Let Me Guess How Old You Are:
Effects of Age, Gender, and Facial Expression on Perceptions of Age

Manuel C. Voelkle
Max Planck Institute for Human Development

Natalie C. Ebner
Yale University

Ulman Lindenberger and Michaela Riediger
Max Planck Institute for Human Development

Perceptions of age influence how we evaluate, approach, and interact with other people. Based on a paramorphic human judgment model, the present study investigates possible determinants of accuracy and bias in age estimation across the adult life span. For this purpose, 154 young, middle-aged, and older participants of both genders estimated the age of 171 faces of young, middle-aged, and older men and women, portrayed on a total of 2,052 photographs. Each face displayed either an angry, fearful, disgusted, happy, sad, or neutral expression (FACES database; Ebner, Riediger, & Lindenberger, 2010). We found that age estimation ability decreased with age. Older and young adults, however, were more accurate and less biased in estimating the age of members of their own as compared with those of the other age group. In contrast, no reliable own-gender advantage was observed. Generally, the age of older faces was more difficult to estimate than the age of younger faces. Furthermore, facial expressions had a substantial impact on accuracy and bias of age estimation. Relative to other facial expressions, the age of neutral faces was estimated most accurately, while the age of faces displaying happy expressions was most likely underestimated. Results are discussed in terms of methodological and practical implications for research on age estimation.

Keywords: age estimation, own-age advantage/bias, own-gender advantage/bias, faces, emotion

As Samuel Klugman noted more than 60 years ago, a "fairly long list could be made of situations in which an estimate of a person's age is necessary" (1947, p. 29). This applies to our daily lives, as well as to various professional disciplines ranging from anthropology over psychology to forensic medicine, law, and law enforcement. In fact, age is one of the most important attributes to describe an unfamiliar person. This is true in everyday situations ("I met someone and I guess he is about my age") as well as under exceptional circumstances, such as eye-witness testimonies ("The suspect is described as 45 years of age"). Oftentimes multiple cues for estimating someone's age are readily at hand, such as his or her behavior, appearance, or background information on someone's vita. However, in some situations a picture of a person's face is the only information available. Examples include photographs taken by surveillance cameras or pictures of missing persons. In particular with the rise of private or business-related social networks like Facebook, flickr, LinkedIn, and many others, it has become common practice to share pictures, often without additional back-ground information. This also applies to many companies and universities, where faculty pictures are routinely displayed on the institution's webpage, usually without information on age. In all of these examples, the perceiver is left estimating the age of the person in question. Apart from few situations (e.g., when determining the legal age of a person for buying alcohol or the legitimacy of receiving a student/junior or senior discount), misestimating someone's age may appear fairly inconsequential at first glance. However, as perceived age constitutes an important determinant of how we approach and interact with other people, this impression may be highly misleading. In almost all societies, a person perceived to be in his or her eighties will be addressed differently than a person perceived to be in his or her fifties or twenties, and it stands to reason that correct age estimation is a prerequisite for adequate social interactions and that—ceteris paribus—people with better age estimation ability should also exhibit better interpersonal skills.

Despite the prominent role of age estimation in everyday life, to date there exits little research and thus little theoretical and empirical knowledge in this area. Furthermore, many prior studies suffered from various limitations as discussed next. Making use of a comprehensive database of faces, the present study set out to change this situation by exploring age estimation across the entire adult life span, considering the influence of characteristics of the perceiver (who is estimating the age of another person), the target person (whose age is estimated), and interactions between the two.

The few existing studies on age estimation fall roughly into two broad categories. Either they focus on the role of *individual* char-

acteristics of the face (e.g., skin texture, wrinkles, or the relative importance of different parts of the face; cf. George & Hole, 1998, 2000; Burt & Perrett, 1995), or they focus on the role of *group* characteristics (e.g., race, gender, or age; cf. Rhodes, 2009) for age estimation. Regarding group characteristics, the precision with which the age of a face is estimated seems to depend on the age of the depicted person. In particular, overestimating the age of young faces (i.e., estimating the face as older than the person's actual age) appears to be a common phenomenon (Henss, 1991; Sörqvist & Eriksson, 2007; Vestlund, Langeborg, Sörqvist, & Eriksson, 2009; Willner & Rowe, 2001). There is also some evidence that the age of older persons is more likely to be underestimated than that of younger persons (e.g., Henss, 1991; Vestlund et al., 2009). Over and above these main effects, individuals seem to provide better age estimates of persons of their own age group as compared with other age groups (George & Hole, 1995; Klugman, 1947). A similar own-group advantage in age estimation was demonstrated for race by Dehon and Brédart (2001). With respect to the effect of perceiver's gender on age estimation, some studies report higher accuracy and lower bias for female than male *perceivers* (Vestlund et al., 2009; in particular when rating older faces; Nkengne et al., 2008), and more precise age estimates for male than female *target persons* (Dehon & Brédart, 2001), but no own-gender advantage. Given this somewhat mixed evidence with respect to the effect of group membership, Rhodes concluded in his recent literature review: "given the dearth of data, the full range and impact of group biases on age estimation has yet to be explored" (2009, p. 8). In addition, none of the previous studies has explored the influence of emotional expressions on age estimation of faces, as another salient feature of a face.

Furthermore, prior research usually assigned different priorities to perceiver and face characteristics. Studies were either conducted with few participants, but many target stimuli, or with many participants and few target stimuli—sometimes only a single picture (e.g., Klugman, 1947). However, unless both factors are considered simultaneously, the possibility of detecting relevant interactions between these factors is limited. To address these limitations, the present study combined a comparatively large sample of 154 young and older male and female raters with an extensive set of 2,052 to-be-rated faces of 171 different young and older male and female posers, each displaying angry, fearful, disgusted, happy, sad, and neutral facial expressions. These unique design characteristics allowed us to simultaneously examine effects of age and gender of perceiver and poser, as well as interactions with facial expressions. The study design also provides sufficient power to detect any of our hypothesized effects assuming medium effect sizes in the population (Cohen, 1988, 1992).

We differentiate between *estimation bias* and e*stimation accuracy* (cf. Dehon & Brédart, 2001). *Estimation bias* reflects the direction of mis-estimation and is computed as the difference between estimated and true age. *Estimation accuracy* is the absolute value of this difference and informs about the overall extent of mis-estimation. Because over- and underestimation may cancel each other out when averaged across different estimates, distinguishing between bias and accuracy is crucial. It is possible that someone who shows no bias (i.e., no tendency to systematically over- or underestimate the age of a given group of faces) is nevertheless highly (nonsystematically) inaccurate in his or her age estimates. At the same time, however, someone with compar-

atively accurate age estimates may still exhibit substantial systematic bias in his or her responses.

A Paramorphic Model of Human Age Estimation

Given the dissatisfying lack of research on age estimation and the virtual nonexistence of a theoretical framework, we propose to conceive of age estimation as a judgment process (sensu Meehl, 1954) that can be represented in terms of a linear paramorphic model (Dawes, 1979; Dawes & Corrigan, 1974; Hoffman, 1960). Linear paramorphic models describe a judgment, such as an age estimate, as a weighted linear combination of different attributes (e.g., skin texture/wrinkles or hair color). The term *paramorphic* was first introduced to the psychological literature by Hoffman (1960). It implies that we do not actually assume that people compute a linear weighted sum prior to making an age estimate, but merely that their behavior (i.e., their age judgments) can be simulated reasonably well by such a model. Figure 1 illustrates the idea. The upper panel of Figure 1 represents the "true" state of the world, with the circle to the left being the true, but usually unknown, age of a target person (j).[1] The age of this person manifests itself in various ways, such as changes in skin texture, cardioidal strain (cf. Mark, Shaw, & Pittenger, 1988), hair color, and numerous other variables as indicated by the boxes. The strength of the effect is represented by β. It is important to note that each person j may differ in the degree to which any of the indicators changes as a function of age. For example, for some people their hair may turn gray as they age (i.e., $\beta_{3j} > 0$), while for others this may not be the case (i.e., $\beta_{3j} = 0$). A different way to make the same statement has been chosen in Figure 1A, by postulating an overall model (fixed β's; i.e., the hair color will change for all people when they age) and person-specific deviations, which are represented by the circles to the right (δ_j). The importance of this reformulation becomes apparent when considering how people make age estimates in everyday life (Figure 1B). Here, the information on skin texture, cardioidal strain, hair color, and so on, is directly observed and combined to arrive at an age estimate. In contrast to the true model, however, all variables are comprised of true information *and* "error" (e.g., dyed hair), and there is usually no way for the perceiver to disentangle the two. The differential weighting (*b*) of the information depends on prior learning, that is, an acquired representation of the ecological model. For this purpose, we assume that people continuously make implicit or explicit age estimates and evaluate their estimates, once they find out the true age of the person in question. Thus, learning takes place by minimizing the prediction error (i.e., minimizing ε, with ε = estimated age minus true age). Finally, in some situations, like the present study, no feedback is provided, so that no learning can take place (Figure 1C). Here, the age estimates are indicative of the quality of the previously calibrated judgment model.

---

[1] We are sympathetic to the idea of conceiving of age as a true latent variable that is not simply the distance from birth or distance to death but an unknown quantity that manifests itself in various indicators. This would highlight the idiosyncratic nature of true age and do justice to the expression "you are only as old as you feel." However, for the purpose of the present article we define true age simply as time from birth.
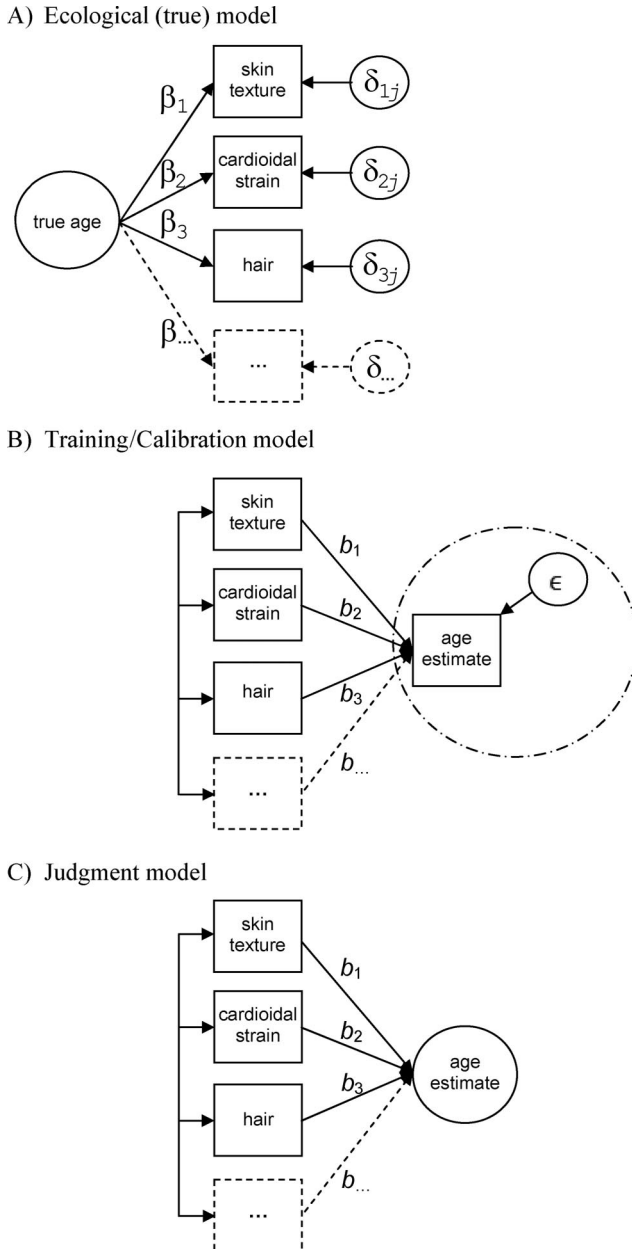
A) Ecological (true) model

B) Training/Calibration model

C) Judgment model

Figure 1. A schematic illustration of human age estimation. (A) Ecological (true) model of how age affects various physiological indicators. (B) Training/Calibration model of age estimation. The dash dotted line represents the true age. (C) Paramorphic model of human age estimation. See text for details.

Under the null hypothesis that there are no systematic group differences in the "calibration" phase, we would expect some perceivers to provide better age estimates than others and some targets to be more difficult to estimate than others, but in the long run (taking the average across perceivers and posers, respectively) these differences should cancel out and age estimates should converge on the true ages. In the present article, however, we put forth the alternative hypothesis that the quality of the age estimation model differs systematically for different groups of perceivers

(young vs. older and male vs. female), different groups of posers (young vs. older and male vs. female), different target expressions (angry, disgusted, fearful, happy, neutral, sad), and different combinations of perceiver and poser characteristics. We presume that these differences derive primarily from different learning opportunities (i.e., differentially well calibrated judgment models).

Study Predictions

Role of Age and Gender of Perceiver

Research on face recognition (Anastasi & Rhodes, 2005; Bäckman, 1991; Harrison & Hole, 2009; He, Ebner, & Johnson, 2011; Isaacowitz et al., 2007) and emotion recognition (Ebner & Johnson, 2009; Ruffman, Henry, Livingstone, & Phillips, 2008; Suzuki, Hoshino, Shigemasu, & Kawamura, 2007) has repeatedly shown that young adults perform better than older adults. This has been explained in terms of age-related differences in cognitive, neurobiological, and/or socioenvironmental factors (cf. Ruffman et al., 2008). Although caution is advised when comparing results on face/emotion recognition with age estimation, in parts similar underlying mechanisms may be at work. For instance, a general decline in cognitive abilities may result in an inferior calibration model (i.e., the *b*-coefficients in Figure 1B) or in "negative learning" in the sense that a formerly good model is replaced by a worse model. In either case, we would expect that the linear combination of information predicts the age of the target person less well for older than for younger perceivers, leading to the hypothesis that the *accuracy of age estimation decreases across the adult life span (Hypothesis 1a)*. That is, we expected older adults to provide less accurate age estimates than young adults. Consistent with prior research on age estimation (Henss, 1991; Vestlund et al., 2009), we also expected *age overestimation by older perceivers and age underestimation by young perceivers (Hypothesis 1b)*. Although reported in prior research on age estimation, by itself the latter hypothesis is difficult to deduce theoretically. It may rather be the inevitable consequence of overestimating young, and underestimating older stimuli, in combination with an own-age advantage in age estimation, as will be outlined in more detail below.

With respect to gender of the perceiver, we expected that *women are more accurate in age estimation than men (Hypothesis 2)*. To date there exist only a few studies on gender differences in age estimation, which report slightly better performance in female raters. For example, Nkengne et al. (2008) found that female rates provide more accurate ratings than male raters, who tended to underestimate the age of faces. However, only female faces were used in this study, so the result may also be the consequence of an underlying own-gender advantage. Likewise, Vestlund et al. (2009) provide some support for a better performance in age estimation in women, by demonstrating that female raters provided less biased age estimates for older faces than men. However, additional support for the hypothesis comes from research on face recognition, which suggests better face recognition by women than men (especially for female as compared with male faces, see Lewin & Herlitz, 2002). In terms of the paramorphic model, this would mean that female perceivers generally exhibit better judgment weights than male perceivers, which we find difficult to explain on theoretical grounds. Nevertheless, in the attempt to replicate previous findings, we expected a small advantage for

female perceivers, but at the same time anticipated the effect to be qualified by the gender of the target person (see below).

Role of Age, Gender, Expression of the Target Face, and Interactions With Perceiver Characteristics

Regarding the effect of the target persons' age, prior findings are mixed. While there is some indication of age-related deficits in evaluating older faces (Fulton & Bartlett, 1991), other studies report no differences in age estimation performance between young and older faces (Burt & Perrett, 1995). Considering the huge range of person-specific internal (e.g., genetic) and external (e.g., living environment) factors that may alter a person's appearance throughout his or her life span, we expect at least two age-related changes in the ecological model of Figure 1A. First, the "signal," that is the relationship between true age and any indicator of age, should become weaker. For example, for babies, weight or height are very good indicators of age (high β's), while these variables are fairly uninformative for estimating the age of adults. With increasing age, other indicators (such as hair color, skin texture) become more important, but the average signal will generally be weaker because of lower correlations between phenotype and genotype. Second, through prolonged interactions with different environments, the "noise" (i.e., δ_j) will increase. For example, an older person who spent most of his or her life outside, exposed to the sun, is likely to have more wrinkles than someone of the same age with an office job. In young age, this difference may be negligible. Taken together, due to age-related decrease in "signal" and increases in "noise," we expected *more accurate age estimates for young than older faces (Hypothesis 3)*.

Empirical evidence suggests that people are better at recognizing and interpreting faces by individuals who are similar to themselves than individuals who are dissimilar (Anastasi & Rhodes, 2005; Armony & Sergerie, 2007; Dehon & Brédart, 2001; Ebner & Johnson, 2009; Elfenbein & Ambady, 2003; Harrison & Hole, 2009; He et al., 2011). Several mechanisms to explain these in-group effects have been proposed, such as better knowledge of individuals belonging to a group with which one self-identifies, or a higher motivation to attend to, and process, the appearances of such individuals. It stands to reason that age-group membership may have similar effects when it comes to estimating the age of other persons (Rhodes, 2009). This would correspond to an age estimation model that is better calibrated for groups of faces one is more familiar with, or has a greater interest in, than for less familiar or less interesting groups of faces. We therefore expected the main effect of the age of the face to be qualified by the perceiver's age such that *older raters provide more accurate and less biased estimates for older than younger faces, whereas young raters provide more accurate and less biased estimates for young than older faces (own-age advantage,[2] Hypothesis 4)*.

The literature also offers some insights regarding possible effects of face's gender on age estimation. Dehon and Brédart (2001) found age estimates for male faces to be more precise than age estimates for female faces. Moreover, research on perceived attractiveness shows stronger negative correlations between perceived age and attractiveness for women than men (Henss, 1991). Thus, perceived age may be more important for women, and women may be willing to take more efforts to look younger. Compared with men, women may also exhibit more inter- and intraindividual variability in their efforts to look younger. Even

when controlling for the most obvious attempts to control or conceal age-related changes in appearance, like putting on make-up, such behavior is likely to increase the noise (δ's) in the true model (Figure 1A) and systematically bias the age estimates downward (i.e., judging a person as younger than he or she actually is) in the judgment model (Figure 1C). Based on these considerations, we expected *less accuracy in age estimation for female than male faces (Hypothesis 5a)*. In addition, it is well-known that women develop faster in earlier life phases (e.g., earlier onset of puberty, Chumlea, 1982) but have a higher life expectancy (Heron et al., 2009). Thus, the effect of increasing noise in female faces is likely to become more pronounced as people age, so that we also assumed that *while the age of faces is overestimated for both genders when young, and underestimated when older, this effect will be stronger for female faces (Hypothesis 5b)*.

As noted above, in addition to results suggesting a somewhat better performance of face recognition by women than men, there is also some tentative evidence that women are particularly good in recognizing female faces (Lewin & Herlitz, 2002). While this is in line with a general own-group advantage (i.e., more learning opportunities in terms of the calibration model of age estimation in Figure 1B), no comparable own-gender advantage could be shown for men. This has been explained in terms of gender differences in prior knowledge of, and interest in, faces of one's own gender (e.g., as a result of exposure to pictures of female faces in magazines and elsewhere; see Lewin & Herlitz, 2002), even though empirical support for these claims is still missing. If, however, these interpretations are correct, they should not be limited to face recognition but should similarly apply to age estimation. Thus, we expected *a significant interaction between perceiver's and face's gender (own-gender advantage), such that women provide more accurate and less biased age estimates for female than for male faces (Hypothesis 6)*.

To our knowledge, there exists no prior research on the relationship between facial expression and age estimation. Based on the paramorphic model of age estimation introduced in Figure 1A, it seems likely, however, that emotional facial expression may be one of the most important "natural"[3] factors affecting the age estimation process. The reasons for this may be twofold: First, different emotional expressions may add noise to the true model by temporarily changing the appearance of the face, such as the extent and location of wrinkles. Under static conditions (one time presentation of a static face with a given facial expression), perceivers cannot distinguish the effect of true age from the effect of the momentary facial expression, which led us to predict *lower accuracy in age estimation of emotional than neutral faces (Hypothesis 7a)*. Second, different facial expressions may activate different cues for age estimates, such as perceived attractiveness. From research on attractiveness it is known that smiling faces appear more attractive (Reis et al., 1990), and that attractiveness is closely related to youth (Henss, 1991). We therefore expected *lower estimated age for happy than any other facial expression (Hypothesis 7b)*. Although the precise causal mechanisms underlying this

---

[2] Often the term *own-age bias* is used instead of *own-age advantage*. In the present study, *bias* is defined as the difference between estimated and true age, thus the term *own-age advantage* has been chosen to avoid confusion. The same applies to the term *own-gender bias*.

[3] As opposed to "artificial" factors such as make-up or plastic surgery.

relationship remain unclear at present, this may be due to aging stereotypes that associate young age with more positive characteristics and older age with more negative outcomes (e.g., Gluth, Ebner, & Schmiedek, 2010; Gruehn, Gilet, Studer, & Labouvie-Vief, 2011; Heckhausen, Dixon, & Baltes, 1989).

Method

Participants

Seventy-eight men and 76 women participated in the study ($N = 154$). All participants were Caucasian and German-speaking. Most of them were recruited through the research participant pool of the Max Planck Institute for Human Development in Berlin, while others had heard about the study and contacted the Institute for participation. The sample consisted of three different age groups ranging from 20 to 31 years ($n = 52$, young), 44 to 55 years ($n = 51$, middle-aged), and 70 to 81 years ($n = 51$, older), with about equal numbers of men and women in each of the age groups (for details see Ebner, Riediger, & Lindenberger, 2010).

Stimuli

Facial stimuli were taken from the *FACES Life span Database of Facial Expressions*[4] (Ebner et al., 2010). In two parallel sets, the database contains 2,052 photographs of 171 men and women displaying angry, fearful, disgusted, happy, sad, and neutral facial expressions. Like the participants, all posers were of Caucasian origin and belonged to three different age groups ranging from 19 to 31 years ($n = 58$, young), 39 to 55 years ($n = 56$, middle-aged), and 69 to 80 years ($n = 57$, older), with approximately equal numbers of men and women in each of the age groups. Pictures were taken following a standardized procedure (see Ebner et al., 2010). In order to identify faces, which are as representative as possible of the Caucasian population, all target persons (faces) were recruited through an agency, which specializes in casting film extras. Out of the agency's database of more than 40,000 people, only people with an "average-type" of look were selected. In addition, pictures of the faces were taken with all jewelry, make-up, glasses, or other eye-catching items removed and posers wore identical standard gray shirts. Some example photographs are shown in Figure 2. Access to the FACES database for scientific purposes can be requested at http://faces.mpib-berlin.mpg.de.

Procedure

Each participant was assigned to one of the two FACES sets of 1,026 pictures. Faces were presented, one at a time, on a 19-inch monitor. Participants estimated (among other things) the age of the person shown on a scale ranging from 0 to 100 years. That is, participants rated the age of a given target face six times, once for each of the six facial expressions. Rating dimension and response options were presented together below the image, and age estimates had to be provided by adjusting a slider. The exact age estimate selected was shown in a box to the left of the slider. After completing responses to all questions, the next face appeared on the screen. Faces were presented in a randomized order, and each face with a given facial expression was shown only once to each participant. Rating sessions were terminated after 100 minutes each on one specific day.

Figure 2. Sample pictures of male and female faces from three different age groups with different facial expressions.

The majority of participants attended as many sessions as needed to rate all 1,026 pictures of their assigned set (average number of test sessions per person $M = 11.28$, $SD = 4.7$). All participants were financially reimbursed for their participation. The amount depended on the number of face pictures rated and on the number of testing sessions attended, and ranged between 50 and 342 EUR (about 60 to 410 USD). More detailed information on participants, stimuli, and procedure are provided in Ebner et al. (2010). The study was approved by the ethics committee of the Max Planck Institute for Human Development in Berlin.

Analyses

From a theoretical maximum of 158,004 age estimates (171 posers $\times$ 6 emotions $\times$ 154 participants) a total of 135,030 ratings were obtained (85%).[5] Directional deviation scores (i.e., *age esti-*

---

[4] In the following referred to as FACES database.

[5] Because the next face picture only appeared after all questions pertaining to the previous picture had been answered, there were no missing values in a strict sense; the missing 15% were solely due to participants terminating the study before they had rated all the pictures. However, with a median number of 1,024 (out of 1,026) faces rated by each participant, this effect is negligible. With so many age estimates, oversights are unavoidable, and the final dataset contained a few clearly accidental ratings (e.g., estimating a person in the 20s to be 100 years old). Because such extreme outliers can distort statistical analyses, we removed 185 age estimates (0.14% of the responses), which deviated more than 30 years from the target person's actual age. This decision was made on visual inspection of the data using scatter- and box-plots. While the decision could have been based on other criteria (e.g., being above or below a certain number times the interquartile range), we consider our approach to be very conservative, and thus optimal for our purposes. Repeating the analyses with the complete dataset did not change any of the main results. Changes in parameter estimates were also negligible.

mation bias) and absolute deviation scores (i.e., *age estimation accuracy*) were computed following the procedure adopted by Vestlund et al. (2009; p. 302; Brown & Siegler, 1993). We then conducted two separate multivariate analyses of variance on accuracy and bias, using face's gender, age group, and facial expression as within-subject factors, and perceivers' gender and age group as between-subjects factors. The decision to use multivariate analysis of variance as the primary data analytic approach was made because our main interest was in group mean differences. As described above, it is self-evident that some perceivers provide better age ratings than others, and some faces are easier to rate than others, but we were primarily interested in whether these differences cancel out *in the long run* (over perceivers and faces, respectively). Nevertheless, although ideal for our purposes in terms of aggregation and parsimoniousness, this approach has certain disadvantages such as reduced power (number of rows = $N = 154$ perceivers for the aggregated analysis vs. number of rows = $154 \cdot 171 = 26,334$ cases for the nonaggregated analysis), problems related to unbalanced designs due to missing values, or the loss of information on the variability between different faces. Furthermore, information on possible order effects due to fatigue, boredom, or the repeated exposition to the same face with six different emotional expressions,[6] gets lost in the aggregated analysis. While the first two shortcomings seem to be negligible in the present study, the latter two aspects may provide additional valuable information on the age estimation process and help to evaluate the robustness of the aggregated findings, in particular in the light of possible fatigue effects. For these reasons, we complemented the multivariate analysis of variance by a crossed random effects analysis with faces and perceivers as two independent random effects (cf. Pinheiro & Bates, 2000). In addition to the factors investigated in the multivariate analysis of variance, we also accounted for linear trends in age estimates due to test-session and stimulus-number, as well as for any first order interactions with these variables in the crossed random effects analysis. The purpose of the complementary crossed random effects analysis was thus twofold. First, it served as a test of the robustness of the findings obtained in the standard analysis of variance. Second, it served as a means to find out whether there were any improvements/ aggravations within or across test sessions. In contrast to linear regression or multivariate analysis of variance, however, the computation of inferential tests is less straightforward for crossed random effects models, since the usual formulas for degrees of freedom are no longer valid (Baayen, Davidson, & Bates, 2008). We thus report significance tests based on 15,000 Markov Chain Monte Carlo (MCMC) simulations using the languageR package (Baayen et al., 2008). The crossed random effects analyses were carried out using the lmer() function (Bates, Maechler, & Bolker, 2011) in R (R Development Core Team, 2010). More information on the model is provided in Appendix A.

Results

Age and Gender of Perceiver

As shown in Table 1, with a marginal mean of 6.83 years (95% CI [6.52, 7.15]) older participants provided less accurate age estimates than middle-aged ($M = 6.30$ years, 95% CI [5.98, 6.62]), or young participants ($M = 5.91$ years, 95% CI [5.59, 6.22]). As

indicated by a partial eta squared (η_p^2) of 0.11, this effect approaches a medium size (Cohen, 1988, 1992) and supports Hypothesis 1a that the accuracy of age estimation decreases across the adult life span.

Interestingly, this age-related decrease in age estimation accuracy was not unsystematic. Overall, older participants *over*estimated the age of a target person by approximately one year ($M = 0.92$ years, 95% CI [0.18, 1.65]), whereas middle-aged and young participants *under*estimated the age of a target person ($M = -1.02$ years, 95% CI [-1.76, -0.27] and $M = -1.69$ years, 95% CI [-2.41, -0.99], respectively) (see Table 1). With $\eta_p^2 = 0.16$, the effect of the perceiver's age on estimation bias was of medium size and supported Hypothesis 1b of an age overestimation by older perceivers and age underestimation by young perceivers.

Contrary to Hypothesis 2, men and women did not differ significantly in the accuracy of their age estimates (see Table 1). There was also no significant overall interaction between the perceiver's gender and age of face, multivariate $F(2, 144) = 3.03$; $p > .05$; $\eta_p^2 = 0.04$. However, when examining only older faces, women were slightly more accurate than men in their age estimates (female perceivers and older faces: $M = 7.01$ years, 95% CI [6.50, 7.53]; male perceivers and older faces: $M = 7.78$ years, 95% CI [7.27, 8.29]). This is in line with findings by Vestlund et al. (2009). At the chosen alpha level of 1%, results were also not significant in the crossed random effects analysis (which has greater power to detect even small effects), thus replicating the results of the multivariate analysis of variance. Overall, given the small effect sizes, Hypothesis 2, predicting women to provide more accurate age estimates than men, was thus not supported by the data.

Age, Gender, Facial Expression of Target Person, and Interactions With Perceiver Characteristics

As shown in Table 2, the older a face the less accurate participants were in their age estimates (older targets: $M = 7.40$ years, 95% CI [7.04, 7.76]; middle-aged targets: $M = 6.05$ years, 95% CI [5.85, 6.26]; young targets: $M = 5.59$ years, 95% CI [5.25, 5.92]). With $\eta_p^2 = 0.29$, this effect was of substantial size and in support of Hypothesis 3 (more accurate age estimates for young than older faces). Importantly, however, the effect was moderated by the age of the perceiver. As illustrated in Figure 3 and Table 3, young perceivers were quite accurate when estimating the age of target faces from their own age group, but they were highly inaccurate in estimating the age of older faces. In contrast, for older participants, the age of the target faces had little impact on the accuracy of their age estimates (average age estimation accuracy of older perceivers when rating young faces: $M = 7.17$ years, 95% CI [6.59, 7.75]; middle-aged faces: $M = 6.61$ years, 95% CI [6.25, 6.97]; older

---

[6] Note, that by treating facial expression as a repeated measures factor in the multivariate analysis of variance, we statistically account for the fact that the six facial expressions were displayed by the same person. Also, because of the completely randomized order of presentation, any order effects will cancel out when computing the mean ratings across faces and perceivers. However, even without feedback on the correct age, it may be that raters improved (or worsened) in their age estimates by the mere fact that they had seen a face before, regardless of the facial expression displayed on the previous picture. Such an effect would go undetected in an aggregated analysis.

Table 1

Effects of Age and Gender of Perceiver on Age Estimation Accuracy and Age Estimation Bias

Criterion	Effect	SS (type III)	df	MS	F	Sig.	η_p^2	Sig. (CR-MCMC)
Absolute difference (Accuracy)	Age of Perceiver	788.06	2	394.03	8.65	.000	.107	.000
	Gender of Perceiver	177.63	1	177.63	3.90	.050	.026	.028
	Age of Perceiver × Gender of Perceiver	63.19	2	31.59	0.69	.501	0^+	.592
	Error	6604.16	145	45.55				
Signed difference (Bias)	Age of Perceiver	6642.99	2	3321.50	13.43	.000	.156	.000
	Gender of Perceiver	330.92	1	330.92	1.34	.249	.009	.231
	Age of Perceiver × Gender of Perceiver	61.74	2	30.87	0.13	.883	0^+	.763
	Error	35863.11	145	247.33				

	Young	Middle	Older	Men	Women
Marginal means (Accuracy)	5.907	6.301	6.832	6.528	6.166
Marginal means (Bias)	−1.687	−1.016	0.915	−0.843	−0.349

Note. $^+$ $F < 1.0$, the unbiased effect size is zero. η_p^2 = partial eta squared, controlling for all other factors (i.e., age and gender of perceiver and target, facial expression of target, and interactions between these factors). CR-MCMC: average *p*-values of 15,000 Markov Chain Monte Carlo samples from the posterior distribution of a fitted crossed random effects model. The crossed random effects model included session and stimulus number, as well as the first order interactions between all factors and these two variables as additional covariates. Marginal means of accuracy and bias are reported for age of perceiver and gender of perceiver.

faces: $M = 6.71$ years, 95% CI [6.09, 7.34]). As also apparent from Figure 3, middle-aged faces and middle-aged perceivers fell somewhere in between these two extremes. The direction of misestimation is illustrated in Figure 4. Independent of the age of the perceiver, the age estimation bias was smallest for middle-aged faces (young perceivers: $M = -1.09$ years, 95% CI [−1.99, −0.21]; middle-aged perceivers: $M = -0.72$ years, 95% CI [−1.64, 0.20]; older perceivers: $M = 0.55$ years, 95% CI [−0.35, 1.46]). Comparing only older and young perceivers in their ratings of older and young faces, we found that older perceivers were least biased when estimating the age of older faces, while young perceivers were least biased when estimating the age of young faces (see Table 3 and Figure 4 for details).

In summary, Hypothesis 4, predicting that older raters provide more accurate and less biased estimates for older than young faces, whereas young raters provide more accurate and less biased estimates for young than older faces, was only partially supported by

the data. As expected, young participants provided the most accurate age estimates for faces from their own age group (own-age advantage). In contrast, for older participants, the age of a face had little effect on age estimation accuracy. With respect to estimation bias, the least biased estimates were obtained for middle-aged faces, regardless of the age group of the perceiver. As expected, age estimation bias for young faces was smallest if rated by young perceivers and largest if rated by older perceivers. The opposite effect was true for older faces (see Figure 4).

Table 2 shows that participants were less accurate in estimating the age of female than of male targets, multivariate $F(1, 145) = 106.5$; $p < .05$; $\eta_p^2 = 0.42$, which is consistent with Hypothesis 5a. In addition, there was a significant difference in estimation bias for male and female targets, multivariate $F(1, 145) = 680.5$; $p < .05$; $\eta_p^2 = 0.82$. On average, the age of male faces was overestimated by 0.64 years, while female faces were estimated 1.83 years younger than their actual age. As shown in Figure 5, however,

Table 2

Effects of Age and Gender of Poser on Age Estimation Accuracy and Age Estimation Bias

Criterion	Effect	Wilks' lambda	F^+	df	Sig.	η_p^2	Sig. (CR-MCMC)
Absolute difference (Accuracy)	Age of Face	0.71	29.70	2,144	.000	.292	.000
	Gender of Face	0.58	106.50	1,145	.000	.423	.002
	Age of Face × Gender of Face	0.36	130.30	2,144	.000	.644	.000
Signed difference (Bias)	Age of Face	0.23	246.40	2,144	.000	.774	.000
	Gender of Face	0.18	680.50	1,145	.000	.824	.000
	Age of Face × Gender of Face	0.33	145.90	2,144	.000	.670	.011

	Young	Middle	Older	Men	Women
Marginal means (Accuracy)	5.589	6.054	7.397	5.908	6.789
Marginal means (Bias)	3.741	−0.423	−5.105	0.635	−1.826

Note. $^+$ Multivariate *F*-test. η_p^2 = partial eta squared, controlling for all other factors (i.e., age and gender of perceiver and target, facial expression of target, and interactions between these factors). CR-MCMC: average *p*-values of 15,000 Markov Chain Monte Carlo samples from the posterior distribution of a fitted crossed random effects model. The crossed random effects model included session and stimulus number, as well as the first order interactions between all factors and these two variables as additional covariates. Marginal means of accuracy and bias are reported for gender and age group of the target person (faces).
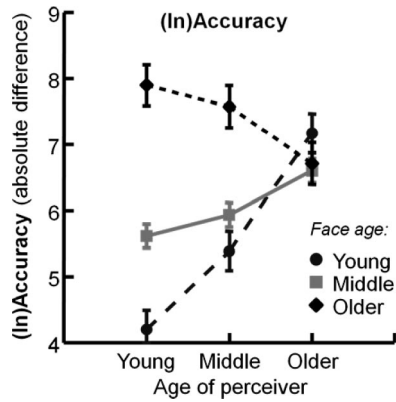
Figure 3. Age estimation accuracy: absolute difference scores (and standard errors) between estimated and true age for young, middle-aged, and older perceivers and targets (in years). The higher the absolute value, the lower the age estimation accuracy.

the gender difference in age estimation bias was only small for young targets (young male targets: $M = 4.08$ years, 95% CI [3.57, 4.59]; young female targets: $M = 3.41$ years, 95% CI [2.87, 3.94]) but much larger for middle-aged and older targets (middle-aged male targets: $M = 1.37$ years, 95% CI [0.83, 1.91]; middle-aged female targets: $M = -2.22$ years, 95% CI [−2.77, −1.66]; older male targets: $M = -3.54$ years, 95% CI [−4.17, −2.92]; older female targets: $M = -6.67$ years, 95% CI [−7.34, −6.00]). On average, older female faces were estimated more than three years younger than older male faces, thus supporting Hypothesis 5b, which predicted that while the age of faces is overestimated for both genders when young, and underestimated when older, this effect would be stronger for female faces (see Table 2 and Figure 5).[7]

The multivariate analysis of variance indicated no own-gender advantage in age estimation accuracy (see Table 3). However, even though the effect was small ($\eta_p^2 = 0.02$), it reached significance in the crossed random effects analysis using MCMC. Regarding age estimation bias, there was virtually no difference between male ($M = -1.85$ years, 95% CI [−2.46, −1.23]) and female perceivers ($M = -1.81$ years, 95% CI [−2.43, −1.19]) in their underestimation of female faces. Male perceivers, however, provided less biased age estimates of male faces, whereas female perceivers provided significantly less accurate and more positively biased age estimates of male faces (male perceivers: $M = 0.16$ years, 95% CI [−0.44, 0.76]; female perceivers: $M = 1.11$ years, 95% CI [0.50, 1.72]). Accordingly, there was an interaction between the perceiver's and face's gender on age estimation bias (crossed random MCMC $p = 0.00$; multivariate analysis of variance $p = 0.01$; see Table 3). Opposite to what we had postulated in Hypothesis 6 (a significant interaction between perceiver's and face's gender, such that women provide more accurate and less biased age estimates for female than for male faces), this was due to an own-gender advantage (less biased estimates) for male, but not female, perceivers.

As apparent from Table 3 and illustrated in Figure 6 and Figure 7, facial expressions of the posers had a large impact on participants' age estimates, age estimation accuracy: multivariate $F(5, 141) = 62.3$; $p < .05$; $\eta_p^2 = 0.69$; age estimation bias: multivariate

$F(5, 141) = 163.09$; $p < .05$; $\eta_p^2 = 0.85$. Age estimation was most accurate for targets with neutral expressions, while any other facial expression resulted in less accurate age estimates (see Figure 6). Helmert contrasts, which compared the estimation accuracy for neutral expressions with the average accuracy across all emotional expressions, indicated a significant difference, $F(1, 145) = 203.3$; $p < .05$; $\eta_p^2 = 0.59$, supporting Hypothesis 7a that it is more difficult to accurately estimate a person's age when he or she displays an emotional expression than when the facial expression is neutral. In addition, Figure 7 shows that the age of targets with happy or neutral expressions was greatly underestimated, while estimation bias for all other facial expressions was relatively small. A direct comparison of age estimation bias for happy targets to the average of all other expressions was significant, $F(1, 145) = 524.47$; $p < .05$; $\eta_p^2 = 0.78$, suggesting that a happy facial expression makes a face look younger. This is in line with Hypothesis 7b, which postulated a lower estimated age for happy than any other facial expression. With the few minor exceptions reported above, results of the multivariate analysis of variance and the crossed random effects analysis are identical. This highlights the robustness of our findings with respect to different methods of variance decomposition (random effects), as well as trends within and across sessions, which may be due to fatigue, increasing inattentiveness/boredom, or changes in age estimation performance due to repeated exposure to the same posers. Nevertheless, a closer look at these effects is insightful as will be shown next.

Trend Analyses

For age estimation bias, the fixed intercept of the crossed random effects analysis was $\hat{\beta}_{bias,intercept} = -0.57$. With all factors included as orthogonal polynomial contrasts, and session number as well as a stimulus number centered, this intercept corresponds to the overall mean in age estimation bias, which was not significantly different from zero ($p > .05$; the corresponding intercept for age estimation accuracy was $\hat{\beta}_{accuracy,intercept} = 6.32$, $p < .01$). Interestingly, however, there was a tendency for more positive age estimates over time ($\hat{\beta}_{accuracy,session} = 0.07, p < .01$). That is, while on average the underestimation of a face's age was higher during the first sessions, it eventually changed to an overestimation toward the end. Albeit small, the effect can also be observed within sessions ($\hat{\beta}_{bias,stimuli} = 0.014, p < .01$). Likewise, the age estimation accuracy improved across sessions ($\hat{\beta}_{accuracy,session} = -0.038, p < .01$), while the within-session accuracy decreased slightly ($\hat{\beta}_{accuracy,stimuli} = 0.001, p < .05$). The latter may be explained in terms of a general decline in attentiveness toward the end of a session, while the former may be due to the fact that perceivers improved their calibration model (cf. Figure 1B), even without direct feedback during the task, by simply having been exposed to the same poser before with different facial expressions. As indicated by the MCMC p values of the crossed random effects analysis in Tables 1 through 3, it is important to

---

[7] With an MCMC p value of 0.01, the interaction is only significant at the 5% level in the crossed random effects analysis. This slight difference in p values is not due to controlling for trends. Even without controlling for trends, the p value remains virtually unchanged in the crossed random effects analysis.

Table 3

Effects of Age and Gender of Perceiver × Age and Gender of Face, and Facial Expression on Age Estimation Accuracy and Age Estimation Bias

Criterion	Effect	Wilks' lambda	F^+	df	Sig.	η_p^2	Sig. (CR-MCMC)
Absolute difference (Accuracy)	Age of Perceiver × Age of Face	0.755	10.865	4, 288	.000	.131	.000
	Gender of Perceiver × Gender of Face	0.985	2.235	1, 145	.137	.015	.002
	Facial Expression	0.312	62.323	5, 141	.000	.688	.000
Signed difference (Bias)	Age of Perceiver × Age of Face	0.916	3.241	4, 288	.013	.043	.000
	Gender of Perceiver × Gender of Face	0.862	23.126	1, 145	.000	.138	.000
	Facial Expression	0.147	163.090	5, 141	.000	.853	.000

Note. $^+$ Multivariate *F*-test. η_p^2 = partial eta squared, controlling for all other factors (i.e., age and gender of perceiver and target, facial expression of target, and interactions between these factors). CR-MCMC: average *p*-values of 15,000 Markov Chain Monte Carlo samples from the posterior distribution of a fitted crossed random effects model. The crossed random effects model included session and stimulus number, as well as the first order interactions between all factors and these two variables, as additional covariates.

note that all previously discussed effects were robust when controlling for possible trends within and across sessions.

Discussion

Correct estimation of a person's age is of great importance in daily life. Together with gender, race, height, and weight, perceived age is one of the most frequently used attributes to describe an unfamiliar person. Although age estimates can often be based on multiple cues, there are many situations in which a picture of a person's face is the only information that is immediately available (e.g., pictures of missing persons, posters of electoral candidates, or Internet profiles of various kinds, ranging from online-dating platforms to faculty-member profiles on universities' Internet presence). Age estimates are important when deciding how to approach and interact with an unfamiliar person, and it stands to reason that higher age estimation ability is also related to socially more appropriate interpersonal behavior. Thus, it is important to understand the processes related to accuracy and possible biases in age estimation, and to identify factors that influence them. Despite the importance of the topic, only few studies have touched on it and there exists little theoretical and empirical knowledge. With a large set of stimulus faces and a large sample size that both cover almost the entire adult life span, the present study set out to change this situation by examining the influence of characteristics of perceivers *and* target faces on accuracy and bias in age estimation. For this purpose we conceptualized the age estimation process as a paramorphic judgment model. We report and discuss several novel findings.

Effects of Perceiver's and Poser's Age on Perception of Age

As expected, we found an age-related decline in age estimation accuracy from young over middle-aged to older perceivers. This is in line with the notion of an inferior calibration model for older adults (see Figure 1B), which may be due to age-related differences in cognitive, neurobiological, and/or socioenvironmental factors. In addition, young perceivers tended to under-, whereas older perceivers tended to overestimate the age of a person. To some degree, this is a consequence of the fact that the age of older faces was generally under-, and the age of young faces generally overestimated, but young perceivers were less biased when estimating young faces, whereas older participants were less biased when estimating older faces. The latter finding is in line with an own-age advantage (Rhodes, 2009). In terms of the paramorphic model, this difference can be explained by quantitatively more feedback, and thus smaller prediction error and better calibrated
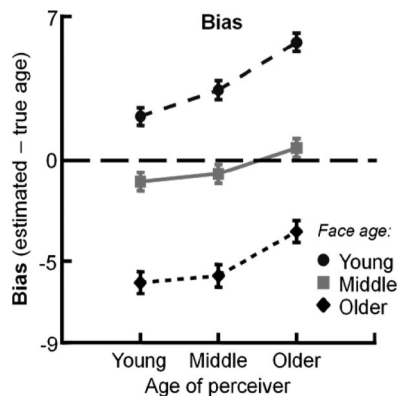

Figure 4. Age estimation bias: directional difference scores (and standard errors) between estimated and true age for young, middle-aged, and older perceivers and targets (in years).
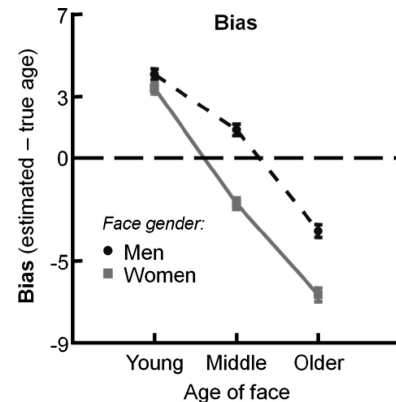

Figure 5. Age estimation bias: directional difference scores (and standard errors) between estimated and true age for young, middle-aged, and older male and female targets (in years).
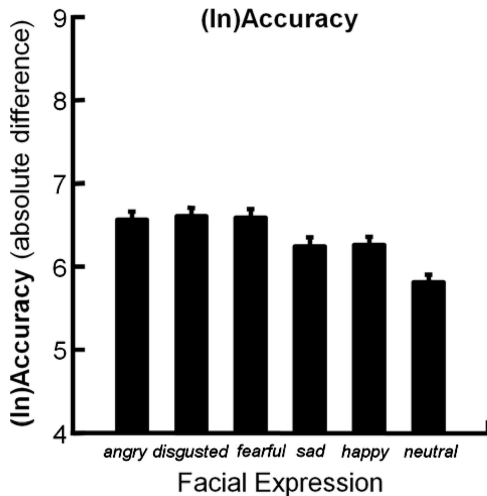
Figure 6. Age estimation accuracy: absolute difference scores (and standard errors) between estimated and true age for six different facial expressions (in years).

weights, due to differences in the amount of experience and social contact with persons of the own age group as opposed to other age groups. An alternative explanation could be qualitative differences in the feedback due to differences in social motivation and interests in members of the "in-" as compared to the "out-" group (Ebner & Johnson, 2009; Harrison & Hole, 2009; He et al., 2011). The present study supports an own-age advantage in age estimation, in that young perceivers were most accurate and least biased in their age estimates for young faces. Interestingly, across all perceiver age groups, age estimation bias was least pronounced for middle-aged faces, suggesting that an *own-age anchor effect* may be at work, such that people have a tendency to adjust their age estimates toward their own age (Vestlund et al., 2009; Ebbesen & Rienick, 1998). Likewise, the fact that for older perceivers the accuracy of age estimation is almost completely independent of the age of the poser is quite remarkable (see Figure 3), and maybe due to their experience of having been young and middle-aged once themselves. One may be inclined to expect that it is easy to be off by 10 years when estimating a 70-year-old person, while this is much less likely when estimating a 30-year-old person. For older perceivers, however, this impression is not supported by our data. We only observed the effect when averaging across all perceivers.

The overestimation of young and underestimation of older faces observed in the present study is a very robust finding not only in psychology but also in other disciplines like forensic medicine or anthropology (Aykroyd, Lucy, Pollard, & Solheim, 1997). In these research fields, age estimates are obtained by combining various indicators (e.g., samples of teeth or bones) via statistical procedures like linear regression. Because human judgment is not involved in these cases, the pattern of overestimating young and underestimating older stimuli leaves little room for a purely psychological explanation. Rather, it results from the fact that indicators for age estimation are not perfectly reliable. This is exactly the situation described in the ecological model in Figure 1A. Without knowledge on the reliability of a measure (i.e., with just a single observation of a face), it is impossible to separate the true part

from the error part. However, when fallible indicators are combined to arrive at an age estimate, the estimate will almost always tend to be too high for young and too low for older individuals—a phenomenon also known as *regression to the mean* (Barnett, van der Pols, & Dobson, 2005). It is thereby largely irrelevant of whether the indicators are combined in a linearly optimal way, like in linear regression, or by human judgment, as described by the paramorphic model in Figure 1 (cf. Dawes, 1979; Dawes & Corrigan, 1974). For a more detailed description of the phenomenon we refer the interested reader to Appendix B.

Effects of Perceiver's and Poser's Gender on Perception of Age

Age estimates for male faces were more accurate and less biased than those for female faces. On average, older female faces were estimated more than three years younger than male faces of the same age group. At this point we can only speculate about the underlying mechanism of this effect. For example, as perceived age may be more important for women than men, women may pay more attention to their physical appearance and will consequently look younger and thus will be estimated as younger. In terms of the paramorphic model, this corresponds to an increase in noise (δ's) in the true model (Figure 1A) and a subsequent downward bias in age estimates. Conceiving of age as distance to death—rather than distance from birth—one may also argue that older women actually *"are"* younger than older men from the same birth cohort because of their higher life expectancy. This interpretation is in line with recent findings showing that perceived age based on facial photographs is a robust "biomarker" (Christensen et al., 2009, p. 1) of aging. For example, Christensen et al. (2009) showed that the perceived age of twins from photographs is significantly associated with survival among those aged over 70 and is correlated with important functional and molecular ageing phenotypes, even after controlling for various background variables, including chronological age (see also Uotinen, Rantanen, & Suutama, 2005).
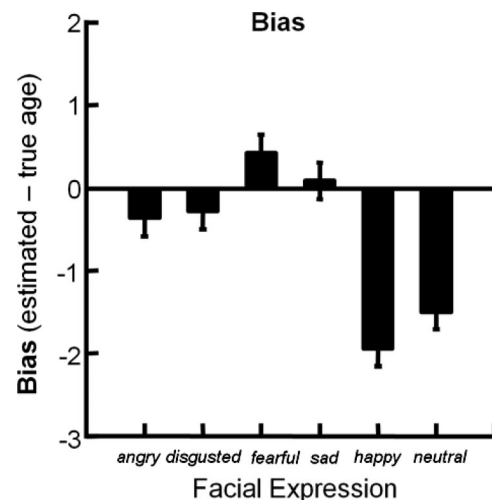

Figure 7. Age estimation bias: directional difference scores (and standard errors) between estimated and true age for six different facial expressions (in years).

In contrast to research on face recognition (Lewin & Herlitz, 2002; Herlitz & Rehnman, 2008), male and female perceivers did not differ in their age estimation ability. Also, in contrast to our expectations based on previous research (e.g., Armony & Sergerie, 2007), we found no evidence for an own-gender advantage. In particular, women did not provide better age estimates for female than male faces. Instead, it was men who were slightly better in estimating the age of male as compared with female faces. This suggests important differences in the underlying processes of gender effects between face recognition and age estimation.

Effects of Facial Expression on Perception of Age

To our knowledge the present study is the first to provide evidence that displayed facial expressions have a large effect on accuracy and bias of age estimates. As compared with neutral faces, emotional facial expressions led to more inaccurate age estimates. This corresponds to adding "noise" to the true model in Figure 3A by temporarily changing the appearance of important indicators like wrinkles. As discussed above, especially under static conditions perceivers are unable to distinguish the effect of true age from the effect of temporary facial expression, which will inevitably result in less accurate age estimates. In applied terms, this finding supports, for example, the current practice of many countries to require pictures with neutral facial expressions on official documents.

However, at the same time we found that the age of happy—but also neutral—faces was significantly underestimated. This may be explained in terms of aging stereotypes that associate young age with more positive characteristics and older age with more negative outcomes (e.g., Gluth et al., 2010; Gruehn et al., 2011; Heckhausen et al., 1989). Furthermore, smiling has been shown to enhance perceived physical attractiveness (Lau, 1982; Otta, Folladore, & Hoshino, 1996; Reis et al., 1990), and young faces are generally perceived as more attractive (e.g., Henss, 1991). Thus it stands to reason that happy facial expressions may also result in lower age estimates. However, given that this is the first study that examined the effect of emotional expression on age estimation, future research is necessary to better understand the unique contributions of such factors as age stereotypes or physical attractiveness. Also, it will be interesting to explore the influence of these factors on the relationship between *perceived* and *true* characteristics of the target person. For example, Abel and Kruger (2010) showed that smile intensity in photographs not only affected the perception of age, but actually predicted longevity. In contrast to the present study, Abel and Kruger (2010) used photographs of baseball players who debuted prior to 1952. Since most of the players had died, they could relate their smile intensity to the age at death. Although it remained unclear whether the facial expressions on these pictures were truly spontaneous, the authors argue that this was likely the case and that smile intensity reflects a general underlying disposition which is related to longevity. Other studies have shown that smile intensity on childhood and college photos is related to marriage stability and well-being up to 30 years later in life (Harker & Keltner, 2001; Hertenstein, Hansel, Butts, & Hile, 2009). To this end it appears promising to study not only posed emotions on static photographs, but also spontaneous facial expressions as brief excerpts of behavior in real-life situations (i.e., "thin slices"; cf. Ambady, Bernieri, & Richeson, 2000).

Trend Analyses

All effects reported in the present paper were robust when controlling for possible trends within and across sessions. There was a slight decrease in age estimation accuracy within sessions, presumably due to increasing fatigue toward the end of the session, but the effect was quite small (even after rating 100 face stimuli, the decrease in age estimation accuracy was only 0.1 year). Interestingly, however, age estimation accuracy improved across sessions despite the fact that the order of stimulus presentation was completely randomized. We did not expect this finding, but when considering the paramorphic model in Figure 1 it is quite plausible. With repeated exposure to different pictures of the same face (i.e., six pictures of the same poser with different facial expressions), perceivers may gain additional information to differentiate between the effect of true age and facial expression ("noise") on a given indicator. Thus, on average, one should expect the perceiver's age estimates to improve.

Conclusions and Outlook

The present study demonstrated with consistently large effect sizes that a face's age and gender impact the precision of age estimates. Furthermore, we showed that age estimation ability decreases across the adult life span. However, we also demonstrated that it is necessary to qualify perceiver effects by face characteristics (own-age advantage), and that a fair assessment of age estimation accuracy/bias requires the simultaneous combination of a large set of stimuli with a large sample size. To our knowledge, the present study is the first to show that in addition to the age of the perceiver and the face, facial expression affects accuracy and bias in age estimation. Apart from the most obvious implications of this finding (e.g., for producing and perceiving application photos), the interplay between perceived age, facial expression, and attractiveness appears to be a promising field for future research, which may improve our understanding of the perception of, attitude toward, and social interaction with, people of different ages.

References

Abel, E., L., & Kruger, M. L. (2010). Smile intensity in photographs predicts longevity. *Psychological Science, 21,* 542–544. doi:10.1177/0956797610363775

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 32, pp. 201–271). San Diego, CA: Academic Press.

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review, 12,* 1043–1047. doi:10.3758/BF03206441

Armony, J. L., & Sergerie, K. (2007). Own-sex effects in emotional memory for faces. *Neuroscience Letters, 426,* 1–5. doi:10.1016/j.neulet.2007.08.032

Aykroyd, R. G., Lucy, D., Pollard, A. M., & Solheim, T. (1997). Technical note: regression analysis in adult age estimation. *American Journal of Physical Anthropology, 104,* 259–265. doi:10.1002/(SICI)1096-8644(199710)104:2<259::AID-AJPA11>3.0.CO;2-Z

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Bäckman, L. (1991). Recognition memory across the adult lifespan: The

role of prior knowledge. *Memory & Cognition, 19,* 63–71. doi:10.3758/BF03198496

Barnett, A. G., van der Pols, J., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology, 34,* 215–220. doi:10.1093/ije/dyh299

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review, 100,* 511–534. doi:10.1037/0033-295X.100.3.511

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

Christensen, K., Thinggaard, M., McGue, M., Rexbye, H., Hjelmborg, J. v. B., Aviv, A., . . . Vaupel, J. W. (2009). Perceived age as clinically useful biomarker of ageing: cohort study. *British Medical Journal, 339,* 1433–1434. doi:10.1136/bmj.b5262

Chumlea, W. C. (1982). Physical growth in adolescence. In B. B. Wolman, G. Stricker, S. J. Ellman, P. Keith-Spiegel, & D. S. Palermo (Eds.), *Handbook of developmental psychology* (pp. 471–485). Englewood Cliffs: Prentice Hall.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. doi:10.1037/0033-2909.112.1.155

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571–582. doi:10.1037/0003-066X.34.7.571

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81,* 95–106. doi:10.1037/h0037613

Dehon, H., & Brédart, S. (2001). An 'other-race' effect in age estimation from faces. *Perception, 30,* 1107–1113. doi:10.1068/p3122

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375–38. Retrieved from http://CRAN.R-project.org/package=lme4

Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology, 83,* 745–762. doi:10.1037/0021-9010.83.5.745

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

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES - A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods, 42,* 351–362. doi:10.3758/BRM.42.1.351

Elfenbein, H. A., & Ambady, N. (2003). Universals and cultural differences in recognizing emotions of a different cultural group. *Current Directions in Psychological Science, 12,* 159–164. doi:10.1111/1467-8721.01252

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

George, P. A., & Hole, G. J. (1995). Factors influencing the accuracy of age estimates of unfamiliar faces. *Perception, 24,* 1059–1073. doi:10.1068/p241059

George, P. A., & Hole, G. J. (1998). The influence of feature-based information in the age processing of unfamiliar faces. *Perception, 27,* 295–312. doi:10.1068/p270295

George, P. A., & Hole, G. J. (2000). The role of spacial and surface cues in the age processing of unfamiliar faces. *Visual Cognition, 7,* 485–509. doi:10.1080/135062800394621

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

Gruehn, D., Gilet, A-L., Studer, J., & Labouvie-Vief, G. (2011). Age-relevance of person characteristics: Persons' beliefs about developmental change across the lifespan. *Developmental Psychology, 47,* 376–387. doi:10.1037/a0021315

Harker, L., & Keltner, D. (2001). Expressions of positive emotion in women's college yearbook pictures and their relationship to personality and life outcomes across adulthood. *Journal of Personality and Social Psychology, 80,* 112–124. doi:10.1037/0022-3514.80.1.112

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

He, Y., Ebner, N. C., & Johnson, M. K. (2011). What predicts the own-age bias in face recognition memory? *Social Cognition, 29,* 97–119. doi:10.1521/soco.2011.29.1.97

Heckhausen, J., Dixon, R. A., & Baltes, P. B. (1989). Gains and losses in development throughout adulthood as perceived by different adult age groups. *Developmental Psychology, 25,* 109–121. doi:10.1037/0012-1649.25.1.109

Henss, R. (1991). Perceiving age and attractiveness in facial photographs. *Journal of Applied Social Psychology, 21,* 933–946. doi:10.1111/j.1559-1816.1991.tb00451.x

Herlitz, A., & Rehnman, J. (2008). Sex differences in episodic memory. *Current Directions in Psychological Science, 17,* 52–56. doi:10.1111/j.1467-8721.2008.00547.x

Heron, M., Hoyert, D. L., Murphy, S. L., Xu, J., Kochanek, K. D., & Tejada-Vera, B. (2009). Deaths: Final data for 2006 *National Vital Statistics Reports* (Vol. 57): U.S. Department of Health and Human Services; Centers for Disease Control and Prevention; National Center for Health Statistics; National Vital Statistics System.

Hertenstein, M. J., Hansel, C. A., Butts, A. M., & Hile, S. N. (2009). Smile intensity in photographs predicts divorce later in life. *Motivation and Emotion, 33,* 99–105. doi:10.1007/s11031-009-9124-6

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 57,* 116–131. doi:10.1037/h0047807

Isaacowitz, D. M., Lockenhoff, C. E., Lane, R. D., Wright, R., Sechrest, L., Riedel, R., & Costa, P. T. (2007). Age differences in recognition of emotion in lexical stimuli and facial expressions. *Psychology and Aging, 22,* 147–159. doi:10.1037/0882-7974.22.1.147

Klugman, S. F. (1947). The effect of age upon estimation of age. *Journal of Social Psychology, 26,* 29–33. doi:10.1080/00224545.1947.9921729

Lau, S. (1982). The effect of smiling on person perception. *Journal of Social Psychology, 117,* 63–67. doi:10.1080/00224545.1982.9713408

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition–Women's faces make the difference. *Brain and Cognition, 50,* 121–128. doi:10.1016/S0278-2626(02)00016-7

Mark, L. S., Shaw, R. E., & Pittenger, J. B. (1988). Natural constraints, scales of analysis, and information for the perception of growing faces. In T. R. Alley (Ed.), *Social and applied aspects of perceiving faces* (pp. 11–49). Hillsdale, NJ: Erlbaum, Inc.

Meehl, P. E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis, MN: University of Minnesota Press. doi:10.1037/11281-000

Nkengne, A., Bertin, C., Stamatas, G. N., Giron, A., Rossi, A., Issachar, N., & Fertil, B. (2008). Influence of facial skin attributes on the perceived age of Caucasian women. *Journal of the European Academy of Dermatology & Venereology, 22,* 982–991. doi:10.1111/j.1468-3083.2008.02698.x

Otta, E., Folladore, A., & Hoshino, R. (1996). Reading a smiling face: Messages conveyed by various forms of smiling. *Perceptual and motor skills, 82,* 1111–1121.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects Models in S and S-plus.* New York, NY: Springer.

R Development Core Team. (2010). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Reis, H. T., Wilson, I. M., Monestere, C., Bernstein, S., Clark, K., Seidl, E., . . . Radoane, K. (1990). What is smiling is beautiful and good. *European Journal of Social Psychology, 20,* 259–267.

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

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and biobehavioral Reviews, 32,* 863–881. doi:10.1016/j.neubiorev.2008.01.001

Sörqvist, P., & Eriksson, M. (2007). Effects of training on age estimation. *Applied Cognitive Psychology, 21,* 131–135. doi:10.1002/acp.1271

Suzuki, A., Hoshino, T., Shigemasu, K., & Kawamura, M. (2007). Decline

or improvement? Age-related differences in facial expression recognition. *Biological Psychology, 74,* 75–84. doi:10.1016/j.biopsycho.2006.07.003

Uotinen, V., Rantanen, T., & Suutama, T. (2005). Perceived age as a predictor of old age mortality: A 13-year prospective study. *Age and Ageing, 34,* 368–372. doi:10.1093/ageing/afi091

Vestlund, J., Langeborg, L., Sörqvist, P., & Eriksson, M. (2009). Experts on age estimation. *Scandinavian Journal of Psychology, 50,* 301–307. doi:10.1111/j.1467-9450.2009.00726.x

Willner, P., & Rowe, G. (2001). Alcohol servers' estimates of young people's ages. *Drugs: Education, Prevention & Policy, 8,* 375–383. doi:10.1080/09687630010019299

Appendix A

Crossed Random Effects Model

The crossed random effects model in its general form, and as used in the present paper, is given in Equation A1 (Baayen et al., 2008).

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \quad with$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}), b \sim N(0, \sigma^2 \boldsymbol{\Sigma}), \mathbf{b} \perp \boldsymbol{\varepsilon} \quad (A1)$$

The vector **y** represents either bias or accuracy of the age estimate of face *j* rated by perceiver *i*. The design matrix is denoted by **X**. It contains all factors and interactions as described in the paper using orthogonal polynomial contrasts. In addition, it includes the session number and stimulus number within each session along with all first order interactions with the factors. Session

and stimulus number were centered prior to inclusion. The fixed effects are represented by **β**. The matrix **Z** is a combination of the subject and faces matrix. Because no random slopes were assumed, the two matrices are essentially two vectors of 1's. Thus, **b** contains the two random effects subject and face, both assumed to be normally distributed with a mean of zero and free variance ($\mathbf{b} \sim N(0, \sigma^2 \boldsymbol{\Sigma})$). Likewise, the errors are assumed to be normally distributed with a scaled identity covariance matrix ($\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$). The error term **ε** and **b** are treated as independent ($\mathbf{b} \perp \boldsymbol{\varepsilon}$). Restricted Maximum Likelihood was used for parameter estimation. The lmer() function of the lme4 package, along with the ancillary languageR package in R version 2.12.1 was used to do the estimation (see Baayen et al., 2008, for details).

Appendix B

Regression to the Mean in Age Estimation

What is the reason for the prevalent pattern of over- and underestimation? When participants enter the experimental situation, they are equipped with an existing calibration model (Figure 1B). When asked to estimate the age of a face, they apply the weights from this model. Other than in the training phase, however, there is no feedback given in the experimental situation. As a consequence, the age estimates will asymptotically correspond to the *estimates* of the calibration model, but not to the true ages of the ecological model. Put differently, the errors are moved from the indicators in the ecological model to the dependent variable in the judgment model. In psychological terms this is due to the fact that people cannot know to which degree a given indicator value is true or due to error. Rather, the observed indicator has to be taken as true and the error is moved to the criterion. In regression parlance, this corresponds to the assumption of error free predictors. As a result, the *difference* between true and predicted age must be positively correlated with true age. The slope of the corresponding regression line (correlation) corresponds to $1 - r^2$, with r^2 being the

amount of explained variance in the calibration model (for derivation of this relationship, see Aykroyd et al., 1997). Thus, if true age is high, residuals are more likely to be positive, while residuals are more likely to be negative if true age is low. In statistics, c*lassical calibration* is a common procedure to avoid this problem.

However, it should be noted that the pattern of over- and underestimation is not imperative. Obviously, if indicators are perfectly reliable and all indicators are available, age estimates will also be exact. Another reason may be that participants employ weights in the judgment model which differ from their calibration model (e.g., when asked to add or subtract something from their best estimate). At a more general level, it may be that the cues are not combined in a linear fashion and we encourage future research on the plausibility of the model.