Age and beauty are in the eye of the beholder

Dylan G Kwart[1], Tom Foulsham[2]§, Alan Kingstone[3]
[1] Department of Experimental Psychology, University of Oxford, Oxford, UK;
[2] Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK;
e-mail: foulsham@essex.ac.uk; [3] Brain and Attention Research Laboratory, Department of
Psychology, University of British Columbia, Vancouver, Canada
Received 2 October 2011, in revised form 31 July 2012

Abstract. How "old" and "attractive" an individual appears has increasingly become an individual concern leading to the utilisation of various cosmetic surgical procedures aimed at enhancing appearance. Using eyetracking, in the present study we aimed to investigate how individuals perceive age and attractiveness of younger and older faces and what "bottom–up" facial cues are used in this process. One hundred and twenty eight digital images of neutral faces of ages ranging from 20 to 89 years were paired and presented to subjects who judged age and attractiveness levels while having their eye movements recorded. There was an effect of face attractiveness on age-rating accuracy, with attractive faces being rated younger than their true age. Similarly, stimulus age affected attractiveness ratings, with younger faces being perceived as more attractive. Judgments of age and attractiveness were tightly linked to fixations on the eye region, along with the nose and mouth. It is thus likely that cosmetic surgical procedures targeted at the eyes, nose, and mouth may be most efficacious at enhancing one's physical appearance.

Keywords: age, aging, attractiveness, cosmetic surgery, eye movement, face perception

1 Introduction

How "old" or "attractive" an individual appears plays a major role in determining the outcomes of one's life-course trajectory including vocational success, social relationships, mental health, and overall well-being (Bleske-Rechek and Buss 2001; Chaiken 1979; Dickey-Bryant et al 1986; Farina et al 1977; Kligman and Graham 1986). Accordingly, an increasing number of individuals are making use of hundreds of invasive and non-invasive cosmetic surgical procedures, hoping to improve how old or attractive they are perceived (American Society for Aesthetic Plastic Surgery 2007; Clarke and Griffin 2008). Though past research has revealed that cosmetic surgical procedures enhancing appearance may improve patients' psychological health (see Castle et al 2002; Honigman et al 2004; Rankin et al 1998), little work has examined the specific aspects of a face that are diagnostic of age or attractiveness.

The idea that age and attractiveness perception are linked arises from evolutionary theories of interpersonal attraction, which purport that attractive mates are those who appear youthful, viable, and physically vital (Berry 2000; Burt and Perrett 1995; Jones 1996). Past experimental research has tested these ideas with studies involving subjective rating of the age and attractiveness of face images. In a variety of related studies it has been shown that subjects tend to rate older faces as unattractive, and, similarly, when rating age, attractive faces are perceived as younger (Ebner 2008; Foos and Clark 2011; McKelvie 1993; Wernick and Manaster 1984). Furthermore, recent literature has revealed that face age and attractiveness cues directly influence the neuropsychological and perceptual processing of facial stimuli (Ebner et al 2011; George and Hole 1998; Rellecke et al 2011; Sui and Liu 2009), suggesting that attentional processes may be tuned to specific diagnostic biomarkers of age and attractiveness contained within a face.

§ Author to whom all correspondence should be sent.

Older and younger faces are physically different in many ways. For example, characteristic age-related features, including skin and hair topography, colour heterogeneity, and texture, have been shown to be "telling" of age (Matts et al 2007). Structural and facial surface features also vary with age through either weight redistribution or growth resulting in stereotyped changes in skull and forehead shape, broadening of the chin, lengthening of the ears and nose, changes in the size of the eyes and surrounding eye region, and retraction of the lips (Albert et al 2007; Bruce and Young 1998; Burt and Perrett 1995; Ebner 2008). Research has also demonstrated that, apart from age, attractiveness tends to correlate most with face averageness and symmetry (Dykiert et al 2012; Valentine et al 2004; Zaidel and Cohen 2005), both of which can be affected by the above structural facial alterations associated with age.

Studies involving stimulus manipulation and subjective reports have revealed that specific age-sensitive facial features including the eye region are important in judging facial qualities such as age (George and Hole 1995; Rexbye and Povlsen 2007; Valentine et al 2004). Facial stimulus manipulation, while important to research control, can result in distorted face images that introduce salience artifacts that may not accurately capture the natural perceptual processes associated with age and attractiveness. Thus, different tools enabling insight into an observer's attentional processes may reveal mechanisms involved in age and attractiveness perception. Eyetracking involves the real-time monitoring of eye movement behaviour to reveal gaze targets which reflect the priorities of visual attention (Duchowski 2007; Parkhurst and Niebur 2002). Eyetracking has been used widely in the field of facial processing to determine visual cues in face recognition, emotion detection, and social scenes (Bindemann et al 2009; Birmingham et al 2007, 2009; Calvo and Lang 2004; Henderson et al 2005; Lundqvist and Ohman 2005; Pelphrey et al 2002; Walker-Smith et al 1977). While it is accepted that eye movement behaviour differs with a given task (Yarbus 1967), or even with gaze starting position (Arizpe et al 2012), fixations towards the eye region are consistently overrepresented (Bindemann et al 2009; Birmingham et al 2007, 2008, 2009; Nguyen et al 2009; Henderson et al 2005). However, this tendency does seem to be modulated by task instruction (Lundqvist and Ohman 2005), suggesting visual cues may be differentially utilised in specific perceptual tasks. On this basis, we sought to determine whether different facial cues are used in the age- and attractiveness-rating paradigms described above.

Recently Nguyen et al (2009) used an eyetracking paradigm to address how people process older facial stimuli while making subjective ratings of face age and fatigue. In both tasks, the authors found that the eyes were looked at most frequently followed by the forehead and nose. Further, they noticed that the glabella, the region between the eyebrows, was looked at more in faces rated the oldest. The authors concluded that cosmetic procedures should be targeted toward the eye region when intended to enhance the appearance of an individual. In addition to their novel approach, the data presented in the Nguyen et al study raise certain methodological questions.

First, the faces in Nguyen et al's study were presented individually, at the centre of the computer monitor facing participants where a fixation cross was first presented. This design makes it possible that the predominance of fixations on the eye region may have been partly because participants initially fixated at the centre of the screen, and were subsequently biased to look at this location once the face was presented. By way of contrast, we presented participants with two faces at a time, separated by a fixation cross at the centre of the screen. Because fixation began on the space between the faces, eye movements were required to select one of the faces, as well as the component features, and thus fixations were not preferentially biased by the starting location (Arizpe et al 2012). An additional advantage to presenting two faces at a time is that the judgment task involves a relative comparison, a strategy common in everyday life (Jones 2002; Tennis and Dabbs 1975).

Further, in our study stimuli represented a wide range of ages (18–89 years) as we predicted that young observers might not respond equally to all facial stimuli, independently of face age (Anastasi and Rhodes 2005). For this reason, we also analysed assessors' ratings made during the judgment tasks to specifically investigate age rating accuracy (since the actual age of facial stimuli was known) and variance in attractiveness ratings across the face-age span. Nguyen et al only presented judgers with older faces (actual ages unknown), yet discuss in their results differences in eye-movement behaviour towards 'older' (>50 years) versus 'younger' (<50 years) faces. By investigating broader trends in rating and eye movement behaviour across faces of varying age and attractiveness (including actual younger adult faces) we aimed to qualify these results.

Since the majority of past research in the field has characterised the perception and effects of beauty or appearance in terms of attractiveness, we asked participants to rate faces along two dimensions: age and attractiveness, instead of fatigue (Berry 1991; Langlois et al 1994). While Nguyen et al may have predicted that perception of fatigue would be closely related to age judgments, our methodological changes provide a bridge to the literature on attractiveness, which along with age is the main dimension which people seek to alter with cosmetic surgery. It is known that eye-movement behaviour changes when viewers are asked different questions about a visual stimulus (Yarbus 1967), thus we can extend Nguyen et al's findings by asking whether there is a bias to fixate the eye region regardless of whether one is estimating someone's age or rating his/her attractiveness. Moreover, given that a bias for people to look at the eyes of others is well established in visual cognition (Birmingham et al 2008, 2009; Henderson et al 2005), it is important to consider whether this bias is moderated by the task.

Nguyen et al were the first to use eyetracking to investigate the perceptually, and potentially clinically important, facial regions used while making assessments of a person's age and physical appearance. With the aforementioned methodological changes we expected to gain greater insight into how people judge age and beauty, thereby replicating and expanding Nguyen et al's original work. We aimed to utilise our participants' subjective ratings to investigate how age judgment-accuracy changes as a function of face age and attractiveness, as well as whether face age affects perceived attractiveness. Furthermore, we used eyetracking to test how facial features are recruited in age and attractiveness judgment tasks, and analysed whether such tasks rendered differences in eye-movement behaviour. Specifically, we predicted that visual fixations would be overrepresented in the eye region; that faces of different ages and attractiveness levels would be rated differently—with age and attractiveness rating levels being inversely related—and that these differences in ratings would be reflected in overall eye-movement behaviour.

2 Materials and methods
2.1 Stimulus preparation
One hundred and twenty eight male and female faces with known ages ranging from 20 to 89 years in age were retrieved from the CAL/PAL Face Database (Minear and Park 2004) for use in the present study. This database consists of 576 face images of Ohio and southern Michigan locals asked to stand in their street clothing with a neutral face in front of a neutral grey background under natural lighting. Photos were taken in two student unions, a shopping mall, and a variety of older adult festivals (Minear and Park 2004). This database has been widely utilised in perceptual research involving neutral facial stimuli across the lifespan (Ebner 2008; Platek and Thomson 2007; Wright et al 2007). In our study Caucasian stimuli were selected and cropped to include the entire head (including hair) and neck region. Extraneous jewelry and clothing were also cropped out with Adobe Photoshop CS. Final images were all around 640 × 480 pixels in size.

These faces were organised into categories based on actual face age (grouped into seven decades: 20–29, 30–39, and so on up to 80–89 years), gender (male or female), and attractiveness (attractive or unattractive), resulting in 28 possible age/gender/attractiveness classifications ($7 \times 2 \times 2$). Faces were first split as attractive or unattractive based on the subjective consensus of our research group, with each face being rated on the 7 point scale by the authors. This initial dichotomization (5–7 = attractive; 1–3 = unattractive) was carefully applied and later validated by participants' actual attractiveness response ratings. Faces were paired into representatives of all possible age, gender, and attractiveness combinations, and 64 pairings were compiled into a single set. A 7 point scale for either age rating (20–29 to 80–89 years) or attractiveness level (1 = not at all attractive, 7 = extremely attractive) was placed under the two images. The images were separated by a central fixation cross. Two example pairs of faces are shown in figure 1.

Two possible variations of each pair were compiled and used equally across participants: one where a given face of the presented pair was on the left and another where it was on
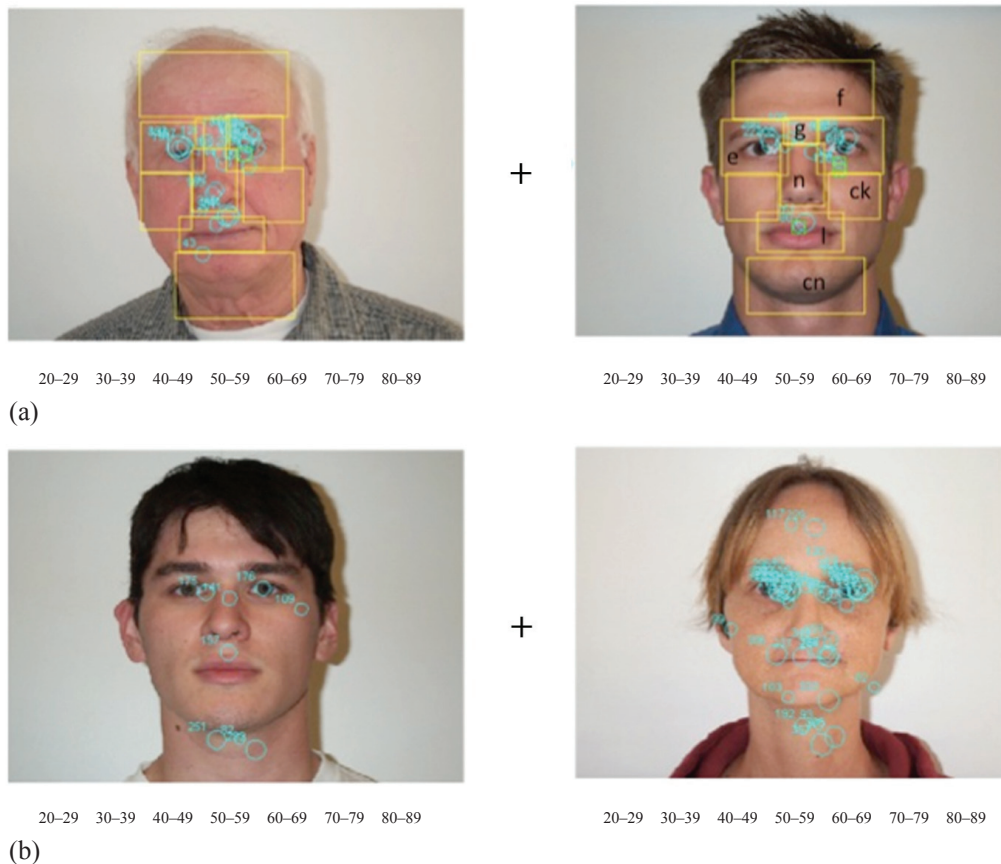

20–29　30–39　40–49　50–59　60–69　70–79　80–89 　　　　　20–29　30–39　40–49　50–59　60–69　70–79　80–89

(a)

20–29　30–39　40–49　50–59　60–69　70–79　80–89 　　　　　20–29　30–39　40–49　50–59　60–69　70–79　80–89

(b)

Figure 1. Stimulus preparation and sample age-rating trial. Split bilaterally by a central fixation cross presented first, two faces were paired and simultaneously presented. Subjects estimated the age of each face based on the age-rating scale presented below each image (20–29, 30–29 … 80–89). When rating attractiveness, the age-rating scale was replaced with a 7 point Likert scale (1 = not at all attractive; 4 = neutral; 7 = extremely attractive). Responses were recorded by a mouse click on the corresponding box. (a) Boxes represent IAs used for looking behaviour analysis (left = 80 year old attractive male, right = 22 year old attractive male). (b) Blue circles represent visual fixations; numerical values represent fixation duration in ms (left = 21 year old attractive male; right = 41 year old unattractive female). Note: f = forehead; g = glabella; e = eye region; n = nose; ck = cheek; l = lips; cn = chin. [In colour online, see http://dx.doi.org/10.1068/p7136]

the right. This procedure ensured that the side of the screen in which a face appeared was balanced across participants, as was the order of the two judgment conditions. Within each condition (age or attractiveness rating), all 64 trials were randomly presented with each face pair presented once per task.

2.2 Equipment and apparatus

Eye-movement behaviour was recorded using a SR Research EyeLink 1000 eyetracker. This equipment recorded monocular eye position throughout the trials, using the pupil image and the corneal reflection at 1000 Hz. Saccades and fixations were defined using the standard EyeLink algorithm based on velocity and acceleration thresholds (of $30°$ s^{-1} and $8000°$ s^{-2}, respectively). Participants were asked to keep their heads fixed on a chin-rest approximately 24 inches from the centre of a 13 inch $\times$ 10.5 inch (17 inch diagonal) Dell LCD Monitor on which the paired facial stimuli were presented with a central fixation cross. The screen had a resulting visual angle of approximately 31 deg by 24 deg, and faces were about 8 deg wide. Facial regions were defined on each face as interest areas (IAs) using the area selection tool on EyeLink's DataViewer software, with regions based on the study of Nguyen et al (2009), and included left and right eye regions, glabella, nose region, cheeks, lips, chin, and forehead, as shown in figure 1a. Bilateral interest areas were equally sized and scaled to the size of each face stimulus' size. Fixation count and dwell times (in ms, and in percent of trial time) were calculated for each IA.

2.3 Procedure and participants

Prior to each task, the eyetracker was aligned with the participant's right eye and a brief 9-point calibration took place. Each participant then completed the two judgment tasks, the order of which was counterbalanced across participants. Thirteen University of British Columbia students (six males; mean age = 22.69 years, SEM = 1.43 years) were tested in this study for course credit, or a small stipend. All participants signed an informed consent form prior to experimentation. Ethics Committee approval was obtained for this study.

2.3.1 *Judging age*. In each trial, participants were presented with a drift correct cross at the centre of a white screen (at which point fixation was confirmed at this location). A face pair then appeared, and participants were instructed to use the computer mouse to rate the age of each face by clicking on the correct age-range box (see figure 1 for a sample of the rating scales) under each face. Participants were allowed to change the rating choice of the face they rated first (rating was acknowledged with the appearance of an asterisk above the scale); when the rating of the second face was completed the trial was terminated and the drift correct cross reappeared, signaling the next trial. There was no time limit, ratings proceeded in a self-paced fashion and it was up to the participant which face was rated first.

2.3.2 *Judging attractiveness*. The same protocol was followed in this condition, however participants were instructed to rate the attractiveness of each face on a 7 point scale, and the position of the faces (left or right side of the screen) was reversed from that used in the previous judgment set. Following the completion of both tasks participants were debriefed about all experimental initiatives.

3 Results

We presented participants with a range of different face stimuli, and in particular we categorised each face according to age and attractiveness. Before our analyses, we first validated our subjective grouping of attractive/unattractive faces by finding a significant effect of predefined face attractiveness on participant attractiveness ratings ($F_{1,12} = 124.48$, $p < 0.001$, $\eta_p^2 = 0.91$), where "attractive" faces were indeed rated higher on the attractiveness scale ($M = 3.75$, SEM = 0.21) compared to unattractive faces ($M = 2.40$, SEM = 0.21).

To investigate the effect of these dimensions on behaviour, we dichotomised face age into old (> 50 years, $N = 64$; $M = 71.32$, SEM = 1.28) and young (< 50 years, $N = 59$; $M = 30.21$, SEM = 1.16). In each case, we then used a 2×2 repeated-measures ANOVA to compare the means for each type of face with factors of age and attractiveness.

3.1 Ratings results

3.1.1 Does attractiveness influence age ratings? First, we calculated an age rating error score by subtracting the known age decade from the age rating, converted into a 7 point scale (20–29 years = 1, 30–39 years = 2, etc) to examine how well our young subjects estimated the age of faces from different age groups and to see whether these estimates were affected by attractiveness. On this measure a positive score (for example, a 30–39 year old face rated as 50–59 years old; error = 2) indicated that the participant overestimated the age of the face, a negative score signified an underestimate, and an error of zero meant that they correctly chose the right decade. We analysed these values with an age of face (young, old) × attractiveness (attractive, unattractive) 2×2 repeated-measures ANOVA.

There was a significant effect of stimulus face age on age rating error ($F_{1,12} = 57.89$, $p < 0.001$, $\eta_p^2 = 0.83$) with mean rating errors of young faces being positive (ie they were rated older than their true ages; $M = 0.21$, SEM = 0.051), and those for older faces being negative (ie they were rated younger than their true ages; $M = -0.396$, SEM = 0.082). Further, a significant effect of face attractiveness on age rating error ($F_{1,12} = 111.82$, $p < 0.001$, $\eta_p^2 = 0.90$) shows that, on average across all ages, attractive faces were rated approximately 2.5 years younger than their actual associated decade of age ($M = -0.24$, SEM = 0.051). Conversely, the age of unattractive faces was slightly overestimated ($M = 0.050$, SEM = 0.064). A marginally significant interaction between stimulus face age and face attractiveness was found on age rating error ($F_{1,12} = 4.36$, $p = 0.059$, $\eta_p^2 = 0.27$). Follow-up, paired-samples t-tests demonstrated that the effect of attractiveness was reliable: attractive faces were rated as younger than unattractive faces, both in young faces ($t_{12} = 8.10$, $p < 0.001$, $d = 2.31$) and in older faces ($t_{12} = 4.70$, $p < 0.001$, $d = 1.33$), but the difference was larger in the former. Young faces were rated 3.7 years younger than their associated decade of age, on average, if they were attractive, while the mean attractiveness benefit for old faces was 2.2 years. A summary of these results is shown in figure 2.

3.1.2 Does face age influence attractiveness ratings? A summary of the mean attractiveness rating results for each face age decade group is depicted in figure 3. For analysis, mean
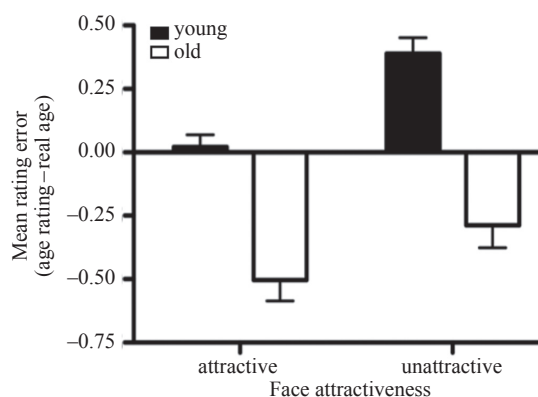

Figure 2. Summary of the mean age-rating error results (age rating – real age) by face age (young: < 50 years old; old: > 50 years old) and attractive level (attractive versus unattractive). Bars represent mean accuracy or error ± SEM.
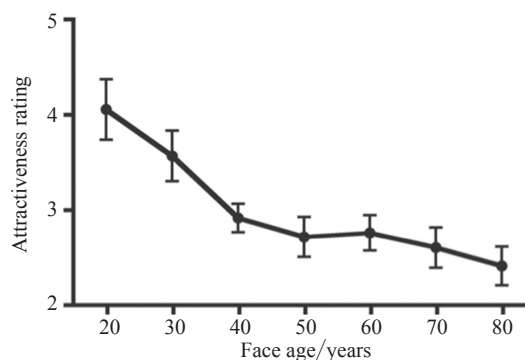
Figure 3. Summary of mean attractiveness rating results by stimulus face age. A significant effect of face age exists on attractiveness ratings. Dots represent mean rating ± SEM.

attractiveness ratings were again entered into a 2×2 repeated-measures ANOVA with face age (young vs old) and attractiveness (attractive vs unattractive) as previously described. There was a significant effect of face age on attractiveness ratings ($F_{1,12} = 41.51$, $p < 0.001$, $\eta_p^2 = 0.78$), where young faces were rated as more attractive ($M = 3.56$, SEM = 0.21) than older faces ($M = 2.61$, SEM = 0.22). Moreover, a significant interaction between face age and face attractiveness on attractiveness ratings was found ($F_{1,12} = 6.23$, $p = 0.03$, $\eta_p^2 = 0.34$). Looking at the ratings in each group of this 2×2 design we can see that the difference in mean attractiveness rating between young and old faces is larger for attractive faces (young: $M = 4.33$, SEM = 0.21; old: $M = 3.17$, SEM = 0.25; $t_{12} = 5.76$, $p < 0.001$) than unattractive faces (young: $M = 2.79$, SEM = 0.23; old: $M = 2.04$, SEM = 0.20; $t_{12} = 6.08$, $p < 0.001$). The standardised effect size in each case is similar (ds = 1.60 and 1.72, respectively).

3.2 *Eye movement results*

Participants took a mean of 10.36 s (SEM = 1.13 s) to rate both faces on age, and they made 29.68 fixations per trial on average (SEM = 3.26). Attractiveness judgments were made more quickly ($M = 7.87$ s, SEM = 0.57 s) and with fewer fixations than age judgments ($M = 22.16$ s, SEM = 1.62 s). Paired-sample t-tests confirmed that the two tasks were different in both duration ($t_{12} = 2.25$, $p < 0.05$, $d = 0.62$) and number of fixations ($t_{12} = 2.30$, $p < 0.05$, $d = 0.63$). The first fixation was always on the centre, due to the appearance of the fixation point before each stimulus. Figure 1b shows an example of the fixations made by one subject viewing a pair of faces. The remainder of our results focus on how often, and how early, different regions of interest were fixated in the different tasks. We performed this analysis by defining seven areas of interest for each face: eyes, glabella, nose, cheeks, mouth, chin and forehead.

3.2.1 *How often were different facial features fixated?* Figure 4 shows the proportion of all fixations made on each of the regions of interest, for the age and attractiveness rating tasks. In each case, fixation frequencies were summed across both faces in the display and expressed as a proportion of the total number of fixations made in each trial.

A 2×7 repeated-measures ANOVA was computed with factors of task (age vs attractiveness rating) and region type. This analysis demonstrated that there was a reliable effect of region type on the proportion of fixations ($F_{6,72} = 21.84$, $p < 0.001$, $\eta_p^2 = 0.65$). There was no main effect of task ($F_{1,12} = 1.66$, $p = 0.22$, $\eta_p^2 = 0.12$), and no interaction ($F_{6,72} = 1.00$, $p = 0.43$, $\eta_p^2 = 0.08$),
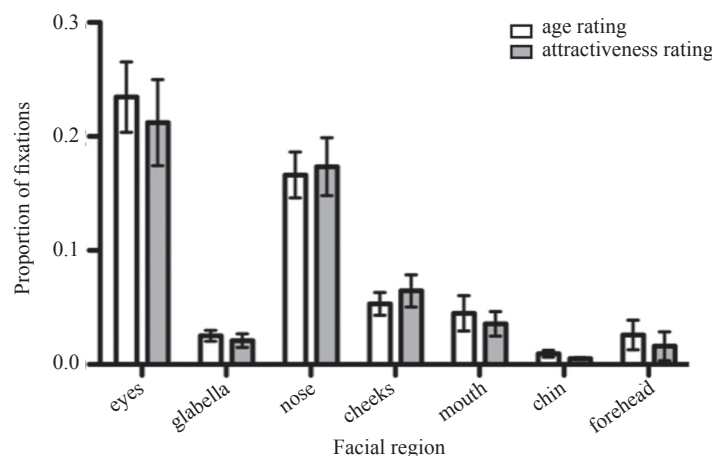

Figure 4. Proportion of visual fixations on given facial interest areas for age and attractiveness rating tasks. Bars represent the proportion of fixations (a numeric decimal of 1) per whole trial ±SEM for both tasks. Both the eyes and nose region are overrepresented as regions fixated on in both age and attractiveness rating tasks.

indicating that similar regions were fixated in both tasks. In both tasks, the eyes were the most frequently inspected regions, and they were looked at significantly more often than all other regions except the nose (Bonferroni-adjusted paired comparisons, vs glabella: $t_{12} = 6.54$, $d = 1.81$; vs cheeks: $t_{12} = 4.22$, $d = 1.16$; vs mouth: $t_{12} = 4.79$, $d = 1.33$; vs chin: $t_{12} = 6.49$, $d = 1.80$; vs forehead: $t_{12} = 5.48$, $d = 1.52$; all $ps < 0.05$; vs nose: $t_{12} = 1.18$, $d = 0.33$, $p = 0.26$). The nose was fixated more often than the glabella ($t_{12} = 5.74$, $d = 1.59$), cheeks ($t_{12} = 4.74$, $d = 1.32$), mouth ($t_{12} = 5.54$, $d = 1.54$), chin ($t_{12} = 7.62$, $d = 2.11$), or forehead ($t_{12} = 4.97$, $d = 1.38$; all $ps < 0.05$). The cheeks were fixated more often than the chin ($t_{12} = 4.73$, $d = 1.32$, $p = 0.01$), but no other comparisons were reliable.

It is also important to consider the relative size of the regions of interest, as we might expect larger regions to be fixated more often by chance alone. In our case, the pixel area of regions could not account for the differences in fixation frequency. For example, on average across the face stimuli, the forehead region covered twice as much area as the eye regions, yet it received far fewer fixations. A series of one-sample t-tests compared the proportion of fixations on each region to the relative areas of that region, expressed as a proportion of the total area of all regions. The eyes and the nose were fixated significantly more often than expected given their area (t_{12}s = 2.55 and 4.09, $ds = 0.71$ and 1.13, respectively, $ps < 0.05$). The cheeks ($t_{12} = 2.86$, $d = 0.79$, $p < 0.05$), mouth ($t_{12} = 6.19$, $d = 1.72$, $p < 0.001$), chin ($t_{12} = 61.94$, $d = 17.12$, $p < 0.001$), and forehead ($t_{12} = 20.41$, $d = 5.66$, $p < 0.001$) were fixated significantly less often than their area would suggest, while the glabella was fixated in accordance with its relative area ($t_{12} < 1$). Moreover, the results from the first fixation were very similar to those from all fixations: the eyes and nose were fixated preferentially, and other regions were rarely selected on the first fixation.

3.2.2 *Does face age or attractiveness affect eye movements?* We then compared participant means for each type of face, looking at the total time spent fixating on the face, as well as the proportion of fixations on different regions of interest. The results for total time spent fixating a face are summarised in figure 5. There was no effect of age on the total time spent fixating a face, in either task (age rating: $F_{1,12} = 1.46$, $p = 0.25$, $\eta_p^2 = 0.11$; attractiveness rating: $F_{1,12} = 1.57$, $p = 0.23$, $\eta_p^2 = 0.12$). However, there was an effect of attractiveness in both tasks. When judging age, unattractive faces were looked at for longer than attractive faces ($F_{1,12} = 6.06$, $p < 0.05$, $\eta_p^2 = 0.34$). Interestingly, the direction of this effect was reversed when judging attractiveness: in this case attractive faces were actually looked at for longer than unattractive faces ($F_{1,12} = 8.79$, $p < 0.05$, $\eta_p^2 = 0.42$). The interaction between face age and face attractiveness was not reliable in either task (age rating: $F_{1,12} = 3.07$, $p = 0.11$, $\eta_p^2 = 0.20$; attractiveness rating: $F_{1,12} < 1$).
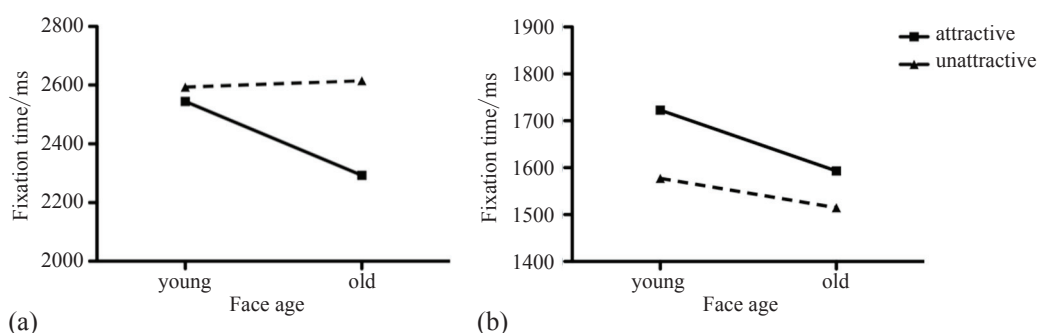

(a)   (b)

Figure 5. Summary of trends depicting time spent fixating on a face. Unattractive faces are looked at longer when judging age (a), whereas attractive faces are looked at longer when rating attractiveness (b).

Next, we computed separate $2 \times 2 \times 7$ repeated-measures ANOVAs for each task, with factors of age, attractiveness, and region of interest. The proportion of fixations on each region of interest interacted with both face age and face attractiveness, in both ratings tasks. These interactions were inspected further with pairwise comparisons between old/young and attractive/unattractive faces, in order to see which regions were inspected differently for each type of face. The results for the age rating task are summarised in table 1. When rating age, region of interest interacted with face age ($F_{6,72} = 2.45$, $p < 0.05$, $\eta_p^2 = 0.17$) and face attractiveness ($F_{6,72} = 2.76$, $p < 0.05$, $\eta_p^2 = 0.19$), but these effects were qualified by a three-way interaction between age, attractiveness and region of interest ($F_{6,72} = 2.59$, $p < 0.05$, $\eta_p^2 = 0.18$). In attractive faces, the only difference between old and young faces lay in the frequency of fixations made on the mouth. Participants spent more fixations on the mouths of young attractive faces than those of old attractive faces ($t_{12} = 2.33$, $d = 0.65$, $p < 0.05$). In unattractive faces, the nose was fixated more often in young faces than in old faces ($t_{12} = 2.79$, $d = 0.77$, $p < 0.05$). No other comparisons were significant.

Table 1. Proportion of fixations on each region of interest on different faces during the age rating task.

Region of interest	Attractive faces				Unattractive faces			
	young		old		young		old	
	M	SEM	M	SEM	M	SEM	M	SEM
Eye	0.110	0.016	0.112	0.016	0.117	0.015	0.126	0.017
Glabella	0.015	0.003	0.012	0.003	0.012	0.002	0.010	0.001
Nose	0.083	0.011	0.079	0.011	0.090*	0.012	0.074*	0.009
Cheeks	0.027	0.007	0.034	0.006	0.021	0.005	0.029	0.005
Mouth	0.026*	0.009	0.020*	0.008	0.023	0.008	0.021	0.007
Chin	0.005	0.002	0.005	0.002	0.004	0.001	0.005	0.002
Forehead	0.013	0.007	0.012	0.006	0.013	0.007	0.013	0.007

Note: Means for old and young faces are reliably different at * $p < 0.05$.

When rating attractiveness, a different pattern was observed. In this task, face age had only a marginal effect on the fixations allocated to each region (interaction with region: $F_{6,72} = 2.14$, $p = 0.060$, $\eta_p^2 = 0.15$), but the interaction between region of interest and face attractiveness was significant ($F_{6,72} = 3.37$, $p < 0.01$, $\eta_p^2 = 0.22$). There was no three-way interaction ($F_{6,72} < 1$), so old and young faces were combined and we compared the regions fixated in attractive and unattractive faces (see table 2). The nose and mouth of attractive faces were fixated more often than unattractive faces (nose: $t_{12} = 3.34$, $d = 0.93$, $p < 0.01$; mouth: $t_{12} = 2.25$, $d = 0.63$, $p < 0.05$). The number of fixations involved here was rather small. However, the standardised effect sizes were medium to large in magnitude and statistically significant. To summarise this analysis, therefore, while the eyes were invariably fixated in all types of face, fixations on the nose and mouth may be particularly diagnostic for age and attractiveness.

3.2.3 *Eye movements and the comparison between faces*. Our design, featuring two faces side-by-side, raises the question of whether participants overtly compared the two faces when making their decision. One way in which to investigate this is to look at the frequency with which saccades moved from one face to the other. In a completely serial strategy, participants would look at one face and decide upon its age/attractiveness before making a single shift to the other face. In contrast, a strongly comparative strategy might result in many more between-face shifts of attention, perhaps between particular features, before making both rating responses towards the end of the trial.

Table 2. Proportion of fixations on each region of interest on different faces during the attractiveness rating task.

Region of interest	Attractive faces		Unattractive faces	
	M	SEM	M	SEM
Eye	0.105	0.019	0.108	0.020
Glabella	0.010	0.003	0.011	0.003
Nose	0.092**	0.014	0.081**	0.012
Cheeks	0.033	0.007	0.034	0.008
Mouth	0.019*	0.006	0.017*	0.005
Chin	0.003	0.001	0.003	0.001
Forehead	0.009	0.007	0.007	0.005

Note: Means for attractive and unattractive faces are reliably different at * $p < 0.05$; ** $p < 0.01$.

Qualitative and quantitative evaluation of participant behaviour indicated a largely serial strategy. On most trials, participants inspected and rated one face before moving on to the other. On average, participants made two between-face saccades per trial, and this did not differ between tasks ($t_{12} < 1$). This was equivalent to an average of only 10% of the total number of saccades in the experiment, with the remainder shifting attention within a single face. Moreover, when a shift from one face to another did occur, in only a minority of cases was this directed between matching interest areas (eg from the nose in face a to the nose in face b, which would be a good indication of a comparison strategy). These matching saccades accounted for only 14% of between-face saccades, and again this did not differ between tasks ($t_{12} < 1$). Most participants started by fixating and rating the face on the left of the screen, before moving on to the one on the right.

4 Discussion

Until recently, no research has investigated how individuals make judgments about physical appearance in relation to the looking behaviours and facial biomarkers utilised in these tasks. Most work has relied on stimulus distortion, subjective ratings, or simple electrophysiological techniques to provide insight as to how people perceive qualities of faces including age and attractiveness (Ebner 2008; Ebner et al 2011; Foos and Clark 2011; George and Hole 1998; Matts et al 2007; Rellecke et al 2011; Wernick and Manaster 1984). Nguyen et al were first to describe the eye region as the primary tell-all facial area used when judging qualities of physical appearance. The present study sought to extend these novel results with a closer inspection of the accuracy of age ratings and their relation to attractiveness.

Presenting a central fixation cross prior to the presentation of the facial stimuli prevented visual fixations from already being localised close to the eye region, thus allowing us to obtain a purer measure of where initial saccades were targeted. Although our stimuli may have encouraged a comparison between faces, eye movement patterns suggested that this was not performed by moving back and forth between the faces. We observed that fixations towards the eyes and nose were significantly overrepresented, as expected based on Nguyen et al's results. Upon further detailed analysis, our results demonstrate how the nose and mouth regions were differentially recruited when judging either age or attractiveness.

Using facial stimuli that spanned from youth to older adulthood allowed us to specifically analyse how eye movements and subjective perceptions change depending on the age of such faces. Viewers showed a negative bias towards older adults, rating them as less attractive than younger people. A negative bias also led unattractive faces (regardless of age) to be

rated older than age-matched attractive faces. This was particularly true in younger people, and attractive young faces elicited the most accurate age ratings. Our choice in selecting attractiveness levels as a judgment task resulted in a similar set of findings that Nguyen et al purported with regards to fatigue levels and age: Age and attractiveness levels are perceptually linked, with older faces being rated less attractive and more fatigued. In our study, attractiveness ratings decreased gradually with face age, and we expect that perceived fatigue levels would increase in the same manner. Judgments of overall attractiveness may be based on a "fatigue-factor", alongside dimensions such as structural symmetry, definition of features and skin tone (Jones et al 2004).

The raw rating results for both age and attractiveness judgment tasks contribute to the growing field of psychology interested in how views and interpretations of human faces are made. We saw that face attractiveness predicts the accuracy of age judgments where attractive faces are rated as younger than unattractive ones. Interestingly, subjects tended to rate younger faces older than they actually were, while older faces were commonly rated younger. The marginal interaction found between face age and attractiveness on age rating error can be interpreted such that observers, who were not good at rating older faces in general, were not as influenced by attractiveness as in the (better discriminated) younger faces. Furthermore, participants favoured younger faces compared to older faces when judging overall attractiveness and the interaction between face age and attractiveness on attractiveness ratings (larger rating differences exist for young vs old faces in attractive people than unattractive people) demonstrates that for unattractive faces the effect of age is negligible.

The observed link between judging age and attractiveness raises some interesting questions about the potential differences reflected in eye movements during these two tasks. In general, both judgment tasks elicited the same eye movement behaviour, suggesting that the facial regions targeted in age and attractiveness perception would be efficacious targets for cosmetic intervention aiming to alter perception of these characteristics. The drive to look at these certain features suggests that these patterns are under "top–down" control, and reflect the fact that these features are important for both tasks. Nguyen et al also found a disproportionate number of fixations on the eyes and nose, a pattern that we replicated for age and attractiveness judgments. On the other hand, our participants did not show significant fixations on the forehead and glabella, as reported in that same study. While the inconsistency in findings for the forehead seems significant, given the large amount of facial area taken up by this region, the small size of the glabella, its proximity to the eyes, and the resolution of the eye tracker used by Nguyen et al (their eyetracker recorded at the rate of 50–60 Hz with an accuracy of 0.5° visual angle whereas ours recorded at 1000 Hz with an accuracy of approximately 0.25°–0.5°), renders the utility of the glabella in perceptual tasks as questionable.

Our analysis of fixations by face type extends the general picture of eye-movement behaviour by showing which regions were inspected in faces that were subsequently rated differently. Interestingly, although the eyes were fixated most frequently, the time spent on this feature was not diagnostic of face age or attractiveness. In fact, the nose and the mouth were most diagnostic for these tasks, with younger and more attractive faces drawing more fixations to these regions. This provides an incentive for further research into salient differences between these regions in old/young and attractive/unattractive faces, as well as into the efficacy of cosmetic interventions targeting these areas.

Unlike previous studies, which have often used a single face presented for a fixed duration, participants in the present study were free to distribute their gaze between the two faces on the screen for as long as they liked. We found a dissociation between the two tasks in terms of the length of time spent on each face. In general, subjects were the least accurate at rating the age of unattractive faces, and this lack of confidence may explain the increased

time spent looking at such faces in the age rating task. Whereas in the attractiveness rating task, attractive faces were looked at for longer, perhaps because when beauty is the selection feature individuals dwell longer on such stimuli that can provide aesthetic pleasure.

This bias toward attractive faces is congruent with evolutionary theories on attractiveness perception and contributes to our discussion of attentional biases to certain facial stimuli. In general, these theories imply that motivational states are conditioned through adaptation and thus create attentional biases towards increasing reproductive opportunities (Maner et al 2007). It has been strongly suggested that attention is specifically directed at individuals of the opposite sex with evolutionarily attractive qualities (young in appearance, fertile-looking, symmetrical, etc) or to other physically attractive members of the same sex who are considered threats or "intrasexual competition" (Li and Kenrick 2006; Maner et al 2003, 2007). Therefore, these evolutionary theories of attention explain our findings that young participants are significantly more attentive to other young attractive faces. Furthermore, these theories also predict that the same biases will be shown across the lifespan (even in older adults) or within gender categories.

It remains possible that the current understanding from the present and other recent studies is limited to the specific younger sample of the population that was used. For example, it is possible that young faces are deemed more attractive because all the viewers rating the facial stimuli were in that age group. When younger individuals view older adults they may make stigmatisations or judgments differently from the way in which older viewers rate similarly aged faces. Factors including social agism, the media's perceptions and other psychological theories may cause raters of different ages to respond to aging faces differently. It is also possible that the facial features recruited by older viewers, and thus the eye-movement behaviour during rating, may differ from those reported. Future research will be directed at teasing out these possible age differences and determining how cosmetic intervention should be specifically applied towards the middle and older adult population.

In sum, our results show how judging age is affected significantly by overall face attractiveness, and that attractiveness rating and face age are tightly correlated. We found that the eye region, along with the nose and mouth, were significantly and differentially overrepresented amongst regions fixated when judging age and attractiveness. Assuming the facial features fixated on reflect the visual information needed for making age and attractiveness judgments, it is likely that cosmetic surgical procedures targeted on the eye region, and potentially the nose region and the mouth region, may be most efficacious at enhancing one's perceived physical appearance.

References

Albert A M, Ricanek K Jr, Patterson E, 2007 "A review of the literature on the aging adult skull and face: implications for forensic science research and applications" *Forensic Science International* **172** 1–9

American Society for Aesthetic Plastic Surgery (ASAPS), 2007 *Cosmetic Surgery National Data Bank Statistics* Retrieved from: http://www.surgery.org. Accessed March 20, 2010

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

Arizpe J, Kravitz D J, Yovel G, Baker C I, 2012 "Start position strongly influences fixation patterns during face processing: difficulties with eye movements as a measure of information use" *PLoS ONE* **7** e31106

Berry D S, 1991 "Attractive faces are not all created equal: Joint effects of facial babyishness and attractiveness on social perception" *Personality and Social Psychology Bulletin* **17** 523–531

Berry D S, 2000 "Attractiveness, attraction, and sexual selection: evolutionary perspectives on the form and function of physical attractiveness", in *Advances in Experimental Social Psychology* Ed. M P Zanna (San Diego, CA: Academic Press) pp 273–342

Bindemann M, Scheepers C, Burton A M, 2009 "Viewpoint and center of gravity affect eye movements to human faces" *Journal of Vision* **9** 1–16

Birmingham E, Bischof W F, Kingstone A, 2007 "Why do we look at people's eyes?" *Journal of Eye Movement Research* **1** 1–6

Birmingham E, Bischof W F, Kingstone A, 2008 "Social attention and real world scenes: The roles of action, competition, and social content" *Quarterly Journal of Experimental Psychology* **61** 986–998

Birmingham E, Bischof W F, Kingstone A, 2009 "Saliency does not account for fixations to eyes within social scenes" *Vision Research* **49** 2992–3000

Bleske-Rechek A L, Buss D M, 2001 "Opposite-sex friendship: sex differences and similarities in initiation, selection, and dissolution" *Personality and Social Psychology Bulletin* **27** 1310–1323

Bruce V, Young A, 1998 *In the Eye of the Beholder: The Science of Face Perception* (Oxford: Oxford University Press)

Burt 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 B* **259** 137–143

Calvo M G, Lang P J, 2004 "Gaze patterns when looking at emotional pictures: Motivationally biased attention" *Motivation and Emotion* **28** 221–243

Castle D J, Honigman R J, Phillips K A, 2002 "Does cosmetic surgery improve psychological wellbeing?" *Medical Journal of Australia* **176** 601–604

Chaiken S, 1979 "Communicator physical attractiveness and persuasion" *Journal of Personality and Social Psychology* **37** 1387–1397

Clarke L H, Griffin M, 2008 "Visible and invisible ageing: Beauty work as a response to ageism" *Aging and Society* **28** 653–674

Dickey-Bryant L, Lautenschlager G J, Mendoza J L, Abrahams N, 1986 "Facial attractiveness and its relation to occupational success" *Journal of Applied Psychology* **71** 16–19

Duchowski A T, 2007 *Eye Tracking Methodology: Theory and Practice* (New York: Springer)

Dykiert D, Bates T C, Gow A J, Penke L, Starr J M, Deary I J, 2012 "Predicting mortality from human faces" *Psychosomatic Medicine* **74** 560–566

Ebner N C, 2008 "Age of face matters: Age-group differences in ratings of young and old faces" *Behavior Research Methods* **40** 130–136

Ebner N C, He Y, Fichtenholtz H M, McCarthy G, Johnson M K, 2011 "Electrophysiological correlates of processing faces of younger and older individuals" *SCAN* **6** 526–535

Farina A, Fischer E H, Sherman S, Smith W T, Groh T, Mermin P, 1977 "Physical attractiveness and mental illness" *Journal of Abnormal Psychology* **86** 510–517

Foos P W, Clark M C, 2011 "Adult age and gender differences in perceptions of facial attractiveness: beauty is in the eye of the older beholder" *Journal of Genetic Psychology* **172** 162–175

George P A, Hole G J, 1995 "Factors influencing the accuracy of age estimates of unfamiliar faces" *Perception* **24** 1059–1073

George P A, Hole G J, 1998 "The influence of feature-based information in the age processing of unfamiliar faces" *Perception* **27** 295–312

Henderson J M, Williams C C, Falk R, 2005 "Eye movements are functional during face learning" *Memory & Cognition* **33** 98–106

Honigman R J, Phillips K A, Castle D J, 2004 "A review of psychosocial outcomes for patients seeking cosmetic surgery" *Plastic and Reconstructive Surgery* **113** 1229–1237

Jones B C, Little A C, Feinberg D R, Penton-Voak I S, Tidderman B P, Perrett D I, 2004 "The relationship between shape symmetry and perceived skin condition in male facial attractiveness" *Evolution and Human Behavior* **25** 24–30

Jones B C, Little A C, Feinberg D R, Tiddeman B P, Pentonvoak I S, Perrett D I, 2004 "The relationship between shape, symmetry, and visible skin condition in male facial attractiveness" *Evolution & Human Behavior* **25** 24–30

Jones D G, 1996 *Physical Attractiveness and the Theory of Sexual Selection* (Ann Arbor, MI: University of Michigan Museum of Anthropology)

Jones D G, 2002 "Social comparison and body image: Attractiveness comparisons to models and peers among adolescent girls and boys" *Sex Roles* **45** 645–664

Kligman A M, Graham J A, 1986 "The psychology of appearance in the elderly" *Dermatologic Clinics* **4** 501–507

Langlois J H, Roggman L A, Musselman L, 1994 "What is average and what is not average about attractive faces?" *Psychological Science* **5** 214–220

Li N P, Kenrick D T, 2006 "Sex similarities and differences in preferences for short-term mates: What, whether and why" *Journal of Personality and Social Psychology* **90** 468–489

Lundqvist D, Ohman A, 2005 "Emotion regulates attentions: the relation between facial configurations, facial emotion, and visual attention" *Visual Cognition* **12** 51–84

McKelvie S J, 1993 "Stereotyping in perception of attractiveness, age, and gender in schematic faces" *Social Behaviour and Personality* **21** 121–128

Maner J K, Gailliot M T, Rouby D A, Miller S L, 2007 "Can't take my eyes off you: Attentional adhesion to mates and rivals" *Journal of Personality and Social Psychology* **93** 389–401

Maner J K, Kenrick D T, Neuberg S L, 2003 "Sexually selective cognition: Beauty captures the mind of the beholder" *Journal of Personality and Social Psychology* **85** 1107–1120

Matts P J, Fink B, Grammer K, Burquest M, 2007 "Color homogeneity and visual perception of age, health, and attractiveness of female facial skin" *Journal of the American Academy of Dermatology* **57** 977–984

Minear M, Park D C, 2004 "A lifespan database of adult facial stimuli" *Behavior Research Methods, Instruments, and Computers* **36** 630–633

Nguyen H T, Isaacowitz D M, Rubin P A D, 2009 "Age- and fatigue-related markers of human faces: An eye-tracking study" *Ophthalmology* **116** 355–360

Parkhurst D K, Niebur E, 2002 "Modeling the role of salience in the allocation of overt visual attention" *Vision Research* **42** 107–123

Pelphrey K A, Sasson N J, Reznick J S, Paul G, Goldman B D, Piven J, 2002 "Visual scanning of faces in autism" *Journal of Autism and Developmental Disorders* **32** 249–261

Platek S M, Thomson J W, 2007 "Facial resemblance exaggerates sex-specific jealousy-based decisions" *Evolutionary Psychology* **5** 223–231

Rankin M, Borah G, Perry A, Wey P, 1998 "Quality-of-life outcomes after cosmetic surgery" *Plastic and Reconstructive Surgery* **102** 2139–2145

Rellecke J, Bakirtas A M, Sommer W, Schacht A, 2011 "Automaticity in attractive face processing: Brain potentials from a dual task" *NeuroReport* **22** 706–710

Rexbye H, Povlsen J, 2007 "Visual signs of ageing: What are we looking at?" *International Journal of Ageing and Later Life* **2** 61–83

Sui J, Liu C H, 2009 "Can beauty be ignored? Effects of facial attractiveness on covert attention" *Psychonomic Bulletin & Review* **16** 276–281

Tennis G H, Dabbs J M Jr, 1975 "Judging physical attractiveness: Effects of judges' own attractiveness" *Personality and Social Psychology Bulletin* **1** 513–516

Valentine T, Darling S, Donnelly M, 2004 "Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces" *Psychonomic Bulletin & Review* **11** 482–487

Walker-Smith G J, Gale A G, Findlay J M, 1977 "Eye movement strategies involved in face perception" *Perception* **6** 313–326

Wernick M, Manaster G J, 1984 "Age and the perception of age and attractiveness" *The Gerontologist* **24** 408–414

Wright C I, Negreira A, Gold A L, Britton J C, Williams D, Barrett L F, 2007 "Neural correlates of novelty and face-age effects in young and elderly adults" *NeuroImage* **42** 956–968

Yarbus A L, 1967 *Eye Movements and Vision* (New York: Plenum Press)

Zaidel D W, Cohen J A, 2005 "The face, beauty, and symmetry: Perceiving asymmetry in beautiful faces" *International Journal of Neuroscience* **115** 1165–1173