SENSITIVITY TO "BAD GENES" AND THE ANOMALOUS FACE OVERGENERALIZATION EFFECT: CUE VALIDITY, CUE UTILIZATION, AND ACCURACY IN JUDGING INTELLIGENCE AND HEALTH

Leslie A. Zebrowitz and Gillian Rhodes

ABSTRACT: The bad genes and anomalous face overgeneralization accounts of facial preferences were tested by examining cue validity, cue utilization, and accuracy in judging health and intelligence from faces in the upper and lower halves of the distributions of attractiveness and its components: averageness, symmetry, and masculinity. Consistent with the bad genes hypothesis, facial attractiveness, averageness, symmetry, and male face masculinity each provided valid cues to intelligence and/or health for faces in the lower but not the upper halves of the distributions of these facial qualities. Consistent with the anomalous face overgeneralization hypothesis, attractiveness and its components were utilized as cues not only for faces in the lower halves of the distributions, but also for those in the upper halves. Intelligence and health were judged accurately for faces in the lower but not the upper half of the attractiveness distribution, and attractiveness mediated this accuracy at all ages except adolescence. Since adolescence is the prime mating age, the latter finding raises questions about the utility of attractiveness as an evolved mechanism to ensure the selection of high quality mates.

KEY WORDS: attractiveness; bad genes; face overgeneralization; impression accuracy.

Research in interpersonal sensitivity and person perception has documented surprising accuracy in judging intelligence and health from facial appearance. For example, intelligence ratings based on facial photographs of a representative sample of individuals at ages 10, 13, and 30 were significantly correlated with IQ scores (Zebrowitz, Hall, Murphy, & Rhodes,

Leslie A. Zebrowitz is affiliated with Brandeis University. Gillian Rhodes is affiliated with University of Western Australia

This research was supported in part by NIH grant R01MH066836 to the first author. We thank the Institute of Human Development (IHD) at the University of California, Berkeley for access to the data archives used in this study. Requests for reprints should be sent to Professor Leslie Zebrowitz, Department of Psychology, Brandeis University, Waltham, MA 02454.

2002), and health ratings of facial photographs of 17-year-olds from the same sample were significantly correlated with their actual health scores during adolescence (Kalick, Zebrowitz, Langlois, & Johnson, 1998). Although the effect sizes for these accuracy indices are modest, even ''small effects are impressive'' when one considers the limited information available to perceivers viewing black and white facial photographs (Prentice, & Miller, 1992).

The finding that intelligence and health can be detected in faces raises two questions. First, ''what specific facial qualities convey these attributes?'' and, second, ''which of the valid facial cues to intelligence and health do perceivers use to achieve accurate judgments?'' As shown in Figure 1, these questions can be framed within the Brunswik lens model (Brunswik,1955, 1956) that has been utilized by other researchers to explore interpersonal sensitivity. The paths on the left of the model show *cue validity*—the relationships between facial qualities and *measured* traits. The paths on the right of the model show *cue utilization*—the relationships between facial qualities and *perceived* traits. The path at the bottom of the model shows *accuracy*—the relationship between measured traits and perceived traits. Previous research has demonstrated that facial attractiveness or its component qualities of averageness, symmetry, and sexual dimorphism provide valid cues to intelligence and/or health and

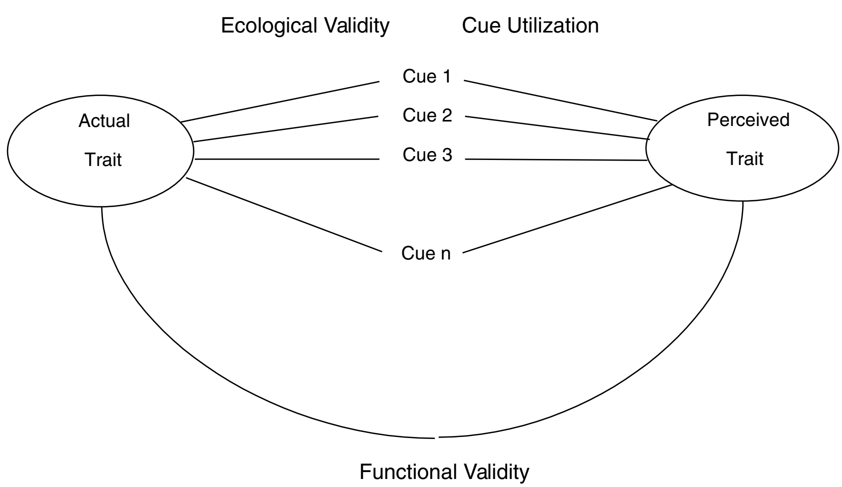

Figure 1. The Lens Model (adapted from Brunswik, 1955, p. 206). ''Ecological validity'' refers to cue validity and ''functional validity'' refers to accuracy.

that they also are related to perceived intelligence and/or health (Kalick et al., 1998; Rhodes, Chan, Zebrowitz, & Simmons, 2003; Rhodes, Zebrowitz, Clark, Kalick, Hightower, & McKay, 2001; Zebrowitz et al., 2002). The goal of the present research was to clarify the relationships between the cue validity of these facial qualities and their utilization.[1]

One explanation for the documented cue validity and utilization effects is provided by the ''good genes'' hypothesis that has been offered by evolutionary theorists to account for facial preferences. According to this hypothesis, attractive faces signal mate quality, and preferences for attractive individuals evolved because they enhance reproductive success (e.g., Hamilton & Zuk, 1982; Thornhill & Gangestad, 1993, 1999; see Berry, 2000 and Zebrowitz & Rhodes, 2002 for pertinent reviews). While previous research investigating whether attractive faces do in fact signal good genes has focused on the relationship between attractiveness and health (e.g., Kalick et al., 1998; Shackelford & Larsen, 1997, 1999), it also has been argued that the preference for certain facial qualities may have evolved because they signal high intelligence (Zebrowitz et al., 2002; cf. Miller, 2000; Miller & Todd, 1998). Such a preference could have enhanced reproductive success because more intelligent mates conferred survival benefits on their offspring through the heritability of intelligence or through their ability to provide better parental care and more resources.

The rationale for the good genes hypothesis is provided by evidence that components of attractiveness, symmetry and averageness, are both associated with the ability to maintain normal development despite environmental and/or genetic stress in non-human animals (Møller & Swaddle, 1997; Parsons, 1990; Thornhill & Møller, 1997), and by evidence that genetic and environmental stress can produce deviations from averageness and symmetry in the human face (Thornhill & Møller, 1997). Facial averageness also may be associated with the ability to resist parasites (Gangestad & Buss, 1993; Thornhill & Gangestad, 1993), because the genetic heterozygosity that is associated with enhanced parasite resistance also tends to produce average forms (Livshits & Kobyliansky, 1991). Sexual dimorphism has been hypothesized to be a marker of ''good genes'' in males because testosterone not only produces secondary sexual traits, such as a large jaw, but also inhibits immune responses with the result that only very fit individuals may be able to ''afford'' these traits (Folstad & Karter, 1992; Møller, Christe, & Lux, 1999).

Consistent with the argument that facial averageness, symmetry, and masculinity are markers of fitness, 10-year-old children with more symmetrical faces and 13-year-old children with more average faces had higher IQ scores than their less symmetrical or less average peers

(Zebrowitz et al., 2002). Also, adolescent women with more average faces and adolescent men with more masculine faces had better health than their less average or less masculine peers (Rhodes et al., 2001; Rhodes et al., 2003).[2] Greater attractiveness also has been associated with higher intelligence (Zebrowitz et al., 2002), but, surprisingly, not with better health (Kalick et al., 1998; Shackelford & Larsen, 1997, 1999).

Implicit in the research generated by the good genes hypothesis has been the assumption of a linear relationship between attractiveness or its components and genetic fitness, and there has been particular emphasis on the greater mate quality of those who are highly attractive (e.g., Buss, 1989; Thornhill & Gangestad, 1999). A refinement of the ''good genes'' hypothesis is the ''bad genes'' hypothesis, which holds that whereas unattractive faces can signal poor genetic fitness, faces that are above average in attractiveness are no more ''fit'' than those in the middle of the attractiveness distribution (Zebrowitz, Fellous, Mignault, & Andreoletti, 2003). On this account, those who avoided mates with extremely unattractive faces would have increased their reproductive success over those who did not. In the extreme case of genetic anomalies, such as Down's syndrome, it is obvious that unattractive faces signal low health and intelligence. However, even more subtle deviations from average attractiveness can signal low fitness. For example, ''minor facial anomalies'' and facial characteristics associated with ''fetal alcohol syndrome'' may be diagnostic of low intelligence and other maladaptive traits (Bell & Waldrop, 1982; Campbell, Geller, Small, Petti, & Ferris, 1978; Cummings, Flynn, & Preus, 1982; Guy, Majorski, Wallace, & Guy, 1983; Hoyme, 1994; Krouse & Kauffman, 1982; Paulhus & Martin, 1986; Streissguth, Herman, & Smith, 1978; Thornhill & Møller, 1997; Waldrop & Halverson, 1972).

Related to the bad genes hypothesis is the anomalous face overgeneralization hypothesis, which holds that the evolutionary importance of recognizing individuals with bad genes may have produced a strong preparedness to respond to their anomalous facial qualities that is overgeneralized to normal adults whose faces merely resemble those who are unfit (Zebrowitz, 1997; Zebrowitz et al., 2003). For example, people whose somewhat asymmetrical faces resemble the more extreme values found in facial anomalies may be perceived as having lower health and intelligence than those with highly symmetrical faces even though this perception is erroneous (Rhodes et al., 2001, Zebrowitz et al., 2002; Zebrowitz et al., 2003). According to ecological theory (McArthur & Baron, 1983; Zebrowitz, 1990; Zebrowitz & Montepare, in press), the errors shown in such overgeneralization effects occur because they are less maladaptive than those that might result from failures to respond to fitness information. However, maladaptive it may be to reject healthy and intelligent, but unattractive, individuals as mates, it would be even more maladaptive to select those with cranio-facial anomalies who do in fact lack health and intelligence (for related discussions, see Kurzban & Leary, 2001; Neuberg, Smith, & Asher, 2000; Park, Faulkner, & Schaller, 2003).

The bad genes and anomalous face overgeneralization hypotheses generate the following predictions regarding judgments of health and intelligence from facial cues:

1. According to the bad genes hypothesis, attractiveness and its components (facial averageness, symmetry, masculinity) will provide valid cues to intelligence or health when faces range from medium to low levels of these facial qualities but not when they range from medium to high levels.
2. According to the anomalous face overgeneralization hypothesis, attractiveness and its components will be utilized as cues when judging intelligence and health both for faces that range from medium to low levels of these qualities and for those that range from medium to high levels.
3. The conjunction of cue validity and cue utilization will result in the accurate judgment of intelligence and health from facial cues when faces range from medium to low levels of attractiveness but not when they range from medium to high levels.

These three hypotheses were tested by reanalyzing archival data that had previously yielded evidence for cue validity, cue utilization, and/or accuracy in judging intelligence or health from facial images that varied in attractiveness, symmetry, averageness, or masculinity (Rhodes et al., 2001, 2003; Zebrowitz et al., 2002). The new analyses replaced comparisons of all faces with comparisons of faces that were low to medium or medium to high in the relevant facial quality. New analyses are reported only for comparisons that had shown accuracy or cue validity in prior research with all faces combined.

Method

Participants

Participants, born between 1920 and 1929, were drawn from the Intergenerational Studies of Development and Aging (IGS), three longitudinal studies archived at U.C. Berkeley Institute of Human Development. These studies included representative samples of individuals born in 1928–1929 in Berkeley, California or enrolled in the 5th and 6th grades in 1932 in Oakland, California.

Facial Stimuli

Images of participants' faces were cropped from black and white whole-body photographs contained in the IGS archives. Facial stimuli were available in childhood (age 10), puberty (ages 12–13 for girls and 14–15 for boys), adolescence (age 17), and adulthood (age 30–40).[3]

Face Ratings

All face ratings, made on 7-point scales, were taken from previous studies. Attractiveness ratings came from Zebrowitz, Olson, and Hoffman (1993). The average alpha coefficient across age for these ratings was .87 for male faces and .90 for female faces. In addition to high reliability, previous research has provided some evidence for the convergent and predictive validity of the attractiveness ratings. Attractiveness ratings of adolescent girls in the present study were significantly correlated with prettiness ratings of the same female faces made in the 1960s, indicating that they may be generalized to another historical time and to another sample of judges (Zebrowitz et al., 1993). Although no cross-cultural data are available for the particular faces used in this study, other research has documented consensual attractiveness judgments across raters from diverse cultures (cf. Dion, 2002; Langlois, et al., 2000; Zebrowitz, Montepare, & Lee, 1993). Finally, pertinent to the evolutionary framework, the adolescent attractiveness ratings used in this study predicted marital outcomes, with less attractive participants marrying later and more likely to never marry (Kalick et al., 1998).

Facial averageness and symmetry ratings came from Rhodes et al. (2001), $\alpha = .84$ for male faces and .86 for female faces, averaged across ratings and age groups. In the latter study ''distinctiveness'' was rated rather than ''averageness'' because the latter can be misinterpreted to mean ''not particularly good looking'' instead of spatially average as intended. Judges rated the ease with which the face could be picked out of a crowd of faces of that age group. Rated distinctiveness and symmetry change systematically with experimental manipulations of averageness and symmetry, respectively, indicating that they are valid measures of these physical traits (Rhodes, Proffitt, Grady, & Sumich, 1998; Rhodes &

Tremewan, 1996). Masculinity ratings came from Rhodes et al. (2003), $\alpha = .92$. Perceived health ratings came from Kalick et al. (1998), $\alpha = .78$ for male faces and .83 for female faces, averaged across age group, and perceived intelligence ratings came from Zebrowitz et al. (2002), $\alpha = .84$ for male faces and .83 for female faces, averaged across age group.

IQ Measures

IQ scores used by Zebrowitz et al. (2002) to study participants in childhood, puberty, adolescence, and adulthood were taken from the IGS archive. Stanford-Binet scores were available for participants in childhood, puberty, and adolescence. Wechsler Adult Intelligence full scale scores (WAIS-R) were available in adulthood. Although the limitations of these measures of intelligence must be acknowledged, we were constrained by the data available in the archive, which did not incorporate more modern, multidimensional assessments. Evidence for the reliability of our measures is provided by stability across time, with the average correlation across all possible pairs of age levels between childhood and adulthood equaling .75. Some evidence for the predictive validity of our measures is provided by significant correlations between adolescent IQ scores and the level of education achieved in adulthood, (Zebrowitz, Andreoletti, Collins, Lee, & Blumenthal, 1998).

Health Measures

Health scores used by Kalick et al. (1998) to study participants in adolescence (averaged across ages 11–18) and adulthood (30–36) were taken from the IGS archive. Health scores for participants at age 17, used by Rhodes et al. (2001), were also taken from the archive. Health in adolescence was assessed from age 11 through 18 by means of clinical exams and detailed histories. 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." Previous research had shown that ratings were largely based on the frequency, duration, and severity of infectious conditions (Bayer & Snyder, 1950). Ratings were averaged from ages 11 through 18 to create the Adolescent Health Measure. The alpha reliability for this composite was .68.

Health scores for people in adulthood were compiled on a similar 5-point scale, but they were based on a single medical exam and history administered between the ages of 30 and 36, supplemented for about a third of the participants with an overnight hospital stay for lab work. Examining physicians prepared charts, which were converted to health ratings independently by two leading IHD medical researchers. The correlation between their ratings was .63. After discussing their independent ratings, the researchers jointly produced a ''conference score'' for each participant that correlated at greater than .80 with each researcher's initial assessments. These conference scores were used for the health ratings in adulthood. As age increased, chronic illnesses tended to have an increasing influence on the health ratings (Bayer, Whissell-Buechy, & Honzik, 1981). Like the IQ measures, the health measures have limitations, but again, we were constrained by the data available in the archive. Evidence for their reliability is provided not only by inter-judge agreement, but also by modest stability from adolescence to adulthood, $r(297) = .18$, $p = .002$.

Results

Overview

The validity and utilization of various facial qualities for judging intelligence and health as well as the accuracy of judging these adaptive attributes were assessed separately for faces ranging from medium to low levels and from medium to high levels of the facial qualities. To this end, faces were divided into two groups—those at or below the median in the facial quality (designated ''below the median'') and those at or above the median (designated ''above the median'') in the facial quality. Table 1 reports the means and standard deviations of the relevant facial quality for faces in the two groups. Cue validity was assessed by correlations between facial qualities and measured intelligence or health. Cue utilization was assessed by correlations between facial qualities and perceived intelligence or health. Accuracy was assessed by correlations between measured and perceived intelligence or health. Finally, partial correlations were performed to determine whether the utilized facial cues mediated accurate judgments. Age was partialed out in correlations involving adult faces, which spanned a number of years that could introduce variations in appearance, health, or IQ scores. Figures 2 and 3 use Brunswik's (1955, 1956) lens model to summarize the results for cue validity, cue utilization, and accuracy. Correlations shown below the lines in these figures represent cue validity, cue utilization, and accuracy for faces below the median in a particular facial quality; values above the lines represent faces above the median.

TABLE 1

Descriptive Statistics for Faces Above and Below the Median in Appearance Qualities

	Face group			
	Above median		Below median	
Variable	M (N)	SD	M (N)	SD
Attractiveness				
Childhood	4.02 (142)	.42	2.79 (156)	.52
Puberty	3.86 (177)	.43	2.68 (179)	.46
Adolescence	3.78 (163)	.44	2.58 (171)	.46
Middle adulthood	3.57 (100)	.59	2.23 (100)	.40
Later adulthood	3.76 (108)	.47	2.56 (113)	.44
Symmetry—Childhood	5.16 (132)	.31	4.06 (132)	.50
Averageness				
Puberty	4.95 (176)	.26	4.06 (180)	.47
Adolescent females	4.92 (79)	.22	4.02 (86)	.53
Masculinity—Adolescent males	4.87 (77)	.40	3.84 (77)	.38

Note: All variables were rated on 7-point scales.

Intelligence Judgments

Validity and utilization of facial cues. As predicted, the correlation between attractiveness and actual IQ scores was significant for participants below the median in attractiveness in childhood, r (103) = .20, p = .04, puberty, r (81) = .30, p = .01, and adulthood, r (62) = .35, p = .01 but not for those above the median in attractiveness in childhood r (101) = −.04, p = .66, puberty, r (98) = .04, p = .69, or adulthood, r (68) = .13, p = .28. The results for participants below the median replicate the findings reported by Zebrowitz et al. (2002) for all participants combined. In adolescence, attractiveness was not significantly correlated with IQ scores for participants below the median, r (113) = .04, p = .71, or above the median, r (119) = .07, p = .44.

Whereas attractiveness was a valid cue to intelligence only for participants below the median in attractiveness, it was utilized as a cue for

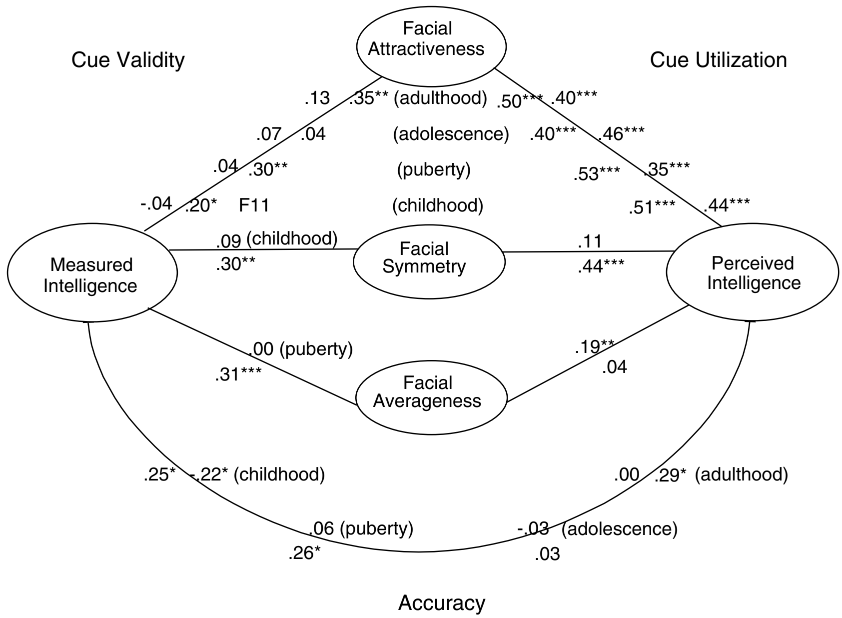

Figure 2. Cue validity, cue utilization, and accuracy in judging intelligence from faces below the median in appearance qualities (correlations shown below the lines) and above the median in appearance qualities (correlations shown above the lines). $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

participants both above and below the median, consistent with the overgeneralization hypothesis: in childhood, $r\,(133) = .51$ for those below the median and $r\,(135) = .44$, for those above the median; in puberty, $r\,(172) = .53$ for those below the median and $r\,(172) = .35$, for those above the median; in adolescence, $r\,(159) = .40$ for those below the median and $r\,(129) = .46$ for those above the median, in adulthood, $r\,(67) = .51$, for those below the median and $r\,(70) = .41$, for those above the median, all ps $< .001$.

Previous research had found that facial symmetry in childhood and facial averageness in puberty each predicted IQ scores (Zebrowitz et al., 2002). These relationships were assessed for faces that were above or below the median in these qualities. The relationship of *perceived* intelligence to facial symmetry in childhood and facial averageness in puberty also was assessed for faces above and below the median in these qualities to test the hypothesis that they would be taken as indicators of intelligence in both cases, as predicted by the overgeneralization hypothesis.

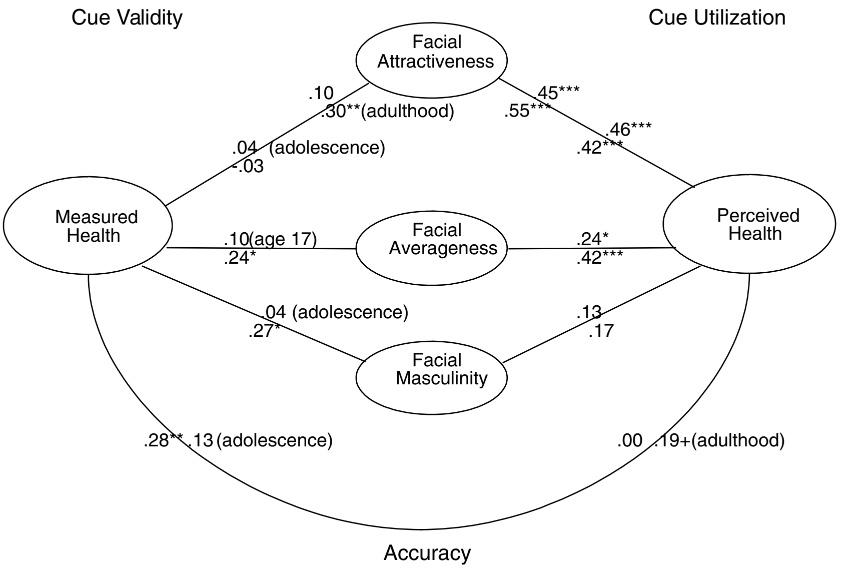

Figure 3. Cue validity, cue utilization, and accuracy in judging health from faces below the median in appearance qualities (correlations shown below the lines) and above the median in appearance qualities (correlations shown above the lines). $+ p < .10$, $*p < .05$, $**p < .01$, $***p < .001$.

As predicted, the correlation between facial averageness and actual IQ scores in puberty was significant for participants below the median in averageness, $r\,(89) = .31$, $p = .003$, replicating the findings of Zebrowitz et al. (2002) for all participants, whereas this correlation was not significant for those above the median in averageness, $r\,(87) = -.00$, $p = .99$[4]. As predicted, the correlation between facial symmetry and actual IQ scores was significant for participants below the median in symmetry in childhood, $r\,(98) = .30$, $p = .003$, replicating the findings of Zebrowitz et al. (2002) for all participants combined, while the correlation was not significant for those above the median in symmetry, $r\,(88) = .09$, $p = .42$.

Just as facial symmetry was a valid cue to children's IQ scores only for those below the median in symmetry, so was it utilized in judging intelligence only for children below the median, $r\,(130) = .44$, $p < .001$, but not for those above the median, $r\,(130) = .11$, $p = .19$, which is contrary to the overgeneralization hypothesis. Whereas averageness was a valid cue to intelligence only for participants below the median, it was utilized as a cue when perceiving the intelligence of participants above the median, r (174) = .19, p = .01, but not participants below the median, r (177) = .04, p = .58.[5]

Accuracy. As predicted, the correlation between perceived intelligence and actual IQ scores was significant for participants below the median in attractiveness in childhood, r (99) = .25, p < .02, puberty, r (80) = .26, p = .02, and adulthood, r (57) = .29, p = .02, but not in adolescence, r (110) = .03, p = .78. These results replicate the findings reported by Zebrowitz et al. (2002) for all participants combined. Also as predicted, the correlation between perceived intelligence and actual IQ scores showed no evidence of accuracy for participants above the median in attractiveness. Rather, there was a negative correlation between perceived intelligence and IQ scores in childhood, r (94) = −.22, p = .03, and no significant correlation at puberty, r (94) = .06, p = .52, adolescence, r (115) = −.03, p = .76, or adulthood, r (60) = −.00, p = .97.

Utilization of the valid attractiveness cues mediated accuracy in judging the intelligence of faces below the median in attractiveness in childhood, puberty, and adulthood. The significant relationships between IQ scores and perceived intelligence were lost when attractiveness was statistically controlled: in childhood, partial r (96) = .12, p = .22; in puberty, partial r (77) = .13, p = .25; in adulthood, partial r (51) = −.06, p = .67. Although facial symmetry in childhood met the criteria necessary for mediation (Baron & Kenny, 1986), the relationship between IQ scores and perceived intelligence remained significant when symmetry was controlled, r (96) = .19, p = .05.

Health Judgments

Validity and utilization of facial cues. Contrary to the bad genes hypothesis, attractiveness did not provide a valid cue to health in adolescence when correlations were examined for adolescent faces below the median, r (169) = −.03, p = .71, as well as for those above the median, r (166) = .04, p = .60, replicating the effects reported by Kalick et al. (1998) for all faces combined. However, consistent with the bad genes hypothesis, attractiveness was a valid cue to health for adult faces below the median, r (79) = .30, p = .01, but not for those above the median, r (83) = .12, p = .29.

Although attractiveness was not a valid cue to adolescents' health, it was utilized as a cue both for those below the median in attractiveness, r (127) = .42, p < .001, and for those above the median, r (129) = .46, p < .001. Whereas attractiveness was a valid cue to adult's health only for participants below the median in attractiveness, it was utilized as a cue both for participants above the median, r (81) = .45, $p < .001$, and those below the median, r (77) = .54, $p < .001$, consistent with the overgeneralization hypothesis.

Previous research had found that women's facial averageness at age 17 and adolescent men's facial masculinity each predicted health scores (Rhodes et al., 2001, 2003). These relationships were assessed for faces that were above or below the median in these qualities. The relationship of *perceived* health to women's facial averageness and men's facial masculinity also was assessed for faces above and below the median in these qualities to test the hypothesis that they would be taken as indicators of health in both cases, as predicted by the overgeneralization hypothesis.[6]

As predicted by the bad genes hypothesis, women's facial averageness and health scores at age 17 were significantly correlated for faces below the median in averageness, r (80) = .24, $p = .03$, replicating the findings of Rhodes et al. (2001) for all female participants combined, but not for faces above the median, r (71) = .10, $p = .42$. Consistent with the overgeneralization hypothesis, facial averageness also was correlated with perceived health for female faces below the median, r (59) = .42, $p < .001$, as well as for those above the median, r (64) = .24, $p = .05$. As predicted by the bad genes hypothesis, men's facial masculinity was correlated with adolescent health scores for male participants below the median in facial masculinity, r (75) = .27, $p = .02$, replicating the findings of Rhodes et al. (2003) for all male participants combined, while it was uncorrelated for those above the median, r (75) = .04, $p = .72$. However, facial masculinity was not significantly utilized when judging the health of participants below the median, r (62) = .17, $p = .17$, or above the median, r (58) = .13, $p = .32$.[7]

Accuracy. As predicted, the correlation between perceived health and adolescent health scores was significant for participants below the median in attractiveness in adolescence, r (127) = .28, $p = .002$, replicating the findings reported by Kalick et al. (1998) for all participants combined. Judgments of health were also marginally accurate for participants below the median in attractiveness in adulthood, r (77) = .19, $p = .10$. Also as predicted, the correlation between perceived and actual health scores showed no significant evidence of accuracy for participants above the median in attractiveness: adolescence, r (129) = .13, $p = .14$; adulthood, r (81) = .00, $p = .99$.

Although perceivers did show accuracy in judging adolescent health when faces were in the bottom half of the attractiveness distribution, this ability had to depend on utilization of cues other than attractiveness, since the latter was unrelated to health. Two possibilities are facial averageness and facial masculinity, which, as described above, provided valid cues to the health of female and male adolescents, respectively. Partial correlations revealed that the marginally significant correlation between women's health scores and perceived health, r (57) = .23, p = .08, lost significance when facial averageness was controlled, r (56) = .14, p = .30. The partial correlation controlling masculinity could not be tested for adolescent male faces because the zero order correlation between their health scores and perceived health was not significant, r (62) = .10, p = .42. In adulthood, facial attractiveness did mediate the relationship between perceived and actual health, which lost significance when attractiveness was controlled, r (76) = .02, p = .89.

Discussion

Consistent with the bad genes hypothesis, the present findings provide a theoretically significant qualification to previous evidence that attractiveness, averageness, symmetry, or facial masculinity provide valid cues to intelligence or health (Rhodes et al., 2001, 2003, Zebrowitz et al., 2002). These facial qualities provided valid cues only for faces in the lower halves of the distributions, not for those in the upper halves. Replicating the findings of Zebrowitz et al. (2002) for all faces combined, attractiveness was significantly correlated with IQ scores in childhoood, puberty, and adulthood for faces in the lower half of the attractiveness distribution, whereas these correlations were not significant for faces in the upper half of the distribution. Similarly, averageness at puberty and symmetry in childhood were significantly correlated with IQ scores for faces in the lower half of the distributions of these facial qualities, but not for faces in the upper half of the distributions. In addition, adolescent females' facial averageness and adolescent males' facial masculinity were each significantly correlated with health scores for faces in the lower half of the distributions, replicating the findings of Rhodes et al. (2001, 2003) for all faces combined, but not for those in the upper half. Finally, adult attractiveness was significantly correlated with health scores for faces in the lower but not the upper half of the attractiveness distribution. Compared with the eight significant correlations showing cue validity for faces in the lower halves of the distributions of attractiveness, averageness, symmetry, or sexual dimorphism, there were no cases in which a significant relationship was found for faces in the upper halves of the distributions. Thus, low attractiveness, low averageness, low symmetry, or low masculinity signal low fitness, as indexed by intelligence or health. On the other hand, high attractiveness, averageness, symmetry, or masculinity do not signal any higher levels of intelligence or health than do moderate levels of these attributes.

Whereas all significant relationships between facial appearance and measures of actual fitness were confined to faces in the lower half of the distributions, supporting the bad genes hypothesis, relationships between facial appearance and *perceived* intelligence or health provided support for the overgeneralization hypothesis. Facial qualities were as likely to be significantly correlated with perceived health and intelligence for faces in the upper half of the appearance distributions as for those in the lower half. Most notably, facial attractiveness was strongly related to perceived intelligence for faces in both the upper and lower halves of the attractiveness distribution despite being related to IQ scores only for faces in the lower half. An exception to the overgeneralization effect occurred for the perception of intelligence from facial symmetry in childhood, which was limited to the subset of faces in the lower half of the distribution, where this quality provided a valid cue.

Accurate perceptions of intelligence and health from facial appearance (Kalick et al., 1998, Zebrowitz et al., 2002) were mediated by the utilization of valid cues. Judgments of intelligence were accurate in childhood, puberty, and adulthood only when the faces were medium to low in attractiveness, and attractiveness mediated this accuracy. Similarly, judgments of health in adolescence and adulthood were accurate only when faces were medium to low in attractiveness. Attractiveness mediated this accuracy in adulthood, but not in adolescence, where there was some evidence for mediation by averageness in the case of female faces.

Although the overall pattern of results provides support for the hypothesis that perceivers are sensitive to facial cues that are valid indicators of ''bad genes'' rather than ''good genes'', and that this sensitivity is overgeneralized in perceivers' cue utilization, the specific results for adolescent faces are problematic for the argument that facial preferences evolved because they enhance reproductive success. Although perceivers did show accuracy in judging health from facial cues at adolescence when faces were in the bottom half of the attractiveness distribution, attractiveness was not the mediator, and it is presumably attractiveness judgments that influence mate choice. It should be recalled that perceivers' ability to accurately judge the health of the entire distribution of adolescent faces also was not mediated by attractiveness (Kalick et al., 1998). If attractiveness is not the mediator, then perceivers' preferences for attractive faces could not have enhanced reproductive success via eschewal of the least healthy mates. Similarly, perceivers' preferences for attractive faces would not have enhanced reproductive success via avoidance of the least intelligent mates, since attractiveness was not a valid cue to adolescents' intelligence for faces in either the bottom or the top half of the distribution.

Although our results are consistent with the bad genes hypothesis, it should be noted that the observed relationships between low attractiveness and low fitness could reflect additional mechanisms. As previously argued by Zebrowitz and her colleagues (Zebrowitz, 1997; Zebrowitz & Collins, 1997; Zebrowitz, Collins, & Dutta, 1998; Zebrowitz et al., 2002), there are multiple paths to a covariation between attractiveness and adaptive traits. These include not only the influence of biological factors on both attractiveness and adaptive traits that is consistent with the bad genes hypothesis, but also: (a) an influence of environmental factors, such as nutrition; (b) an influence of adaptive traits on attractiveness, such as the use of more flattering grooming by more intelligent or healthier individuals; and (c) an influence of attractiveness on people's social environments, which in turn may influence their intelligence and health.

Regardless of the mechanisms that contribute to relationships between low attractiveness and low fitness, the present findings reveal that people accurately judge the traits of intelligence and health in faces from the lower half of the attractiveness distribution, but not in those from the upper half, that there are valid facial cues to low but not high intelligence or health, and that people not only correctly utilize these cues when they are valid, but they also overgeneralize, utilizing these cues in the upper half of the distribution, where they are not valid.

Notes

1. The data provided to assess ''utilization'' are typically correlations between cues and perceptions, and this term is therefore not meant to imply that a causal relationship has been determined.
2. There was also a lagged relationship between health and facial averageness: adolescent men with more average faces had better health during childhood.
3. Facial stimuli were also available in later adulthood (ages 50–62). However, analyses at this age are omitted because previous research had found no significant evidence for cue validity or accuracy in judging health or intelligence at this age, and the present study also, found no evidence for effects when analyses were performed separately for faces above and below the median.
4. These correlations were based on ratings of facial distinctiveness. The signs have been reversed so that they reflect the effect of facial averageness (i.e., non-distinctiveness).
5. Consistent with Zebrowitz et al's (2002) results for all participants, facial averageness was not significantly correlated with IQ scores in childhood or the 30s for participants below

or above the median in averageness, all $ps > .18$. Nevertheless, it was utilized as a cue when perceiving intelligence of children both below the median in averageness, r (126) = .35, and above the median, r (138) = .21, $p = .01$, as well as when perceiving the intelligence of participants in their 30s both below the median, r (87) = .25, $p = .02$, and above the median, r (82) = .24, $p = .03$.

6. The good genes hypothesis makes no clear prediction about the relationship between facial femininity (in female faces) and health, and Rhodes et al. (2003) found none.
7. Health at age 17 was not significantly related to averageness for male faces above or below the median, consistent with the results of Rhodes et al. (2001) for all male faces combined; adolescent health also was not significantly related to femininity for female faces above or below the median, consistent with the results of Rhodes et al. (2003) for all female faces combined.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality & Social Psychology*, *51*, 1173–1182.

Bayer, L. M., & Snyder, M. M. (1950). Illness experience of a group of normal children. *Child Development*, *21*, 93–120.

Bayer, L. M., Whissell-Buechy, D., & Honzik, M. P. (1981). Health in the middle years. In D. H. Eichorn, J. A. Clausen, N. Haan, M. P. Honzik, & P. H. Mussen (Eds.), *Present and past in middle life* (pp. 55–88). New York: Academic Press.

Bell, R. Q., & Waldrop, M. F. (1982). Temperament and minor physical anomalies. In R. Porter & G. M. Collins (Eds.), *Temperamental differences in infants and young children: CIBA Symposium No. 89*. (pp. 206–220). London: Pitman.

Berry, D. S. (2000). Attractiveness, attraction, and sexual selection: Evolutionary perspectives on the form and function of physical attractiveness. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 32, pp. 273–342). San Diego: Academic Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioural and Brain Sciences*, *12*, 1–49.

Campbell, M., Geller, B., Small, A. M., Petti, T. A., & Ferris, S. H. (1978). Minor physical anomalies in young psychotic children. *American Journal of Psychiatry*, *135*, 573–575.

Cummings, C. D., Flynn, D., & Preus, M. (1982). Increased morphological variants in children with learning disabilities. *Journal of Autism and Developmental Disorders*, *12*, 373–383.

Dion, K. K. (2002). Cultural perspectives on facial attractiveness. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 239–260). Westport, CT: Ablex.

Folstad, I., & Karter, A. J. (1992). Parasites, bright males, and the immunocompetence handicap. *American Naturalist*, *139*, 603–622.

Gangestad, S. W., & Buss, D. M. (1993). Pathogen prevalence and human mate preferences. *Ethology and Sociobiology*, *14*, 89–96.

Guy, J. D., Majorski, L. V., Wallace, C. J., & Guy, M. P. (1983). The incidence of minor physical anomalies in adult male schizophrenics. *Schizophrenia Bulletin*, *9*, 571–582.

Hamilton, W. D., & Zuk, M. (1982). Heritable true fitness and bright birds: A role for parasites? *Science*, *218*, 384–387.

Hoyme, H. E. (1994). Minor anomalies: Diagnostic clues to aberrant human morphogenesis. In T. A. Markow (Ed.), *Developmental instability: Its origins and evolutionary implications* (pp. 309–317). The Netherlands: Kluwer Academic Publishers.

Kalick, S. M., Zebrowitz, L. A., Langlois, J. H., & Johnson, R. M. (1998). Does human facial attractiveness honestly advertise health? Longitudinal data on an evolutionary question. *Psychological Science*, *9*, 8–13.

Krouse, J. P., & Kauffman, J. M. (1982). Minor physical anomalies in exceptional children: A review and critique of research. *Journal of Abnormal Child Psychology*, *10*, 247–264.

Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin*, *127*, 187–208.

Langlois, J. H., Kalakanis, L. E., Rubenstein, A. J., Larson, A. D., Hallam, M. J., & Smoot, M. T. (2000). Maxims and myths of beauty: A meta-analytic and theoretical review. *Psychological Bulletin*, *126*, 390–423.

Livshits, G., & Kobyliansky, E. (1991). Fluctuating asymmetry as a possible measure of developmental homeostasis in humans: A review. *Human Biology*, *63*, 441–466.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*, 215–238.

Miller, G. F., & Todd, P. M. (1998). Mate choice turns cognitive. *Trends in Cognitive Sciences*, *2*, 190–198.

Miller, G. (2000). *The mating mind*. New York: Doubleday.

Møller, A. P., & Swaddle, J. P. (1997). *Asymmetry, developmental stability, and evolution*. Oxford, England: Oxford University Press.

Møller, A. P., Christe, P., & Lux, E. (1999) Parasitism, host immune function, and sexual selection. *Quarterly Review of Biology*, *74*, 3–74.

Neuberg, S. L., Smith, D. M., & Asher, T. (2000). Why people stigmatize: Toward a biocultural framework. In T. Heathereton, R. Kleck, J. G. Hull, & M. Hebl (Eds.), *The social psychology of stigma* (pp. 31–61). New York: Guilford.

Park, J. H., Faulkner, J., & Schaller, M. (2003). Evolved disease-avoidance processes and contemporary anti-social behavior: Prejudicial attitudes and avoidance of people with physical disabilities. *Journal of Nonverbal Communication*, *27*, 65–88.

Parsons, P. A. (1990). Fluctuating asymmetry: An epigenetic measure of stress. *Biological Reviews*, *65*, 131–145.

Paulhus, D. L., & Martin, C. L. (1986). Predicting adult temperament from minor physical anomalies. *Journal of Personality and Social Psychology*, *50*, 1235–1239.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, *112*, 160–164.

Rhodes, G., Chan, J., Zebrowitz, L. A., & Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health? *Proceedings of the Royal Society of London, Series B (Suppl.)*, 270, S93–95.

Rhodes, G., Proffitt, F., Grady, J., & Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychonomic Bulletin & Review*, *5*, 659–669.

Rhodes, G., & Tremewan, T. (1996). Averageness, exaggeration, and facial attractiveness. *Psychological Science*, *7*, 105–110.

Rhodes, G., Zebrowitz, L. A., Clark, A., Kalick, S. M., Hightower, A., & McKay, R. (2001). Do facial averageness and symmetry signal health? *Evolution & Human Behavior*, *22*, 31–46.

Shackelford, T. K., & Larsen, R. J. (1997). Facial asymmetry as an indicator of psychological, emotional, and physiological distress. *Journal of Personality and Social Psychology*, *72*, 456–466.

Shackelford, T. K., & Larsen, R. J. (1999). Facial attractiveness and physical health. *Evolution & Human Behavior*, *20*, 71–76.

Streissguth, A. P., Herman, C. S., & Smith, D. W. (1978). Intelligence, behavior, and dysmorphogenesis in the fetal alcohol syndrome: A report on 20 patients. *The Journal of Pediatrics*, *92*, 363–367.

Thornhill, R., & Gangestad, S. W. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human Nature*, *4*, 237–269.

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences*, *3*, 452–460.

Thornhill, T., & Møller, A. P. (1997). Developmental stability, disease and medicine. *Biological Reviews*, *72*, 497–548.

Waldrop, M. F., & Halverson, C. F. (1972). Minor physical anomalies: Their incidence and relation to behavior in a normal and a deviant sample. In R. C. Smart & M. S. Smart (Eds.), *Readings in child development and relationships* (pp. 146–155). New York: Macmillan.

Zebrowitz, L. A. (1990). *Social perception*. Buckingham, England: Open University Press.

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul*? Boulder, CO: Westview Press.

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, *1*, 204–223.

Zebrowitz, L. A., Andreoletti, C., Collins, M. A., Lee, S. Y., & Blumenthal, J. (1998). Bright, bad, babyfaced boys: Appearance stereotypes do not always yield self-fulfilling prophecy effects. *Journal of Personality and Social Psychology*, *75*, 1300–1320.

Zebrowitz, L. A., Collins, M. A., & Dutta, R. (1998). Appearance and personality across the lifespan. *Personality and Social Psychology Bulletin*, *24*, 736–749.

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin*, *28*, 238–249.

Zebrowitz, L. A., Fellous, J. M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review*, *7*, 194–215.

Zebrowitz, L. A., & Montepare, J. M. (in press). The ecological approach to person perception: Divergence and convergence with evolutionary psychology. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology*. New York: Psychology Press.

Zebrowitz, L. A., Montepare, J. M., & Lee, H. K. (1993). They don't all look alike: Individuated impressions of other racial groups. *Journal of Personality and Social Psychology*, *65*, 85–101.

Zebrowitz, L. A., Olson, K., & Hoffman, K. (1993). Stability of babyfaceness and attractiveness across the lifespan. *Journal of Personality and Social Psychology*, *64*, 453–466.

Zebrowitz, L. A., & Rhodes, G. (2002). Nature let a hundred flowers bloom: The multiple ways and wherefores of attractiveness. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 261–293). Westport, CT: Ablex.