, as shown in Table 1. On the other hand, accuracy was moderated by face age. Whereas health ratings were more accurate for older faces, as predicted, competence ratings tended to be more accurate for younger faces. In addition, the accuracy of health and competence ratings was generally greater for faces below than above the median in attractiveness, as predicted. However, the greater accuracy for faces below than above the median was not weaker for older raters, contrary to previous evidence that they are less sensitive to negatively valenced cues. Unlike rater age, face age did moderate some of the effects of attractiveness level, as predicted, with greater accuracy when rating younger faces below than above the median in attractiveness, but not older faces.

The absence of any age-related decline in the ability to accurately discern the relative health and competence of people based merely on their facial appearance indicates that the adaptive value of knowing who might be unhealthy, disabled, or weak in cognitive capacities is preserved in older adulthood. These results further indicate that the well-

documented older adult positivity effect (Mather & Carstensen, 2005; Murphy & Isaacowitz, 2008; Reed, Chan, & Mikels, 2014), does not reduce the accuracy of first impressions of health and competence. Indeed, health ratings by older raters were just as accurate as those of younger raters in the present study even though previous research using the same faces and raters found more positive health ratings by older than younger raters (Zebrowitz et al., 2013).

Our finding of no age-related decline in the ability to detect the relative health and competence of people based merely on their facial appearance contrasts with evidence that older raters are less accurate than younger raters when judging people's deceitfulness (Stanley & Blanchard-Fields, 2008) or social gaffes (Halberstadt et al., 2011), effects attributed to older adults' lower likelihood of using negatively valenced emotion cues in their judgments. On the other hand, the present findings are consistent with recent research demonstrating that older and younger adults' ratings of the aggressiveness of faces not only predict the targets' actual aggressive behavior with equal accuracy, but also make equal use of facial resemblance to anger and low attractiveness (Boshyan et al., 2013). What may account for variations in older raters' accuracy across the studies are differences in the nature of the tasks. Whereas the studies examining age differences in the detection of deception and social gaffes likely engaged controlled processing of the social information, as raters tried to make an accurate judgment, the studies asking raters to judge the aggressiveness, health, or competence of the faces likely engaged automatic processing which is implicated in younger adults' trait impressions from faces (Bar, Neta, & Linz, 2006; Carré, McCormick, & Mondloch, 2009; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006) and has been shown to be relatively preserved in aging (Mather & Knight, 2006). Indeed, participants in the present study were instructed to just give their first impression of the faces, and they were told there were no right or wrong answers.

Our prediction that there would be greater accuracy for older than younger faces was supported in the case of health ratings, which predicted doctors' health scores, functional abilities, and fitness more accurately for older than younger faces. In contrast, competence ratings predicted IQ scores, reasoning, and STM more accurately for younger than older faces. Although accuracy predicting vocabulary and processing speed didn't vary significantly with face age, the prediction of reasoning differed from chance only for younger faces. The face age effects are particularly notable for doctors' health and IQ scores, which were assessed for the same individuals at two different ages. While raters could discern variations in the general intelligence of people when they were 17–18 years of age, they could not do so when the same people were in their 50s and 60s, and they could better discern variations in the same people's health in older adulthood than at 17–18.⁵

Face age differences in the accuracy of health and competence ratings may reflect more valid cues in faces of one age than the other, an increase in the use of valid cues, or both. Both attractiveness and looking older for one's age were utilized in health ratings of older faces as well as younger ones. However, they provided valid cues to actual health only for

⁵Although we have reported the doctors' health score results using all of the available data, the face age effect remains significant when using data only from the 70 participants who had health data in both younger and older adulthood, $F(1, 61) = 4.78, p = .033$.

older faces, consistent with previous evidence that older-looking older adults are actually less healthy and die sooner than younger-looking individuals of the same age (Christensen et al., 2009). The utilization of these cues when judging health of younger faces despite their lack of validity is consistent with the conclusion that raters may be ‘blinded by beauty’ when judging the health of younger faces (Kalick et al., 1998). Thus, the greater validity of the cues that were utilized when judging the health of older faces may explain the greater accuracy of these ratings. The fact that utilized cues were more valid predictors of older than younger faces’ health and the corresponding greater accuracy of health ratings also may reflect differences in the kind of health problems experienced by the two age groups. For example, variations in chronic illness are likely to be more common among older targets, and it may be that such variations are more evident in facial appearance than whatever variations in health exist among younger targets.⁶

In the case of competence ratings, both attractiveness and a positive expression were utilized when rating both older and younger faces, and these cues were often valid predictors of cognitive indices for faces from both age groups, although some effects for positive expression were only marginally significant. Moreover, whenever one of these cues was valid, competence ratings accurately predicted a criterion measure of competence, whether for older or younger faces. In addition, the finding that competence ratings were accurate predictors of processing speed and short-term memory for younger, but not older, faces was paralleled by validity of attractiveness and/or expression cues for younger, but not older faces. Thus, lesser validity of attractiveness and positive expression when predicting these cognitive abilities in older faces may have contributed to the absence of accuracy for competence ratings. Another contribution to inaccuracy predicting these measures for older faces may be the utilization of an invalid cue. Specifically, older-looking older faces were given lower competence ratings even though looking older was not a valid cue to cognitive scores. It is thus possible that the lower accuracy of competence ratings for older faces was attributable in part to raters’ being misled by an older-looking appearance.

In evaluating the results shown in the Brunswikian lens models, it should be noted that Brunswik argued that it is important to use an ecologically representative sample of stimuli when assessing cue utilization, cue validity, and accuracy (Brunswik, 1955). Thus, it is possible, for example, that a different sample of younger faces would show that attractiveness is a valid cue to health, as it was to the health of older faces in our study or that health can be as accurately discerned in younger faces as older ones or that competence can be as accurately discerned in older faces as younger ones. The fact that attractiveness was a valid cue when predicting health measures for older MIDUS faces, but not for older IGS faces, is consistent with the suggestion that the sample can make a difference. However, as described in the Method section, both the IGS and the MIDUS samples of faces are arguably highly representative. Thus, the differential validity of cues cross the two samples of older faces may reflect the younger mean age of older IGS faces than older MIDUS faces, the fact that the MIDUS faces were in color whereas IGS faces were in black and white, or

⁶We thank an anonymous reviewer for making this suggestion.

differences in the criterion measures. These possibilities should be explored in future research.

The one exception to the tendency for competence ratings to be more accurate for younger than older faces was vocabulary scores, where accuracy was equal and significant across face age. It is noteworthy that this was also the only cognitive measure for which attractiveness was a valid cue for older as well as younger faces. This result suggests that attractiveness conveys 'crystallized' abilities in both older and younger faces, whereas only in younger faces does it convey 'fluid' abilities, as captured by reasoning and STM. This result suggests an important qualification to the evolutionary psychology argument that attractiveness advertises genetic quality (Rhodes, 2006). Whereas fluid abilities are arguably an index of genetic quality, crystallized abilities are associated with lifelong learning (Horn & Cattell, 1966, 1967). Thus, the finding that attractiveness predicted fluid abilities for younger but not older faces suggests that attractiveness is more strongly associated with genetic quality in younger adults, consistent with our suggestion that factors unrelated to fitness may contribute more to low attractiveness in older faces, such as skin changes that are not associated with pathology (Pariser, 1998). In addition, the fact that attractiveness predicts crystallized abilities in both age groups is consistent with evidence that more attractive people are advantaged in the educational system and have more enriching life experiences (for reviews see Langlois et al., 2000; Zebrowitz & Montepare, 2013).

The finding that health and competence ratings were generally more accurate for younger faces below than above the median in attractiveness, whereas this effect was less consistently found for older faces also supports the argument that variations in attractiveness are more linked to genetic fitness in younger than in older adults. More specifically, greater accuracy for younger faces below the median supports the 'bad genes' hypothesis that, compared with faces of average attractiveness, those low in attractiveness may signal genetic anomalies associated with low fitness (Foroud et al., 2012; Cummings et al., 1982). The failure to find comparable effects for older faces is consistent with the suggestion that factors unrelated to fitness may contribute to low attractiveness in older faces more so than in younger ones. An exception to the tendency for accuracy in judging younger faces to be greater when faces were below than above the median in attractiveness was found when predicting functional abilities from health ratings, in which case, accuracy was greater for younger faces above than below the median just as it was for older faces. We do not have an explanation for this unexpected finding.

In summary, face age, but not rater age, moderated the significant accuracy shown by health and competence ratings. Variations in health were more accurately discerned in older than younger faces, with the reverse true for variations in competence. The former effect reflected the dearth of valid cues to variations in the health of younger targets, and the latter effect seemed to reflect the use of the misleading cue of an older-looking face when rating the competence of older faces. Although older targets who looked older for their age had poorer health than their younger-looking peers, they were no less competent. Level of attractiveness of the faces also moderated accuracy, with ratings of health and competence generally showing greater accuracy for faces below than above the median in attractiveness. As was true for overall accuracy, these effects were moderated by face age, but not rater age.

Raters were generally more accurate in judging the health and competence of younger faces below than above the median in attractiveness, whereas this pattern was not shown for older faces. Whereas research investigating younger adult faces suggests that very low attractiveness may be indicative of low fitness, our results suggest that factors unrelated to fitness may contribute to low attractiveness in older faces. Overall, the findings show that both younger and older adults are able to discern important information regarding health and competence from a brief look at faces. The significant correspondence between ratings and actual measures is both encouraging and enlightening. Yet, there also was a failure to achieve accuracy when judging the health of younger people and the competence of older people that could be linked to a dearth of relevant cues or utilization of irrelevant ones. The implications of variations in the accuracy of these appearance-based judgments for interactions in social, work, and professional settings should be explored in future studies.

Acknowledgments

We express appreciation to the Institute of Human Development at the University of California, Berkeley for providing access to the Intergenerational Studies (IGS) data archives used in this study. These archives are now housed at the Henry A. Murray Research Archive, Harvard University, Cambridge, MA. This research was supported by National Institutes of Health Grant AG38375 to the first author.

References

- Alosco ML, Brickman AM, Spitznagel MB, van Dulmen M, Raz N, Cohen R, Gunstad J. The independent association of hypertension with cognitive function among older adults with heart failure. *Journal of the Neurological Sciences*. 2012; 323:216–220.10.1016/j.jns.2012.09.019 [PubMed: 23026535]
- Anastasi JS, Rhodes MG. An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review*. 2005; 12:1043–1047.10.3758/BF03206441 [PubMed: 16615326]
- Bar M, Neta M, Linz H. Very first impressions. *Emotion*. 2006; 6:269–278.10.1037/1528-3542.6.2.269 [PubMed: 16768559]
- Bayer, LM.; Whissell-Buechy, D.; Honzik, MP. Health in the middle years. In: Eichorn, DH.; Clausen, HJA.; Haan, N.; Honzik, MP.; Mussen, PH., editors. *Present and past in middle life*. New York, NY: Academic Press; 1981. p. 55-88.
- Bayer LM, Snyder MM. Illness experience of a group of normal children. *Child Development*. 1950; 21:93–120. [PubMed: 15434876]
- Benedict C, Brooks SJ, Kullberg J, Nordenskjold R, Burgos J, Le Greves M, Schiöth HB. Association between physical activity and brain health in older adults. *Neurobiology of Aging*. 2013; 34:83–90.10.1016/j.neurobiolaging.2012.04.013 [PubMed: 22592017]
- Block, J. *Lives through time*. Berkeley, CA: Bancroft; 1971.
- Boshyan J, Zebrowitz LA, Franklin RG Jr, McCormick CM, Carré JM. Age similarities in recognizing threat from faces and diagnostic cues. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*. 2013 Advance online publication. 10.1093/geronb/gbt054
- Brand A, Bradley MT. More voodoo correlations: When average-based measures inflate correlations. *Journal of General Psychology*. 2012; 139:260–272.10.1080/00221309.2012.703711 [PubMed: 24837177]
- Brunswik E. Representative design and probabilistic theory in a functional psychology. *Psychological Review*. 1955; 62:193–217.10.1037/h0047470 [PubMed: 14371898]
- Carré JM, McCormick CM, Mondloch CJ. Facial structure is a reliable cue of aggressive behavior. *Psychological Science*. 2009; 20:1194–1198.10.1111/j.1467-9280.2009.02423.x [PubMed: 19686297]

- Christensen K, Thinggaard M, McGue M, Rexbye H, Hjelmborg JV, Aviv A, Vaupel JW. Christmas 2009: Young and old: Perceived age as clinically useful biomarker of aging cohort study. *British Medical Journal*. 2009; 339:Article Number: B5262.10.1136/bmj.b5262
- Cummings CD, Flynn D, Preus M. Increased morphological variants in children with learning disabilities. *Journal of Autism and Developmental Disorders*. 1982; 12:373–383.10.1007/BF01538325 [PubMed: 7161238]
- Daniels, K.; Toth, J.; Jacoby, L. The aging of executive functions. In: Bialystok, E.; Craik, FIM., editors. *Lifespan cognition: Mechanisms of change*. New York, NY: Oxford University Press; 2006. p. 96-111.
- Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12:189–198.10.1016/0022-3956(75)90026-6 [PubMed: 1202204]
- Foroud T, Wetherill L, Vinci-Booher S, Moore ES, Ward RE, Hoyme HE, Jacobson SW. Relation over time between facial measurements and cognitive Outcomes in fetal alcohol-exposed children. *Alcoholism: Clinical and Experimental Research*. 2012; 36:1634–1646.10.1111/j.1530-0277.2012.01750.x
- Franklin RG, Adams RB. A dual-process account of female facial attractiveness preferences: Sexual and nonsexual routes. *Journal of Experimental Social Psychology*. 2009; 45:1156–1159.10.1016/j.jesp.2009.06.014
- Franklin RG Jr, Zebrowitz LA. Sensitivity of trait impressions to subtle facial resemblance to emotions is preserved in healthy aging. *Journal of Nonverbal Behavior*. 2013; 37:139–151.10.1007/s10919-013-0150-4 [PubMed: 24058225]
- Fulton A, Bartlett JC. Young and old faces in young and old heads: The factor of age in face recognition. *Psychology and Aging*. 1991; 6:623–630.10.1037/0882-7974.6.4.623 [PubMed: 1777151]
- Gibson, JJ. *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin; 1979.
- Halberstadt J, Ruffman T, Murray J, Taumoepeau M, Ryan M. Emotion perception explains age-related differences in the perception of social gaffes. *Psychology and Aging*. 2011; 26:133–136.10.1037/a0021366 [PubMed: 21280951]
- Horn JL, Cattell RB. Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*. 1966; 57:253–270.10.1037/h0023816 [PubMed: 5918295]
- Horn JL, Cattell RB. Age differences in fluid and crystallized intelligence. *Acta Psychologica*. 1967; 26:107–129.10.1016/0001-6918(67)90011-X [PubMed: 6037305]
- Kalick SM, Zebrowitz LA, Langlois JH, Johnson RM. Does human facial attractiveness honestly advertise health? Longitudinal data on an evolutionary question. *Psychological Science*. 1998; 9:8–13.10.1111/1467-9280.00002
- Kotter-Grühn D, Hess TM. So you think you look young? Matching older adults' subjective ages with age estimations provided by younger, middle-aged, and older adults. *International Journal of Behavioral Development*. 2012; 36:468–475.10.1177/0165025412454029
- Lachman, ME. *Midlife Development in the United States (MIDUS): Boston Study of Management Processes, 1995–1997*. Inter-university consortium for Political and Social Research, University of Michigan; 1997. <http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/203/studies/3596?q=MIDUS&searchIn=ALL&paging.startRow=1>
- Lachman ME, Agrigoroaei S. Promoting functional health in midlife and old age: Long-term protective effects of control beliefs, social support, and physical exercise. *PLoS ONE*. 2010; 5:e13297.10.1371/journal.pone.0013297 [PubMed: 20949016]
- Langlois JH, Kalakanis L, Rubenstein AJ, Larson A, Hallam M, Smoot M. Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*. 2000; 126:390–423.10.1037/0033-2909.126.3.390 [PubMed: 10825783]
- Luevano, VX. Unpublished dissertation. Brandeis University; 2007. Truth in advertising: The relationship of facial appearance to apparent and actual health across the lifespan.
- Malatesta CZ, Izard CE, Culver C, Nicolich M. Emotion communication skills in young, middle-aged, and older women. *Psychology and Aging*. 1987; 2:193–203.10.1037/0882-7974.2.2.193 [PubMed: 3268208]

- Mather M, Carstensen LL. Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences*. 2005; 9:496–502.10.1016/j.tics.2005.08.005 [PubMed: 16154382]
- Mather M, Knight MR. Angry faces get noticed quickly: Threat detection is not impaired among older adults. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*. 2006; 61:54–57.10.1093/geronb/61.1.P54
- McArthur LZ, Baron RM. Toward an ecological theory of social perception. *Psychological Review*. 1983; 90:215–238.10.1037/0033-295X.90.3.215
- Miller LMS, Lachman ME. Cognitive performance and the role of control beliefs in midlife. *Aging, Neuropsychology, and Cognition*. 2000; 7:69–85.10.1076/1382-5585(200006)7:2;1-U;FT069
- Murphy NA, Isaacowitz DM. Preferences for emotional information in older and younger adults: A meta-analysis of memory and attention tasks. *Psychology and Aging*. 2008; 23:263–286.10.1037/0882-7974.23.2.263 [PubMed: 18573002]
- Pariser RJ. Benign neoplasms of the skin. *The Medical Clinics of North American*. 1998; 82:1285–1307.10.1016/S0025-7125(05)70416-8
- Prentice DA, Miller DT. When small effects are impressive. *Psychological Bulletin*. 1992; 112 s:160–164.
- Raven, J.; Raven, JC.; Court, JH. *Manual for Raven's progressive matrices and vocabulary, scales: Section 1*. Oxford, UK: Oxford Psychologists; 1991.
- Reed AE, Chan L, Mikels JA. Meta-analysis of the age-related positivity effect: Age differences in preferences for positive over negative information. *Psychology and Aging*. 2014; 29:1–15.10.1037/a0035194 [PubMed: 24660792]
- Rhodes G. The evolutionary psychology of facial beauty. *Annual Review of Psychology*. 2006; 57:199–226.10.1146/annurev.psych.57.102904.190208
- Salthouse TA. The processing-speed theory of adult age differences in cognition. *Psychological Review*. 1996; 103:403–428. [PubMed: 8759042]
- Salthouse TA, Babcock RL. Decomposing adult age differences in working memory. *Developmental Psychology*. 1991; 27:763–776.10.1037/0012-1649.27.5.763
- Schaie, KW. *Manual for the Schaie-Thurstone Adult Mental Abilities Test (STAMAT)*. Palo Alto, CA: Consulting Psychologists Press; 1985.
- Stanley JT, Blanchard-Fields F. Challenges older adults face in detecting deceit: The role of emotion recognition. *Psychology and Aging*. 2008; 23:24–32.10.1037/0882-7974.23.1.24 [PubMed: 18361651]
- Syddall HE, Martin HJ, Harwood RH, Cooper C, Sayer AA. The SF-36: A simple, effective measure of mobility-disability for epidemiological studies. *Journal of Nutrition, Health, and Aging*. 2009; 13:57–62.10.1007/s12603-009-0010-4
- Todorov A, Pakrashi M, Oosterhof NN. Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*. 2009; 27:813–833.10.1521/soco.2009.27.6.813
- Voelkle MC, Ebner NC, Lindenberger U, Riediger M. Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age. *Psychology and Aging*. 2012; 27:265–277.10.1037/a0025065 [PubMed: 21895379]
- Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): Conceptual framework and item selection. *Medical Care*. 1992; 30:473–483.10.1097/00005650-199206000-00002 [PubMed: 1593914]
- Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*. 1988; 54:1063–1070.10.1037/0022-3514.54.6.1063 [PubMed: 3397865]
- Wechsler, D. *Wechsler Adult Intelligence Scale-III (WAIS-III) manual*. New York, NY: The Psychological Corporation; 1997.
- Willis J, Todorov A. First impressions: Making up your mind after 100 ms exposure to a face. *Psychological Science*. 2006; 17:592–598.10.1111/j.1467-9280.2006.01750.x [PubMed: 16866745]

- Wright DB, Stroud JN. Age differences in lineup identification accuracy: People are better with their own age. *Law and Human Behavior*. 2002; 26:641–654.10.1023/A:1020981501383 [PubMed: 12508699]
- Zebrowitz, LA. Ecological and social approaches to face perception. In: Calder, AJ.; Rhodes, G.; Haxby, JV.; Johnson, Mark H., editors. *The handbook of face perception*. Oxford, UK: Oxford University Press; 2011. p. 31-50.
- Zebrowitz LA, Fellous JM, Mignault A, Andreoletti C. Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from Connectionist modeling. *Personality and Social Psychology Review*. 2003; 7:194–215.10.1207/S15327957PSPR0703_01 [PubMed: 12788687]
- Zebrowitz LA, Franklin RG, Hillman S, Boc H. Older and younger adults' first impressions from faces: Similar in agreement but different in positivity. *Psychology and Aging*. 2013; 28:202–212.10.1037/a0030927 [PubMed: 23276216]
- Zebrowitz LA, Hall JA, Murphy NA, Rhodes G. Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin*. 2002; 28:238–249.10.1177/0146167202282009
- Zebrowitz, LA.; Montepare, JM. Faces and first impressions. In: Bargh, J.; Borgida, G., editors. *Handbook of personality and social psychology, Vol. 1: Attitudes and social cognition*. Washington, DC: American Psychological Association; 2013.
- Zebrowitz LA, Rhodes G. Sensitivity to 'bad genes' and the anomalous face overgeneralization effect: Accuracy, cue validity, and cue utilization in judging intelligence and health. *Journal of Nonverbal Behavior*. 2004; 28:167–185.10.1023/B:JONB.0000039648.30935.1b
- Zebrowitz LA, Voinescu L, Collins MA. 'Wide eyed' and 'crooked-faced': Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin*. 1996; 22:1258–1269.10.1177/01461672962212006

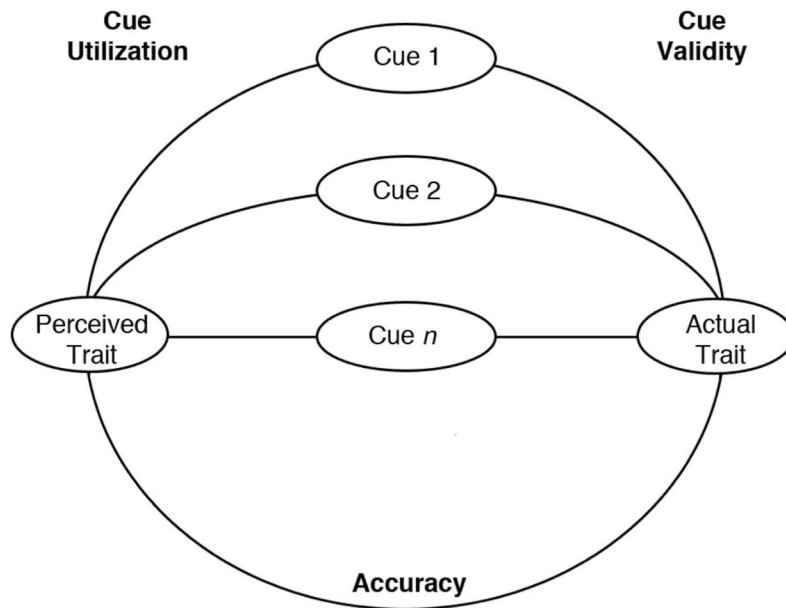


Figure 1. The Lens Model. Adapted from Brunswik (p. 206) with permission of the publisher. Copyright ©1955, American Psychological Association. Authorization for this adaptation has been obtained both from the owner of the copyright in the original work and from the owner of copyright in the translation or adaptation.

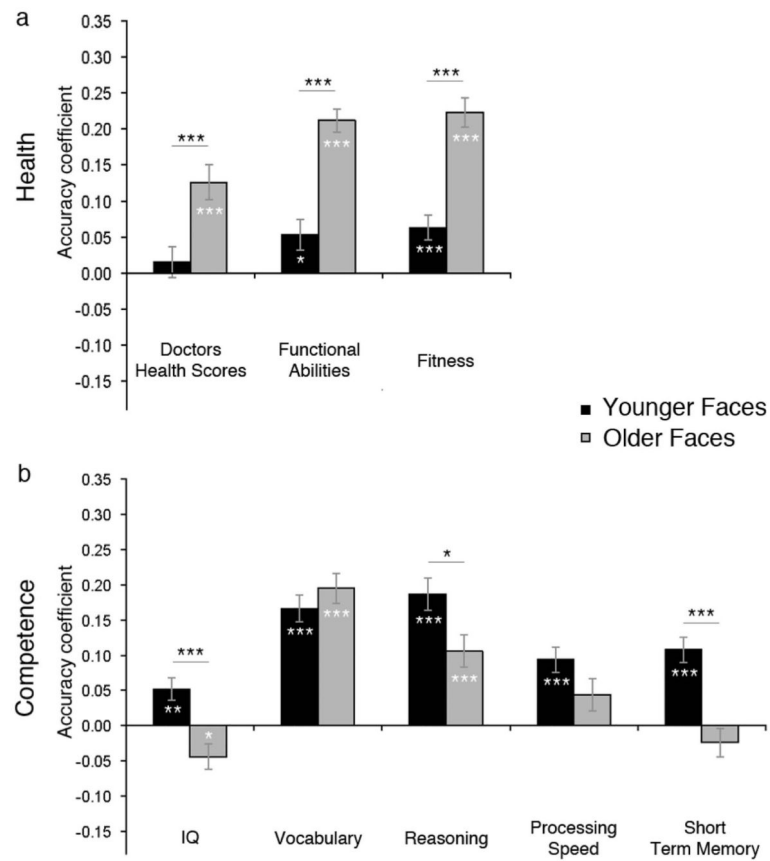


Figure 2.

Panel a shows accuracy with which ratings of health predicted indices of actual health for younger and older faces. Panel b show accuracy with which ratings of competence predicted indices of actual competence for younger and older faces. Asterisks in bars show significance of difference from chance accuracy; asterisks above bars show significance of face age differences.

* $p < .05$. ** $p < .01$. *** $p < .001$.

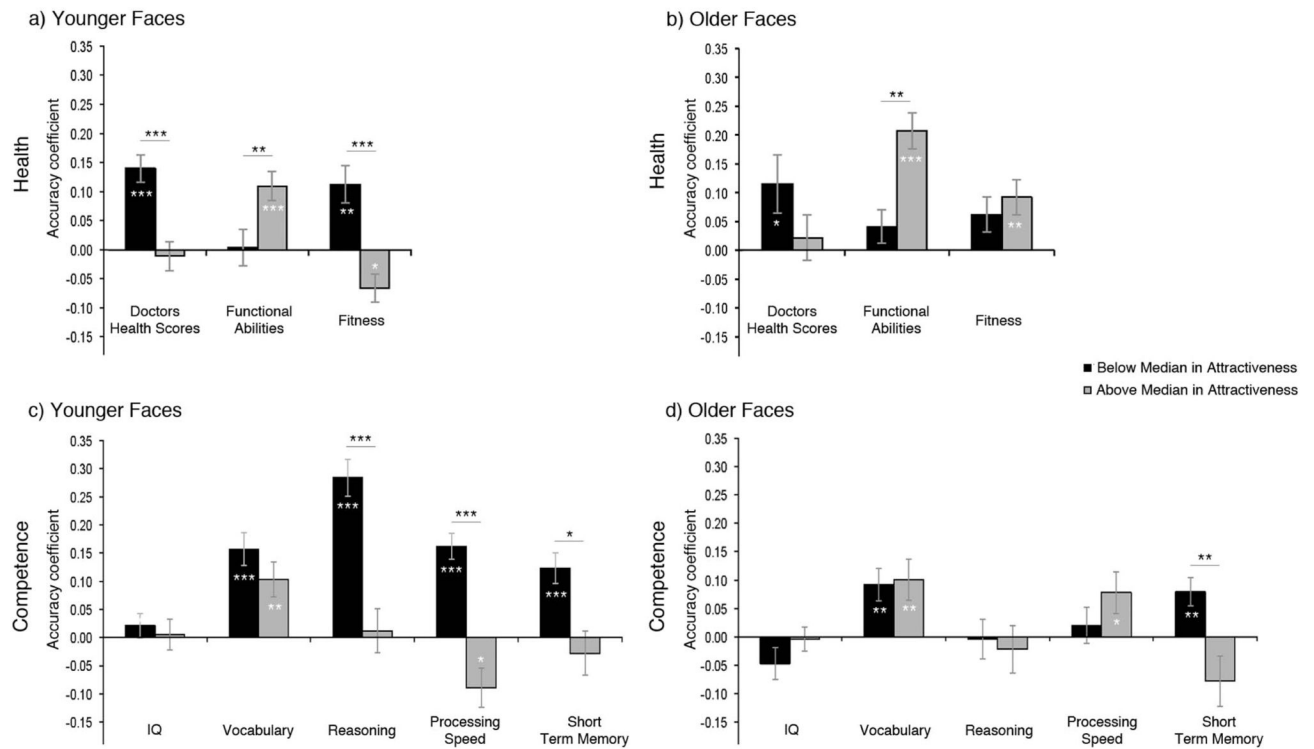


Figure 3.

Panel a (younger faces) and panel b (older faces) show accuracy with which health ratings of faces below and above the median in attractiveness predicted indices of actual health. Panel c (younger faces) and panel d (older faces) show accuracy with which competence ratings of faces below and above the median in attractiveness predicted indices of actual competence. Asterisks in bars show significance of difference from chance accuracy; asterisks above bars show significance of attractiveness level differences.

* $p < .05$. ** $p < .01$. *** $p < .001$.

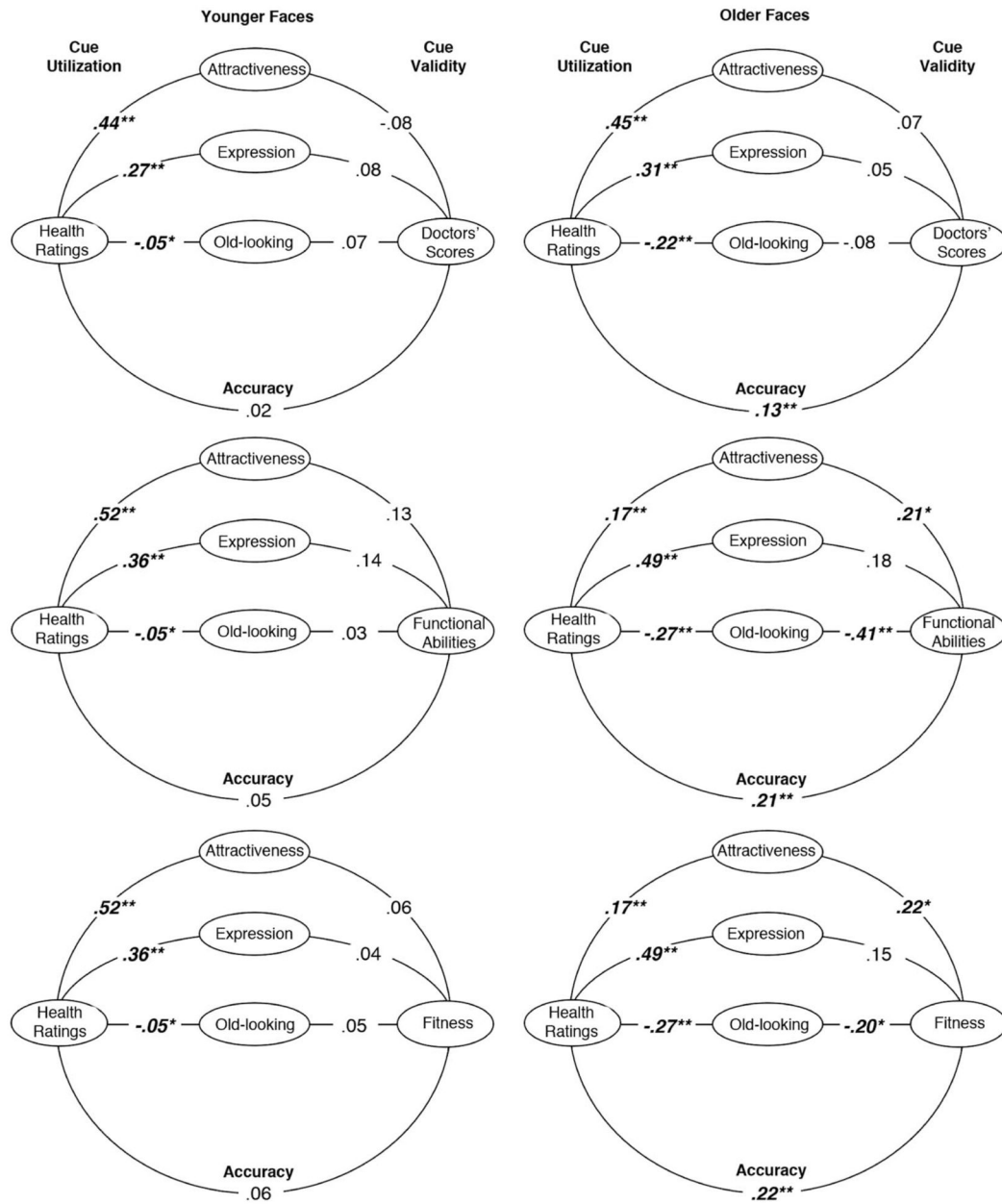


Figure 4. Cue utilization, cue validity, and accuracy in judging health from younger faces (models on left) and older faces (models on right). Cues are coded so that higher values signify greater attractiveness, more positive expression, and older-looking.
 * $p < .05$. ** $p < .01$.

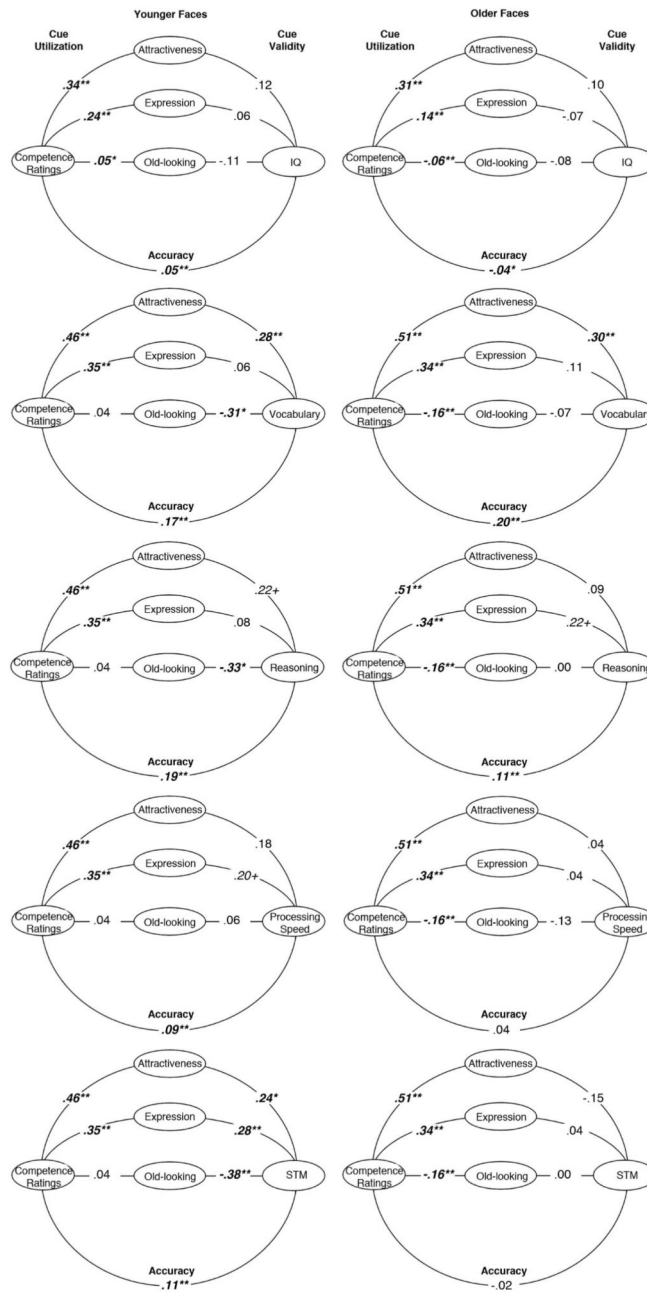


Figure 5. Cue utilization, cue validity, and accuracy in judging competence from younger faces (models on left) and older faces (models on right). Cues are coded so that higher values signify greater attractiveness, more positive expression, and older-looking. $+p < .10$. $*p < .05$. $**p < .01$.

Table 1

Older Raters (OR) and Younger Raters (YR) Scores on Control Measures

Measure	Younger raters			Older raters		
	M	SE	t value	M	SE	p value
PANAS (Positive Affect)	2.78	0.84	2.90	0.70	0.79	0.434
PANAS (Negative Affect)	1.92	0.66	1.73	0.67	1.37	0.173
Snellen Visual Acuity (denominator)	19.75	5.47	34.69	14.27	6.77	<.001
Mars Letter Contrast Sensitivity	1.73	0.06	1.56	0.19	8.87	<.001
Ishihara's Test for Color Deficiency	13.48	1.65	13.46	1.87	0.06	0.954
Benton Facial Recognition Test	42.21	8.58	41.35	9.31	0.47	0.641
Pattern Comparison Test	42.56	5.85	28.17	5.16	12.79	<.001
Shipley Vocabulary Test	31.77	3.89	35.48	3.37	4.99	<.001
Wisconsin Card Sorting Test	36.40	5.56	29.03	8.92	3.59	0.001
Letter-Number Sequencing Test	12.56	2.80	8.19	1.52	5.49	<.001
Education ^a	3	2-7	3	2-3		

^aLevel of Education was coded for highest level attained: 1 – no high school diploma, 2 – high school diploma, 3 – some college, 4 – Bachelor's degree, 5 – some graduate work, 6 – Masters degree, 7 – Doctorate degree. Medians and range are reported.

Table 2

Results of 2 Rater Age (RA) \times 2 Face Age (FA) ANOVAs on Accuracy Coefficients^a

Variable	Overall accuracy			YR			OR			Main effect of RA			YF			OF			Main effect of FA				
	F	P	M	F	P	M	F	P	M	F	P	M	F	P	M	F	P	M	F	P	M	F	P
Doctors' health score	19.65	<.001	0.07	0.02	0.08	0.02	0.01	.916	0.02	0.02	0.02	0.01	.916	0.02	0.02	0.02	0.13	0.02	0.13	0.02	0.02	13.73	<.001
Functional abilities	122.48	<.001	0.14	0.02	0.12	0.02	0.51	.482	0.05	0.02	0.02	0.51	.482	0.05	0.02	0.02	0.21	0.02	0.21	0.02	0.02	31.65	<.001
Physical fitness	107.58	<.001	0.16	0.02	0.13	0.02	0.71	.406	0.06	0.02	0.02	0.71	.406	0.06	0.02	0.02	0.22	0.02	0.22	0.02	0.02	41.15	<.001
IQ scores	0.12	.731	0.00	0.02	0.01	0.02	0.21	.649	0.05	0.02	0.02	0.21	.649	0.05	0.02	0.02	-0.04	0.02	-0.04	0.02	0.02	16.81	<.001
Vocabulary	115.13	<.001	0.16	0.02	0.20	0.02	1.83	.186	0.17	0.02	0.02	1.83	.186	0.17	0.02	0.02	0.20	0.02	0.20	0.02	0.02	1.72	.199
Reasoning	73.70	<.001	0.14	0.02	0.15	0.02	0.01	.916	0.19	0.02	0.02	0.01	.916	0.19	0.02	0.11	0.02	0.11	0.02	0.02	6.94	.013	
Processing speed	27.63	<.001	0.09	0.02	0.05	0.02	1.72	.200	0.09	0.02	0.02	1.72	.200	0.09	0.02	0.04	0.02	0.04	0.02	0.02	2.68	.112	
STM	8.71	<.01	0.06	0.02	0.03	0.02	0.89	.354	0.11	0.02	0.02	0.89	.354	0.11	0.02	-0.02	0.02	-0.02	0.02	0.02	25.07	<.001	

^aYR and OR refer to younger and older raters; YF and OF refer to younger and older faces. Mean accuracy coefficients are correlations between each participant's ratings of the younger or older faces with the targets' actual scores, controlling face sex and actual face age. For Doctors' health score and IQ scores, $df = 1, 62$; for all other variables, $df = 1, 30$.

Table 3

Results of 2 Rater Age (RA) \times 2 Face Age (FA) \times 2 (Face Attractiveness: Below/Above Median) ANOVAs on Accuracy Coefficients^a

Variable	Below median		Above median		Main effect of median		Attractiveness by RA		Attractiveness by FA	
	M	SE	M	SE	F	P	F	P	F	P
Doctors' health score	0.13	0.03	0.01	0.03	13.79	<.001	0.70	.406	0.91	.343
Functional abilities	0.02	0.02	0.16	0.02	25.45	<.001	0.01	.950	1.17	.287
Physical fitness	0.09	0.02	0.01	0.02	7.33	.011	3.48	.072	11.51	.002
IQ scores	-0.01	0.02	0.00	0.02	0.32	.575	3.94	.052	1.31	.257
Vocabulary	0.13	0.03	0.10	0.02	0.38	.544	0.00	.990	1.22	.279
Reasoning	0.14	0.03	-0.01	0.03	12.24	.001	0.06	.802	14.04	.001
Processing speed	0.09	0.02	-0.01	0.03	8.49	.007	0.04	.846	30.85	<.001
STM	0.10	0.02	-0.05	0.03	13.18	.001	0.95	.338	0.01	.926

^aMean accuracy coefficients are correlations between each participant's ratings of older or younger faces above or below the median in attractiveness with the targets' actual scores, controlling face sex and actual face age. For Doctors' health score and IQ scores, $df = 1, 62$; for all other variables, $df = 1, 30$.