Age-dependent face detection and face categorization performance

Claus-Christian Carbon\textsuperscript{1,2,*}, Martina Grüter\textsuperscript{1}, & Thomas Grüter\textsuperscript{1}

\textsuperscript{1} University of Bamberg, Department of General Psychology and Methodology, Bamberg, Germany

\textsuperscript{2} Graduate School of Affective and Cognitive Sciences, Bamberg, Germany

Correspondence should be addressed to CCC, e-mail: ccc@experimental-psychology.com,

Keywords: holistic processing; face perception; development; configural; identification; Mooney face; featural; aging; age-related; old age; visual closure
Abstract (133 words)

Empirical studies on the development of face processing skills with age show inconsistent patterns concerning qualitative vs. quantitative changes over time or the age range for peak cognitive performance. In the present study, we tested the proficiency in face detection and face categorization with a large sample of participants ($N = 312$; age range: 2-88 yrs). As test objects, we used so-called Mooney faces, two-tone (black and white) images of faces lacking critical information of a local, featural and relational nature, reflecting difficult real world face processing conditions. We found that performance in the assessment of gender and age from Mooney faces increases up to about age 15, and decreases from 65 years on. The implications of these findings are discussed in the light of classic and recent findings from face development literature.
Introduction

Visual face perception and processing in the human brain is generally believed to be an efficient and fast-running multistage process (Carbon, Schweinberger, Kaufmann, & Leder, 2005), though this view has recently been challenged by Rossion, Dricot, Goebel, and Busigny (2011) who see evidence for a non-hierarchical face perception process. Many studies have tried to clarify the mechanisms of face detection, processing, categorization (male/female, old/young, emotion, attractiveness), memory and recognition, but still, many details of the functional and anatomical foundations remain obscure (for an overview see Grüter, Grüter, & Carbon, 2008).

Several studies have shown that facial information processing involves the encoding of the shape and size of individual features like the eyes, nose and mouth (“featural” face processing, also called “analytic”, “componential” or “piecemeal”) as well as the encoding of the spatial relations and angles (Edelman & O’Toole, 2001) between those features (“configural”, “configurational” or „relational“ face processing) (Leder & Carbon, 2006; Rakover, 2002) and the holistic quality of the face (“holistic” face processing) (Goffaux & Rossion, 2006) — for a review and clarification of these terms see Piepers and Robbins (2012).

The developmental changes in face encoding and recognition have been studied extensively (for a comprehensive overview see McKone, Crookes, & Kanwisher, 2009). Face detection and face categorization, though, have not received as much attention. There is evidence that even newborns detect a frontally presented face (Goren, Sarty, & Wu, 1975) and may also be able to perform across view change (Turati, Bulf, & Simion, 2008). This is quite remarkable because the features of a prototypical frontal face — two eyes above a nose in the middle above a mouth — no longer apply when then view is changed by angular deviations of 30° degrees or more implying that even newborns have a more sophisticated way of recognizing faces. Alternatively, they may learn that facial objects continuously change their aspect without losing their perceived and processed qualities.
Doi, Koga, and Shinohara (2009) studied the detection abilities of Mooney faces in infants. In Mooney faces, named after Craig Mooney, who first used the faces to test visual closure (see Mooney, 1957), graduated colour or grayscale information is dichotomized into white or black pixels. This transformation conceals most local, featural (e.g., eyes, nose, mouth) and relational information (e.g., reliable spacing between cardinal features; also termed “2nd-order relations”). Mooneyized pictures have been used to investigate (neuro-)cognitive processes (Latinus & Taylor, 2005), neurological diseases and developmental aspects of normal and psychopathological behaviour (Webb & Quinn, 2006), and expertise-based face perception (Jemel, Pisani, Calabria, Crommelinck, & Bruyer, 2003). Doi, et al. (2009) found that 18 month-olds preferred upright Mooney faces to inverted ones, while 6 and 12 month-olds did not (yet) show any preference. In another study, Leo and Simion (2009) reported that even newborns could detect Mooneyized faces as faces. Still, other studies show that further sophisticated processing of Mooney face, e.g. recognition, is quite unmatured in 6-8 year old children and progresses until 18-21 years of age (Uhlhaas, et al., 2009). In a study with pupils of the second to eighth grade (age about 7 to 13) Mooney (1957) found a significant positive relationship between participants’ school grade and recognition performance for age categories and gender.

It has long been held that featural face processing develops in early infancy while configural and holistic face processing matures later until the age of about 10 years (Mondloch, Le Grand, & Maurer, 2002; Schwaninger, Carbon, & Leder, 2003; Schwarzer, 2006; Schwarzer, Zauner, & Korell, 2003). In a recent critical review of the literature, however, McKone, et al. (2009) found that even young children use all standard adult face recognition abilities. The authors concluded that “there is no qualitative chance in face perception beyond 4-5 years of age” (p.474). This does not say that development of face detection and recognition is finished it only states that the structures and rules for most face processing stages already exist in early childhood. Only recently, it could be shown that under certain circumstances even the quantitative face recognition ability
may be at adult level in early childhood (Crookes & McKone, 2009). Cortex specialization for face recognition emerges only gradually over the first decade of life, though, as fMRI studies have shown (for an overview see Kadosh & Johnson, 2007). Of course, cortex development does not stop then, cortical networks keep changing all life (Wu, et al., 2012). Therefore, changes in face recognition capabilities can be expected to extend well into the adult age. In a very extensive internet-based study, Germine, Duchaine, and Nakayama (2011) found that face learning abilities indeed peak not before the early thirties. This study, however, focused on specific face recognition issues tested by the “Cambridge Face Memory Test” (CFMT, Duchaine & Nakayama, 2006). No comparable study exists for face detection and categorization.

In order to shed some light on the development of face detection and face categorization, we employed high-contrast two-tone face images made of black and white providing only low residuary information and thus carrying face processing skills to the limits (Mooney, 1957). Despite this strong reduction of facial information, most people are still able to reconstruct from Figure 1b the image of a face, the angle of view, the gender, emotional expression and even identify it as a portrait of Julius Caesar — and such capabilities have been shown to be trainable (Latinus & Taylor, 2005). If and only if such “Mooneyized” displays are recognized as faces, they activate the fusiform face area (Kanwisher, Tong, & Nakayama, 1998; McKeeff & Tong, 2007; Rossion, et al., 2011) which indicates face-related processing. Specifically, performance in any Mooney face test can be assumed to reflect holistic face detection skills. It does not correlate with performance in the Gollin Incomplete Figures Test for objects and the Poppelreuter overlapping figures test (see Foreman, 1991). The fMRI results and Foreman's study indicate that Mooney face detection and categorization does not primarily employ general closure abilities, but also rely on face-specific cognitive skills. A conceivable delay of several seconds is observed with Mooney face recognition. We may assume that in this time span the brain tries to match the sparse visual information with mental images of objects or faces. Given the neural network character of these
stores, a sufficient similarity will generate a threshold potential. Due to the indeterminacy of Mooney pictures, several rivalling alternatives will be tested and discarded, before the target face is recognized. The missing information will be filled in from memory and the result will then be processed by the face processing network.

The present study

If Mooney faces are processed by a face-related neural network, this network has to be extensively trained with faces at all angles and lighting conditions before it can interpret the relatively sparse data from a Mooney face. This would require a long training phase posing the question: at what age can we expect performance to peak? Intelligence research has hypothesized that some cognitive abilities peak early in adulthood (fluid intelligence) while others increase well into late adulthood because they rely on experience (crystallized intelligence) (Cattell, 1971) — this dissociate concept of intelligence is even part of prominent intelligence scales (see McGrew, 2009). Also, one may expect performance to be quite sensitive to any degradation of the neural network, for example resulting from old age. In the present study we used Mooney’s original material and oriented towards his original procedure to study holistic face processing in a large number of participants \(N=312\) aged between 2 and 88 years in order to provide an age-adjusted standard for further studies.
Experiment

Method

Participants. We recruited 312 participants ($M_{age}=37.9$ yrs, range: 2-88 yrs; 184 female; see further details on the distribution of the ages in Table 1) who had no reported vision impairment or difficulties in recognizing faces.

<table>
<thead>
<tr>
<th>Table 1 about here</th>
</tr>
</thead>
</table>

Stimuli. Stimuli consisted of the original set of 40 Mooney faces (Mooney, 1957) which were presented on paper with a picture size of $4.7 \times 3.1$ cm (H × W). The whole set comprised 15 female faces, 25 male, 20 faces were “young” and 20 were “old”.

Procedure. Participants were asked to evaluate the face set in a pseudo-randomised sequence. First they decided whether or not the displayed stimulus was a face (“face detection task”: yes, no), then they evaluated gender (“gender decision task”: male, female) and age (“age decision task”: young, old). They were not allowed to revise their decision. In case they did not perceive any face-like object, we instructed them to omit the picture. If and only if they saw a face they were instructed to specify age (forced choice: young/old) and gender (forced choice: male/female). The procedure took about 10-15 minutes for each participant. Note: All children (approx. below 12 years of age) had an assistant, usually one of the parents, who supported the children, especially to explain the task verbally. For very young children, additional breaks were introduced as their attention was often diverted after a couple of trials. Nevertheless, even very young children were able to complete the task.

Ethical statement. After the experiment had ended participants were fully informed about the study and allowed to ask questions. Due to the simple and easy-to-conduct nature of the study,
only verbal, unwritten consent was obtained from each participant prior to the experimental session; in the case of minors, caretakers gave their approval, again, by verbal consent. Persons who did not consent were not included in the study. For each participant our records indicate that verbal consent had been obtained. As all data was collected anonymously and no harming procedures were used, ethical approval (specifically for the present) was not sought for the execution of this specific study, but the experiment was part of a more extensive project of testing participants with facial stimuli for which the local ethics committee (“Ethikrat der Otto-Friedrich-Universität Bamberg”; dated 11 February 2011) gave an umbrella approval before any of the targeted studies were conducted.

Results

We found a high recognition rate for the face detection task (Table 1): Even in the youngest and oldest group we found rates higher than 82%. For the young up to middle-aged adults, the recognition rate exceeded 92% in all age groups. Gender and age assessment performance from Mooney faces increases up to about 15 years of age, and decreases from 70 years on (Figure 2 and Table 1). We determined the percentage of correct answers from the number of detected faces, not from the total number of Mooney pictures shown. Gender assessment from recognized faces was lower than expected. From the literature we know that based on original face photographs, gender recognition is nearly perfect in adults (Cellerino, Borghetti, & Sartucci, 2004), and is about 80% correct in seven-year-old children (Wild, et al., 2000). With Mooney faces, the performance of the gender decision task was at chance level in the youngest group, and slightly above chance in the group from 6-10 years and in the oldest group. Even in the adult group, it did not exceed 85%.

Data analysis was based on three different variables: (a) rate of non-recognised faces in the face detection task, (b) correctness of age decision, and (c) correctness of gender decision. Data was assigned to age groups with a gradation of 5 years (e.g., 06-10, 11-15, etc.), with the exception of the lowest (02-05) and the highest age class (81-88).
Mooney face processing from a developmental perspective. First, the data from the participants was analysed by 3 independent one-way factorial Analyses of Variance (ANOVA), with age of the participants as between-participants factor. Rate of non-recognised faces, correctness of age and gender decision were used as separate dependent variables. Concerning the rate of non-recognised faces (on average, only 5.7% of the stimuli were not detected as faces at all), there was a main effect of participant age, $F(16, 295) = 3.22, p < .0001, \eta_p^2 = .149$, with a strong decrease in the non-detection rate in the first 15 years of life, a plateau until about the age of 70 and a strong increase from about age 70 on. Similar, but even more pronounced effects were obtained for the age and gender decision task. We found a medium-to-large main effect of age of participants, $F(16, 295) = 14.74, p < .0001, \eta_p^2 = .444$, for the age decision task. As illustrated in Figure 2 (see also Table 1), there was a strong developmental component for the first 15 years with a massive increase in performance from approx. 50% to approx. 80%. The results for the gender decision task corresponded accordingly. The ANOVA revealed significant main effect of age of participants, $F(16, 295) = 16.81, p < .0001, \eta_p^2 = .477$. Again, there was a pronounced performance improvement within the first 15 years, then a plateau of performance up to 80 years, followed by a marked deterioration from 81 years on. Further pairwise comparisons of participant age categories were conducted for the age decision task and the gender decision task, respectively. Due to the number of possible comparisons for each age category, the $\alpha$-level was Bonferroni-adjusted to $0.05/16 = .003125$. For both the age decision and the gender decision task, we have calculated the significance levels for all pairwise comparisons documenting a fully consistent data pattern for both tasks, graphically shown in Figure 2. No significant effect was found until the age of 81 and above.

-----------------------------------------------------------------------

Figures 2 about here

-----------------------------------------------------------------------
General Discussion

In the present paper, we assessed face detection and categorization performance of 312 participants aged between 2 and 88 years. We asked the participants to decide whether they saw a face in a two-tone picture („Mooney face“) and if so, to identify gender and age of the faces. Face detection as well as gender and age categorization turned out to be age-dependent. For face detection, the results show a perfect detection of 95% beginning with the age group 6-10. For the face categorization tasks (gender as well as age), we revealed a steep performance increase up to about 15 years. There is a plateau until about 65 years with a steep and ongoing decline from 81 years on, though the ceiling effect in the detection rate may hide more subtle differences in this age range. For age and gender decisions, the pairwise comparison did not reveal significant performance changes between 11 and 80 years.

The large number of participants allowed us to set up an age-dependent performance chart that can be used as a reference table for age-related performance data for Mooney face tests.

From publications on gender recognition from facial photographs (e.g., Cellerino, et al., 2004), we would have expected a better recognition rate for gender in all groups. There are several possible explanations for this phenomenon. Visual closure demands that a mental image of the face is formed and gender will be evaluated from this image. At an early age this image may be somewhat poorly defined and therefore the gender decision may be difficult. As not everyone has well-developed mental imagery (Grüter, Grüter, Bell, & Carbon, 2009), the precision of gender evaluation from mental imagery may be lower for all ages. And above all this, the participants may have expected to find a face in each Mooney picture they were shown. Given the inclination of the human visual processing system for pareidolia (seeing Gestalten in random or ambiguous stimuli, e.g., faces in clouds, the moon, or the bark of trees), the participants may have seen some face in the pictures, but not necessarily the one that was depicted. Because we asked the
participants to only evaluate gender and age if they saw a face, erroneous face identification would result in a random gender and age assignment. This would, of course, bias the results towards the lower end. While the effect is certainly not negligible, the extent is difficult to assess. In a study with children and adults, Mooney (1957) found that the participants tended to falsely recognize nonsensical black-and-white drawings as faces. Adults erroneously categorized 63% of nonsensical pictures as faces, while the percentage in children increased with their age from 54% at the second grade (about seven years) to 85% and more in the sixth to the eighth grade (about 11 to 14 years). Others, though, found much lower values for false positive identifications of nonsensical two-tone drawings. Uhlhaas, et al. (2009) reported less than 30% for all ages and Gruetzner, et al. (2010) merely 17.5% in adults. In our study, the excellent rate of correct gender categorization in the groups from 11 to 65 years (between 81.4% and 87.4%; see Table 1) does not leave much room for pareidolia. At earlier ages, the categorization results are not much better than chance. Even if there is some effect of pareidolia, the results are still surprisingly poor.

The basic face categorization mechanisms have been shown to be in place very early in life (McKone, et al., 2009), but face processing under real world conditions is often based on very incomplete visual data. Faces must be detected and categorized under difficult lighting conditions and different viewing angles. These conditions conceal details, distort the real face, cover major parts of faces and strain the visual apparatus to its limits. This is not really reflected in typical face processing test that show perfectly lit frontal views of faces. The original Mooney faces, which we used, are also shown at different angles and require visual closure. Therefore, while all qualitative face processing mechanisms may be fully functional at 4 or 5 years, the detection and categorization of faces under real-life conditions involving visual closure matures somewhat later. Visual closure comprises more than just the completion of interrupted lines. A recent MEG study (Wibral, et al., 2010) suggested that when perceiving closure in Mooney faces, brain areas are activated that relate to the processing of three-dimensional structures from shading cues and
associated with the activation of long-term memory templates. The neural networks involved can be expected to need long years of training to perfect their response. The fusiform gyrus only gets involved when the visual closure processing stage identifies the two-tone image as a face, indicating that mental images of faces are used to complete the missing parts. The perceived image of a face seems to be a combination of the picture processed from early visual fields and the mental image generated from memory.

The drop in holistic face performance at higher ages has been observed before (Read, 1988; Salthouse & Prill, 1988) but cannot be fully explained by a general deterioration of cognitive and intellectual abilities in healthy people, which is known to proceed much slower (Ritchie, Ledesert, & Touchon, 1993; Tranel, Benton, & Olson, 1997). An event-related potential study revealed that older people show delayed latencies for recognising faces, and even more important, that the distribution of respective potentials changes (Chaby, George, Renault, & Fiori, 2003). In line with this, a meta-analysis of functional neuroimaging experiments revealed a reduction of brain activity in memory-related brain areas accompanied by greater activity in the left prefrontal cortex during face recognition (Grady, 2002). If we assume that the set of cognitive functions used for Mooney face processing shifts with age, a disproportionately strong loss of performance will be observed when the cognitive reserves are exhausted. In addition, eyesight tends to worsen at advanced age (Kini, et al., 1978). In the present study, participants confirmed that their eyesight was normal or corrected-to-normal, but in older people, deterioration may go unnoticed (Tejeria, Harper, Artes, & Dickinson, 2002). Therefore, the observed performance deterioration in the Mooney face test in older people may be caused by a combination of deteriorating eyesight and overall changes in cerebral function.

Conclusions

We have shown that Mooney face recognition proficiency takes a long time to learn and decreases
markedly from about the age of 81. While the basic processes for face processing under easy perceptual conditions are in place in early childhood, it takes about an additional decade of experience to perfect the abilities for face detection and categorization under difficult conditions. The long time span of peak performance may be upheld by shifting the cognitive functions employed for the task.
Author’s note

This research was supported by a research grant from the University of Bamberg (FNK) and a grant from Doktor Robert Pfleger-Stiftung to CCC. We appreciate the participants’ efforts in taking part in the study and the great and flexible support of Doktor Robert Pfleger-Stiftung. We would like to thank Alun Brown for proofreading, Galit Yovel for editing and two anonymous reviewers for reviewing the manuscript.
References


Germine, L. T., Duchaine, B. C., & Nakayama, K. (2011). Where cognitive development and
age-dependent face processing


McKone, E., Crookes, K., & Kanwisher, N. (2009). The cognitive and neural development of face recognition in humans. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 467-482).


Rakover, S. S. (2002). Featural vs. configuralational information in faces: A conceptual and


Table 1: Correctness data (means plus SDs in parentheses) for the tasks on face detection and Mooney gender and age decision, split by the age of participants; the \( N \) of the given samples per age is displayed in extra columns.

<table>
<thead>
<tr>
<th>Age</th>
<th>( N )</th>
<th>detection</th>
<th>gender task</th>
<th>age task</th>
</tr>
</thead>
<tbody>
<tr>
<td>02-05</td>
<td>12</td>
<td>.825 (.380)</td>
<td>.506 (.221)</td>
<td>.485 (.198)</td>
</tr>
<tr>
<td>06-10</td>
<td>36</td>
<td>.956 (.205)</td>
<td>.629 (.129)</td>
<td>.566 (.103)</td>
</tr>
<tr>
<td>11-15</td>
<td>21</td>
<td>.988 (.109)</td>
<td>.846 (.076)</td>
<td>.781 (.096)</td>
</tr>
<tr>
<td>16-20</td>
<td>19</td>
<td>.976 (.152)</td>
<td>.832 (.083)</td>
<td>.750 (.117)</td>
</tr>
<tr>
<td>21-25</td>
<td>18</td>
<td>.979 (.143)</td>
<td>.850 (.045)</td>
<td>.785 (.060)</td>
</tr>
<tr>
<td>26-30</td>
<td>23</td>
<td>.979 (.142)</td>
<td>.874 (.068)</td>
<td>.778 (.073)</td>
</tr>
<tr>
<td>31-35</td>
<td>24</td>
<td>.985 (.120)</td>
<td>.874 (.053)</td>
<td>.782 (.065)</td>
</tr>
<tr>
<td>36-40</td>
<td>18</td>
<td>.972 (.164)</td>
<td>.865 (.074)</td>
<td>.801 (.058)</td>
</tr>
<tr>
<td>41-45</td>
<td>20</td>
<td>.946 (.226)</td>
<td>.829 (.108)</td>
<td>.742 (.089)</td>
</tr>
<tr>
<td>46-50</td>
<td>27</td>
<td>.973 (.162)</td>
<td>.824 (.057)</td>
<td>.766 (.062)</td>
</tr>
<tr>
<td>51-55</td>
<td>21</td>
<td>.955 (.208)</td>
<td>.814 (.092)</td>
<td>.754 (.070)</td>
</tr>
<tr>
<td>56-60</td>
<td>16</td>
<td>.956 (.205)</td>
<td>.817 (.089)</td>
<td>.769 (.071)</td>
</tr>
<tr>
<td>61-65</td>
<td>13</td>
<td>.956 (.206)</td>
<td>.848 (.098)</td>
<td>.765 (.094)</td>
</tr>
<tr>
<td>66-70</td>
<td>13</td>
<td>.952 (.215)</td>
<td>.756 (.139)</td>
<td>.675 (.112)</td>
</tr>
<tr>
<td>71-75</td>
<td>13</td>
<td>.985 (.123)</td>
<td>.794 (.093)</td>
<td>.740 (.102)</td>
</tr>
<tr>
<td>76-80</td>
<td>7</td>
<td>.914 (.280)</td>
<td>.700 (.221)</td>
<td>.661 (.211)</td>
</tr>
<tr>
<td>81-88</td>
<td>11</td>
<td>.830 (.376)</td>
<td>.605 (.196)</td>
<td>.570 (.203)</td>
</tr>
</tbody>
</table>
Figure Captions

**Figure 1.** Illustration of a Mooney face based on the manipulation proposed by Craig M Mooney — note: this specific illustration is not part of the original set of Mooney faces and was not used in the present experiment. On the left (a), an original photograph of a bust of Gaius Julius Caesar (bust is situated in Naples National Archaeological Museum; the photo stems from the Wikimedia Commons), on the right (b), a two-tone (“Mooney”) image of it with high contrasts, which was previously smoothed by a Gaussian filter.

**Figure 2.** Correctness data split by age classes for the age (red, unfilled dots) and the gender (black, filled dots) decision task, respectively. Error bars indicate 95%-confidence intervals. Additional information given by “n.s.” show non-significant pairwise comparisons, e.g. the pairwise comparison between age group “01-05” was not significant when compared with “06-10” and “81-88”, meaning all other comparisons were significant at the Bonferroni adjusted level (see body text for more information on the alpha-error correction).
Figures

Figure 1.

(a)       (b)
Figure 2

M(total)=312

- gender decision
- age decision

This paper is "in press" (PlosOne)