The Ability of Face Viewpoint Transformation Remains Immature at Age of 12: Evidence from Recognition of Faces with Different Familiarity in Children

Article in Japanese Psychological Research - May 2018
DOI: 10.1111/jpr.12190

CITATIONS
0

READS
14

2 authors, including:

Adrian Schwaninger
University of Applied Sciences and Arts Northwestern Switzerland
145 PUBLICATIONS 2,043 CITATIONS
SEE PROFILE
The Ability of Face Viewpoint Transformation Remains Immature at Age of 12: Evidence from Recognition of Faces with Different Familiarity in Children

JISIEN YANG* National Quemoy University
ADRIAN SCHWANINGER University of Applied Sciences and Arts Northwestern Switzerland

Abstract: Children are immature in face recognition, particularly in face configural decoding. This study examines the developmental difference of face recognition in another mechanism: viewpoint transformation processing. Adults and sixth-grade children were instructed to match a series of one-tone black face silhouettes to their corresponding front-view faces. This task involves sophisticated calculation in pictorial information, precise viewpoint transformation, and most probably a 3-D face representation. The results showed that although the performances were well above chance level in the recognition of familiar faces, for children, they were at chance level in the recognition of unfamiliar faces. The results indicate that children, at least to the age of about 12 years, are still immature in the processing of face viewpoint transformation compared to adults.

Key words: developmental difference, 3-D face representation, face processing in children.

Face recognition is an important and sophisticated cognitive ability that humans exhibit. Humans indispensably need rapid face recognition for effective interpersonal interaction. It is not surprising that humans are able to process faces exceedingly early in life (Goren, Sarty, & Wu, 1975; Johnson, Dziurawiec, Ellis, & Morton, 1991). However, despite the early employment of face recognition, this ability remains immature and continues to develop throughout childhood. It takes more than a decade for a child to reach the adult’s level of face recognition expertise (Carey, 1992; Feinman & Entwisle, 1976; Itier & Taylor, 2004a, 2004b; for a review see Chung & Thomson, 1995).

Face processing has been distinguished into multiple types of processing. At least three of these types have been deliberately explored to understand how humans decode the various kinds of information within faces. Component processing refers to an analytic process of deconstructing a face into a set of internal features (Freire & Lee, 2001; Freire, Lee, & Symons, 2000; Schwaninger, Lobmaier, & Collishaw, 2002; Sergent, 1984). Holistic processing refers to processing faces as whole templates without explicit part representations (Bruce, Doyle, Dench, & Burton, 1991; Diamond & Carey, 1986; Freire et al., 2000; Mendl, Le Grand, & Maurer, 2002; Rhodes,
Brake, & Atkinson, 1993). Holistic processing refers to treating faces as a gestalt regardless of the basis comprising the whole (Hole, 1994; Tanaka & Farah, 1993; Young, Hellawell, & Hay, 1987).

For decades, considerable effort has been made to explore the major aspects that can best explain the slow development in face recognition. A general consensus in the literature has been that component and holistic processing develop faster than configurational processing (Carey & Diamond, 1977, 1994; Freire & Lee, 2001; Mondloch, Dobson, Parson, & Maurer, 2004; Mondloch et al., 2002). Children appear to rely primarily on featural and holistic information to recognize a face (Carey & Diamond, 1977; Mondloch et al., 2002; Mondloch, Pathman, Le Grand, Maurer, & de Schonen, 2007). They are much less sensitive to the subtle differences in the spatial relationships between facial features.

However, although a large amount of research has been conducted to investigate the development of face recognition, this issue seems to focus mainly on the discussion regarding configurational processing (Itier & Taylor, 2004b; Mondloch et al., 2002). Comparatively less effort has been devoted to exploring other aspects of developmental difference in face recognition; for example, how well can children perform the viewpoint-transformation task in face recognition?

Davidenko (2007) developed a new methodology for face perception research. He proposed the use of two-tone silhouettes as surrogate stimuli for front-view faces. He found that the silhouettes provide enough information for sex, age, and attractiveness judgment. The most striking demonstration is in “Experiment 5” of his research. He asked adult participants to match a front-view face to four possible silhouettes, or to match a silhouette to four possible front-view faces (i.e., “four alternative force choices”). The results showed that participants’ performance (overall accuracy: 70.1%) was substantially higher than chance level (25%).

The viewpoint-transformation competence that participants exhibit is especially notable considering that the silhouette representation of the unfamiliar faces is not available in memory. Meanwhile, the pictorial information, such as texture and intermediate views, is not provided. In addition, the silhouette view is difficult to generate from a front-view face. In the domain of computer vision, recognition algorithms using 3-D face models were developed two decades ago (Blanz & Vetter, 1999; Vetter, 1998). Although one sole non-accidental 2-D view might be sufficient for the construction of a computer 3-D face model (Beymer & Poggio, 1995; Poggio, 1991; Poggio & Vetter, 1992), Poggio and Vetter (1992) suggest that the sole face view cannot be the exact front view. They stated:

One should avoid to use in the database a model view which is a fixed point of the symmetry transformations (since the transformation of it generates an identical new view). In the case of faces, this implies that the model view in the data base should not be an exactly frontal view. (p. 15)

Basically, the front-view face is singular and a second view cannot be computed from it (Schyns & Bulthoff, 1993). Accordingly, the high performance that participants demonstrated in the experiment of Davidenko (2007) must have involved a more complex face representation. This representation might involve sophisticated calculation of the distances between face features. It would also need to process the shading and shadow information contained in the face photographs to construct the shape of face features and the whole face (Bruce & Langton, 1994). In addition, it would demand precise and highly proficient viewpoint-transformation competence. Finally, it might also implicate explicit 3-D face representation of the feature shape as well as the whole face (Tong & Nakayama, 1999). In other words, the complicated calculation and representation being applied in the task might take years to fully develop before reaching a high proficient level.

In our previous research (Schwaninger & Yang, 2011), we found evidence to support the application of 3-D face representation in
adults. However, it is not clear how children perform in this 3-D face-transformation task, so we decided to explore this topic in order to better understand the development of face recognition.

This research aims to explore children’s maturity of 3-D face representation in face recognition. The “face silhouette versus front-view faces matching” paradigm developed by Davidenko (2007) was adopted in this study. Three groups of sixth-grade children and one adult group with different extents of familiarity to the stimulus faces were recruited to participate in the experiment. We recruited students from multiple schools to impose different extents of familiarity with the stimulus faces. The schools are far away from each other and the probability of the students in each school knowing each other is extremely low.

**Method**

**Participants**

We recruited 59 sixth-grade students for three child groups and 30 adults for one adult group to participate in this experiment. Twenty children (nine girls and 11 boys, $M_{age} = 11.6$ years) were recruited from class A and 21 (10 girls and 11 boys, $M_{age} = 11.7$ years) were from class B in a primary school in Taiwan. Another 16 students (nine girls and seven boys, $M_{age} = 12$ years) were recruited from a different primary school in Taiwan. The 30 adults were recruited from a university in Taiwan (20 females and 10 males, $M_{age} = 18.69$ years).

Based on the students’ familiarity with the face stimuli used in this experiment, the three different child groups were designated as “same class,” “neighbor class,” and “novel class,” respectively. It is worth noting that classes A and B were from a small primary school. Students of this school had been randomly assigned to class A or B when they had initially enrolled, and there are only two classes in each grade. Students study with the same classmates throughout their 6-year primary education. Despite being in different classes, students of the same grade have many chances to encounter and interact with the students in their neighbor class in the small campus since there are only two classes in each grade. By contrast, the novel class group attended a primary school in a different county and these students all claimed to be unfamiliar with the face stimuli used in the experiment. Thirty adult participants were recruited from a university located on an outlying island of Taiwan and all were also unfamiliar with the face stimuli used in the experiment.

**Materials**

**Face stimuli.** Using a Nikon Coolpix 3100, we took gray-scale front-view and right-profile-view photographs of the 21 participants from class B of the sixth-grade, Wen-Shen Primary School, and of two adult men and two adult women from National Chung-Chung University. The photographs of these 25 persons were used as stimuli for compiling the test booklet used for this research. The right profile view photographs were transferred into one-tone black silhouette images using Photoimpact 10 software. The silhouettes were trimmed in four different ways: whole silhouette, head silhouette, front silhouette, and facial silhouette. Figure 1 displays examples of the stimuli. The whole silhouette included the top breast, neck, and the whole head (Figure 1b). The head silhouette included just the head (Figure 1c). The front silhouette included only the front part of the head (Figure 1d). The facial silhouette included just the facial part image beneath the forehead (Figure 1e).

**Test booklets.** Each booklet was composed of nine pages in A4 landscape format. One target page contained 25 grayscale front-view face images (as illustrated in Figure 2) numbered from 1 to 25. The eight test pages had the same $5 \times 5$ grid in which one-tone silhouette stimuli were displayed as illustrated in Figure 2. The four types of one-tone silhouettes were shown in upright and inverted
orientations resulting in eight conditions (A–H). There were a total of 200 one-tone silhouette stimuli, 25 (faces) × 8 (one-tone silhouette conditions), that were randomly distributed into the 25 cells in each of the eight test pages. The target page was placed above and the eight test pages were placed below the target page in a pile. Participants could turn over the test pages from the right to the left while the target page remained stationary.

**Procedure**

Participants were instructed to match the one-tone silhouettes on the test pages to the grayscale front-view photographs on the target page by writing down the corresponding number of the target face (from 1 to 25) inside the test face cell. Participants were encouraged to guess if they were not able to recognize the faces and were informed that there was no time constraint for completing the experiment.

**Figure 1** Examples of the stimuli used. (a) Front view (target face), (b) whole silhouette, (c) head silhouette, (d) front silhouette, and (e) facial silhouette.

**Figure 2** Example of one test page. The dashed line denotes the A4 size of the original test page.
All participants completed the experiment within 1 hr.

**Results**

Accuracy was subjected to a split plot analysis of variance (ANOVA) with orientation (upright vs. inverted) and silhouette type (as illustrated in Figure 1) as within-participant factors, and group (same class, neighbor class, novel class, and adult) as between-participant factor. Figure 3 displays means and standard errors of percent correct responses (accuracy) broken up by group, orientation, and silhouette type. Chance performance equals 0.04. Bonferroni-corrected pair-wise comparisons were carried out for post hoc multiple comparisons when the ANOVA revealed significant effects. The ANOVA revealed main effects of group, $F(3, 83) = 73.499, \text{MSE} = 2.480, p < .001$, $\eta^2 = .727$, orientation, $F(1, 83) = 802.472, \text{MSE} = 8.235, p < .001$, $\eta^2 = .906$, and silhouette type, $F(3, 249) = 689.946, \text{MSE} = 3.897, p < .001$, $\eta^2 = .893$. Bonferroni-corrected pair-wise comparisons revealed that the performance of the same class group ($M = .413$) was equivalent to that of the neighbor class group ($M = .369; p = .190$). The performance of both the same class and neighbor class was better that that of the adult group ($M = .271; p < .001$). Finally, the adult group performed better than the novel class group ($M = .116; p < .001$). Performance for upright stimuli ($M = .404$) was better than for inverted stimuli ($M = .180$), that is, the inversion effect was clearly demonstrated (Yang & Schwaninger, 2010; Yin, 1969). Finally, Bonferroni-corrected pair-wise comparisons revealed that performance was equivalent between the whole silhouette ($M = .428$) and head silhouette ($M = .411$) conditions ($p = .129$). Performance in the above two conditions was better than that in the front silhouette condition ($M = .217; ps < .001$). And performance in all three conditions was significantly better than that in the facial silhouette condition ($ps < .001$). The ANOVA also revealed a significant interaction between group and orientation, $F(3, 83) = 83.231, \text{MSE} = .854, p < .001$, $\eta^2 = .751$, group and silhouette type, $F(9, 249) = 35.792, \text{MSE} = .202, p < .001$, $\eta^2 = .564$, and orientation and silhouette type, $F(3, 249) = 53.889, \text{MSE} = .310, p < .001$, $\eta^2 = .394$. The three-way interaction was also significant, $F(9, 249) = 9.291, \text{MSE} = .053, p < .001$, $\eta^2 = .251$. To clarify the source of three-way interaction, separated ANOVAs were conducted for upright and inverted conditions.

![Figure 3](image-url) **Figure 3** Accuracy in different conditions. The asterisks denote the significance of the t test between the condition and the chance level.
Upright
The ANOVA for upright trials revealed main effects of group (same, neighbor, novel class, and adult), \( F(3, 83) = 117.743, \ MSE = 2.999, \ p < .001 \), \( \eta^2 = .810 \), and silhouette type, \( F(3, 249) = 570.247, \ MSE = 3.182, \ p < .001, \ \eta^2 = .873 \). The interaction between group \( \times \) silhouette type was also significant, \( F(9, 249) = 33.960, \ MSE = .190, \ p < .001, \ \eta^2 = .551 \). For the main effect of group, Bonferroni-corrected pair-wise comparisons revealed that performance was comparable between the same class group (\( M = .590 \)) and the neighbor class group (\( M = .539; \ p = .275 \)). However, the performance of these two groups was better than that of the adult group (\( M = .334; \ p < .001 \)).

Finally, the performances of these three groups were all better than that of the novel class group (\( M = .169; \ p < .001 \)). For the main effect of silhouette type, performance in the whole silhouette (\( M = .569 \)) and head silhouette conditions (\( M = .562 \)) were comparable (\( p = 1 \)), while both of them were significantly better than that in the front silhouette condition (\( M = .315; \ p < .001 \)). And all of them were significantly better than the performance in the facial silhouette condition (\( M = .169; \ p < .001 \)). For the class \( \times \) silhouette type interaction, analysis of simple main effects using Bonferroni-corrected revealed that for the whole silhouette and head silhouette conditions, performances between same class (\( M = .891/\text{whole silhouette}; \ M = .813/\text{head silhouette} \)) and neighbor class (\( M = .774/\text{whole silhouette}; \ M = .752/\text{head silhouette} \)) were comparable (\( p > .351 \)), while both of them were significantly better than that in the adult group (\( M = .472; \ p < .001 \)). And all of them were significantly better than the performance in the novel class (\( M = .21; \ p < .001 \)).

For the front silhouette condition, the performances between the same class (\( M = .497 \)) and neighbor class (\( M = .412 \)) were comparable (\( p = .058 \)), while both of them were significantly better than that in the adult group (\( M = .247; \ p < .001 \)). And all of them were significantly better than the performance in the novel class (\( M = .105; \ p < .001 \)). For the facial silhouette condition, the performance of the same class, neighbor class, and adult groups were comparable (\( p > .458 \)) and the performances in these three groups were all better than that in the novel class (\( p < .001 \)).

Inverted
The ANOVA for inverted trials revealed main effects of group (same, neighbor, novel class, and adult), \( F(3, 83) = 18.113, \ MSE = .336, \ p < .001, \ \eta^2 = .396 \), and silhouette type, \( F(3, 249) = 176.040, \ MSE = 1.025, \ p < .001, \ \eta^2 = .680 \). The interaction between group \( \times \) silhouette type was also significant, \( F(9, 249) = 11.356, \ MSE = .066, \ p < .001, \ \eta^2 = .291 \). For the main effect of class, Bonferroni-corrected pair-wise comparisons revealed that performance was comparable between the same class group (\( M = .236 \)), the neighbor class group (\( M = .198 \)), and the adult group (\( M = .208; \ p > .457 \)). The performances of these three groups were all better than that of the novel class group (\( M = .079; \ p < .001 \)).

For the main effect of silhouette type, performances in the whole silhouette (\( M = .287 \)) and the head silhouette conditions (\( M = .260 \)) were comparable (\( p > .097 \)), while both of them were significantly better than that in the front silhouette condition (\( M = .118; \ p < .001 \)). And all of them were significantly better than the performance in the facial silhouette condition (\( M = .056; \ p < .001 \)). For the class \( \times \) silhouette type interaction, analysis of simple main effects using Bonferroni-corrected revealed that for the whole silhouette and head silhouette conditions, performances of the same class, neighbor class, and adult groups were all better than that of the novel class (\( p < .001 \)). For the front silhouette condition, only the difference between the adult group and the novel group was significant (\( p < .001 \)). For the facial silhouette condition, the performances of the four groups were comparable (\( p = .05 \)).

Tests against Chance Performance
Two-tailed \( t \)-tests were conducted for the 32 conditions (4 groups \( \times 4 \) types \( \times 2 \) orientations) to examine whether the performance was better than chance level (0.04). Performance in all
conditions was higher than chance level \((p < .04)\) except for the recognition of the inverted facial silhouette for the same class group \((M = .052; p = .134)\) and the novel class group \((M = .03; p = .216)\), and the recognition of the upright facial part silhouette \((M = .0475; p = .549)\) for the novel class group, as shown in Figure 3.

**Discussion**

Research reveals that face configural processing develops much slower than feature or holistic processing (Mondloch et al., 2002). This study explored the developmental difference of face recognition in a different mechanism (i.e., viewpoint-transformation processing). It has been demonstrated that adults can readily match a front-view face to their corresponding silhouette or vice versa (Schwaninger & Yang, 2011). This processing requires complex calculation and representation of the 3-D information embedded in faces. The current study examined the maturity of this viewpoint-transformation processing.

Our results show that sixth-grade children (age range 11.09–12.09 years, \(M_{\text{age}} = 11.85\) years) were still ineffective in viewpoint-transformation processing when they were required to recognize unfamiliar faces from only the facial part. The performances of the two groups familiar with the faces were better than those of the adult group, who were unfamiliar with the faces; and the performances of all these three groups were better than that of the novel class group. Sixth-grade children from the same class and the neighbor class were excellent at recognizing the different types of silhouettes. Children with high familiarity with the faces performed even significantly better than adults. However, this high performance was limited only to the children who were familiar with the faces. The performance declined significantly for the children who were entirely unfamiliar with the faces. These results suggest that children familiar with the faces might match the silhouettes to the faces stored in their brain instead of rotating the front faces. Therefore, the performances of the children from the novel class were equal to chance level when they were required to recognize unfamiliar faces from only the facial part where no memorized silhouette faces could be provided. The results from the adults showed that the adults had developed an algorithm to mentally rotate the faces. However, this ability was not highly proficient, even in adults. As a result, performance in the adults group was still lower than that in children from the same class and neighbor class. Apparently, viewpoint-based experience for the children from the same class and neighbor class dominates the 3-D face transformation.

For students from either the same class or neighbor class group, the performance significantly declined when the shape of the head and the hairstyle cues were cropped out. However, although the decline is conspicuous, the performance was still much better than chance level when only the subtle facial curve was retained in silhouettes. These results suggest that with extensive experience in the recognition of familiar faces, participants might have stored the profile view in memory and might have developed a robust profile mental representation for familiar faces. However, this representation might be mainly based on the shape of the head and the hairstyle. The contribution from the sophisticated facial outline curve was relatively low. Moreover, this profile representation was largely constrained to the upright orientation of which faces were normally encoded.

For participants from the novel class, although the performance was higher than chance level in the recognition of the whole silhouette, head silhouette, and front silhouette, it was at chance level in the recognition of the facial silhouette. Performance in the recognition of the facial silhouette for the novel class provides a crucial evaluation to assess whether participants can apply viewpoint-transformation processing. For participants from the novel class, the profile views of the target faces were not available for them. Meanwhile, viewpoint-based calculations, such as linear interpolation
between views (Poggio & Edelman, 1990), multiple views plus transformations (Tarr & Pinker, 1989), or linear combination of views (Ullman & Basri, 1991), were not applicable because only one singular view was provided. However, the performance in the recognition of the facial silhouette for adults was higher than chance level and was comparable to the performance of students from the same class and neighbor class. This implies that adults can apply the 3-D transformation and reach the performance of the students with almost robust view-based representations.

In the experiment of Davidenko (2007), the face silhouettes that participants were required to match were manipulated similarly to the facial silhouette adopted in the current experiment. In addition, the face stimuli were novel to the participants. This condition is identical to the condition in the recognition of the facial silhouette for the adult and novel class group participants in the current study. The adult participants exhibited high proficiency in the viewpoint-transformation task in the study of Davidenko (2007) as well as in the current study. By contrast, novel class group children were completely incapable of solving this task in the current study. This suggests that participants from the novel class simply relied on non-facial external information to recognize the faces. They were unable to process the 3-D information embedded within the faces. As discussed previously, participants might need the deployment of multiple mechanisms to accomplish this task. It includes recovering the 3-D features from shading and shadow information (Bruce & Langton, 1994), calculating the distances between face features, precision of viewpoint transformation, and so forth. This competence apparently takes years to reach proficiency.

In sum, the results of this study indicate that in addition to configural processing, the viewpoint-transformation skill in face recognition also develops substantially slowly. It is still far from proficient in early adolescence. The results of this research provide notable insight into the inspection of developmental difference in face recognition. Apparently, face recognition involves divergent and complicated mechanisms. The slow development in face recognition might be due not only to the deficit in configural processing but also to the immaturity in other associated cognitive skills, such as the capacity of transforming the face view from a different viewpoint.

References


(Received May 4, 2017; accepted January 25, 2018)