Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
J. Experimental Child Psychology 89 (2004) 159–181
Journal of Experimental Child Psychology
www.elsevier.com/locate/jecp

Why 8-year-olds cannot tell the difference between Steve Martin and Paul Newman: Factors contributing to the slow development of sensitivity to the spacing of facial features

Catherine J. Mondloch[a,b,*], Kate S. Dobson[b], Julie Parsons[b], Daphne Maurer[b]

[a] *Department of Psychology, Brock University, St. Catharines, Ont., Canada L2S 3A1*
[b] *Department of Psychology, McMaster University, Hamilton, Ont., Canada L8N 1E9*

Received 12 February 2004; revised 17 June 2004
Available online 23 August 2004

Abstract

Children are nearly as sensitive as adults to some cues to facial identity (e.g., differences in the shape of internal features and the external contour), but children are much less sensitive to small differences in the spacing of facial features. To identify factors that contribute to this pattern, we compared 8-year-olds' sensitivity to spacing cues with that of adults under a variety of conditions. In the first two experiments, participants made same/different judgments about faces differing only in the spacing of facial features, with the variations being kept within natural limits. To measure the effect of attention, we reduced the salience of featural information by blurring faces and occluding features (Experiment 1). To measure the role of encoding speed and memory limitations, we presented pairs of faces simultaneously and for an unlimited time (Experiment 2). To determine whether participants' sensitivity would increase when spacing distortions were so extreme as to make the faces grotesque, we manipulated the spacing of features beyond normal limits and asked participants to rate each face on a "bizarreness" scale (Experiment 3). The results from the three experiments indicate that low salience, poor encoding efficiency, and limited memory can partially account for 8-year-olds' poor performance on

* Corresponding author. Fax: +1 905 688 6922.
E-mail address: cmondloch@brocku.ca (C.J. Mondloch).

doi:10.1016/j.jecp.2004.07.002

face processing tasks that require sensitivity to the spacing of features, a kind of configural processing that underlies adults' expertise. However, even when the task is modified to compensate for these problems, children remain less sensitive than adults to the spacing of features.

Keywords: Face processing; Face recognition; Expertise

Introduction

Despite abundant exposure to faces during infancy and childhood, children do not reach adult levels of expertise in recognizing facial identity until adolescence (Carey, Diamond, & Woods, 1980), even in matching tasks that eliminate memory demands (Bruce et al., 2000). Because all faces share the same first-order relations among features (e.g., two eyes above a nose, a mouth below the nose), recognizing facial identity requires processing the shape of individual features (e.g., eyes, mouth, chin) and/or processing the spacing among features (e.g., distance between the eyes). Identification based on these cues involves featural and second-order relational processing, respectively (for a review, see Maurer, Le Grand, & Mondloch, 2002).

Two lines of research indicate that adults are able to use both featural and second-order relational cues to identify individual faces. In the first approach, adults' ability to recognize faces from a study set is measured under conditions in which featural information is reduced by blurring the face or second-order relational information is reduced by scrambling the features and/or by inverting the face so as to make it harder to process second-order relations (Freire, Lee, & Symons, 2000; Murray, Yong, & Rhodes, 2000). Adults are able to recognize faces so long as one type of information is available: when faces are blurred, inverted, scrambled, or inverted and scrambled, adults' performance remains well above chance. However, when faces are both blurred (reducing featural cues) and inverted/scrambled (reducing spacing cues to second-order relations), performance drops to chance, presumably because neither featural nor second-order relational processing can be used effectively (Collishaw & Hole, 2000).

In the second approach, sets of faces have been created that differ only in the shape of individual features or only in the spacing among features (Freire et al., 2000; Mondloch, Le Grand, & Maurer, 2002). Mondloch et al. (2002) modified a single female face (called "Jane") to create 12 new versions (called "sisters")—4 that differed in the spacing of internal features (spacing set), 4 that differed in the shape of internal features (featural set), and 4 that differed in the shape of the external contour (contour set). In separate blocks, stimuli were presented either upright or inverted to verify that the spacing set taps second-order relational processing, which should be disrupted by inversion, and that the featural set taps featural processing, which should be less disrupted by inversion (Diamond & Carey, 1986; Freire et al., 2000; Rhodes, Brake, & Atkinson, 1993). Pairs of faces were presented sequentially, and participants indicated whether the two faces were the same or different. Adults' accuracy was high (>80%) on all face sets. The decrement in performance following inversion was much higher for the spacing set (18%) than for the featural and contour sets (8%), confirming that the spacing set taps second-order relational processing. The high accuracy in all conditions also indicated that adults can judge identity accurately based solely on featural cues or solely on second-order relational cues—at least when faces are upright (see also Leder & Bruce, 2000; Rhodes et al., 1993).

Children's poor performance on face recognition tasks, relative to that of adults, has been attributed to immature second-order relational processing. Although 5-month-olds can detect exaggerated changes in spacing in a schematic face (Bhatt, Bertin, & Hayden, 2004) and 6-year-olds are above chance on the spacing set of the Jane task, even 14-year-olds make more errors than do adults (Mondloch et al., 2002; Mondloch, Le Grand, & Maurer, 2003). In contrast, by 6 years of age (the youngest age tested), children are nearly as accurate as adults on the external contour and featural sets in the Jane task, with no significant difference for the former and only a small difference in means for the latter. Furthermore, unlike older children and adults, the magnitude of the inversion effect is not greater for the spacing set than for the featural and contour sets in 6- and 8-year-olds. The developmental pattern does not correspond to variations across sets in their difficulty for adults. Although adults' accuracy is higher on the featural set ($M = 89\%$ correct) than on the spacing set ($M = 82\%$ correct), even 6-year-olds' performance is adultlike on the contour set—the set on which adults' performance is lowest ($M = 80\%$ correct). Using a similar paradigm, Freire and Lee (2001) also found that 4- to 7-year-olds perform poorly—although better than chance—when asked to detect a target face that is presented among distracters that differ only in the spacing of features. Their accuracy is higher when the distracters differ from the target in the shape of individual features.

The spacing variations in the Jane set cover most of the natural variation among adults' faces; hence, children's poor performance likely affects their processing of faces outside the laboratory and contributes to their documented errors on other tasks. For example, 6- and 8-year-olds have difficulty in matching faces when they differ in point of view, clothing, or lighting (Benton & Van Allen, 1973, as cited in Carey et al., 1980; Bruce et al., 2000; Mondloch, Geldart, Maurer, & Le Grand, 2003). Such changes render second-order relational processing more important because faces no longer can be matched accurately based on the appearance of individual features (e.g., Mondloch, Geldart et al., 2003).

The experiments reported here were designed to test three hypotheses as to why children perform poorly on tasks that require second-order relational processing. In each of these experiments, we compared the performance of 8-year-olds with that of adults. In Experiment 1, we tested the hypothesis that 8-year-olds fail to attend to the spacing among features because they are distracted by the shape of individual internal features (e.g., the eyes) and the external contour. To test this hypothesis, we reduced the salience of featural cues by blurring faces in the spacing set of the Jane task and occluding the features (i.e., eyes, nose, and mouth). In Experiment 2, we tested the hypothesis that 8-year-olds can encode second-order relations but performed worse than adults on the sequential version of the Jane task either because the children were unable to encode spacing information rapidly or because they failed to adequately store and retrieve information about second-order relations.

To test this hypothesis, we presented stimuli in pairs for an unlimited amount of time to provide children with an adequate amount of time to encode second-order relations and to eliminate memory demands. Because neither of these manipulations resulted in adultlike performance, in Experiment 3 we asked whether 8-year-olds would demonstrate adultlike sensitivity to second-order relations if the manipulations were so extreme as to make the faces appear bizarre to adults. To facilitate comparison among the hypotheses, all of the experiments were conducted at 8 years of age, an age when performance on tasks requiring sensitivity to second-order relations is above chance but much poorer than that of adults (Mondloch et al., 2002).

Experiment 1

Children differ from adults in the facial characteristics to which they attend when engaged in face processing. Unlike adults, young children rely more on external features (e.g., hair) than on internal features when recognizing familiar faces (Campbell & Tuck, 1995; Campbell, Walker, & Baron-Cohen, 1995; Ellis, Shepherd, & Davies, 1979), and they are influenced more by paraphernalia (e.g., glasses, hats)—at least when faces are highly similar (Baenninger, 1994; Carey & Diamond, 1977). When asked to categorize ambiguous drawings based on age (child/adult), 7-year-olds do so based on one individual feature (e.g., eyes), whereas adults do so based on overall similarity (e.g., the number of features that could be assigned to either category) (Schwarzer, 2000). Collectively, these results suggest that 8-year-olds' ability to make same/different judgments among faces that differ in the spacing among features might be masked because the children fail to attend to relevant cues. Instead, attention is focused on individual internal features (e.g., eyes, mouth) and the facial contour. This hypothesis is consistent with evidence that children are less able to ignore distracters than are adults (e.g., Goldberg, Maurer, & Lewis, 2001; Ridderinkhof & van Molen, 1995).

To test this hypothesis, in Experiment 1b we altered the faces in the spacing set of the Jane task in two ways so as to make the features and external contour less salient. The first manipulation was to blur the faces so as to reduce featural information (Collishaw & Hole, 2000). The second manipulation was to both blur the faces and occlude the internal features with white ovals (Baenninger, 1994). After these two manipulations, children were given the original task to determine whether testing with faces in which the internal features and external contour were less salient allows 8-year-olds to more accurately discriminate among the original faces that differed in the spacing among features.

Prior to testing children, we validated our stimuli by testing adults in Experiment 1a. Because the goal of testing adults was to validate our stimuli rather than to provide a comparison group for children, adults were tested under different conditions than were children. We tested adults on two versions of the featural set: the original set and blurred versions of those faces. Lower accuracy when the stimuli were blurred would confirm that our blurring manipulation reduced featural cues. We also presented the blurred and occluded spacing sets both upright and inverted. High accuracy in the upright condition and much lower accuracy in the inverted condition would confirm that adults were able to make same/different judgments using second-order cues with the altered upright stimuli.

Experiment 1a

Method

Participants

The 24 Caucasian adults (12 men and 12 women, 18–28 years of age) all were right-handed, as determined by a handedness test adapted from Peters (1988). All were undergraduate students participating for credit in a psychology course at McMaster University. None of the participants had a history of eye problems, and all met our criteria on a visual screening exam: at least 20/20 Snellen acuity in each eye without optical correction, worse acuity with a +3 diopter lens (to rule out farsightedness greater than 3 diopters), fusion at near on the Worth Four dot test, and stereoacuity of at least 40 arc s on the Titmus test. An additional 6 participants were excluded from the final analysis because they failed visual screening ($n = 5$) or were not right-handed ($n = 1$).

Stimuli and apparatus

The stimuli were modified versions of grayscale Caucasian female faces used previously to study the development of face recognition (Mondloch et al., 2002). We modified a single face (called Jane) to create new versions (called Jane's sisters). All stimuli were 10.2 cm wide and 15.2 cm high (5.7×9.1 visual degrees from the testing distance of 100 cm). The four faces in the spacing set (Fig. 1A) were created by moving the eyes 12.5 pixels (4.4 mm) up or down from the original, 9 pixels (3.2 mm) closer together, or 12 pixels (4.2 mm) farther apart. The mouth was moved 4 pixels (1.4 mm) up or down. According to anthropomorphic norms (Farkas, 1994), we moved Jane's eyes up or down 1.3 standard deviations (*SD*s), closer together 2.4 *SD*s, and farther apart 3.2 *SD*s; we moved her mouth up or down 0.79 *SD*s. These spacing changes cover most of the variations in spacing among adult Caucasian female faces in the population. The four faces in the featural set were created by replacing the model's eyes and mouth with the features of different females. We chose features of the same length to minimize resulting changes in the spacing among features (Fig. 1B). We used modified versions of these stimuli in the current study.

Two new versions of the spacing set were presented. The blurred set was created using the Gaussian filter in Adobe Photoshop; as in Collishaw and Hole (2000), the filter had a radius of 10 pixels. We first blurred the eye region and then blurred the entire face (Fig. 1D). The occluded set was created by placing white circular shapes over the eyes (1.90×2.29 cm), nose (1.90×1.90 cm), and mouth (1.27×2.54 cm) (Fig. 1E) of the blurred faces. The same occluders were used for all faces. The inner point of the eye occluder was aligned with the inner edge of each eye, the mouth occluder was aligned with the central top of the mouth, and the nose occluder was positioned

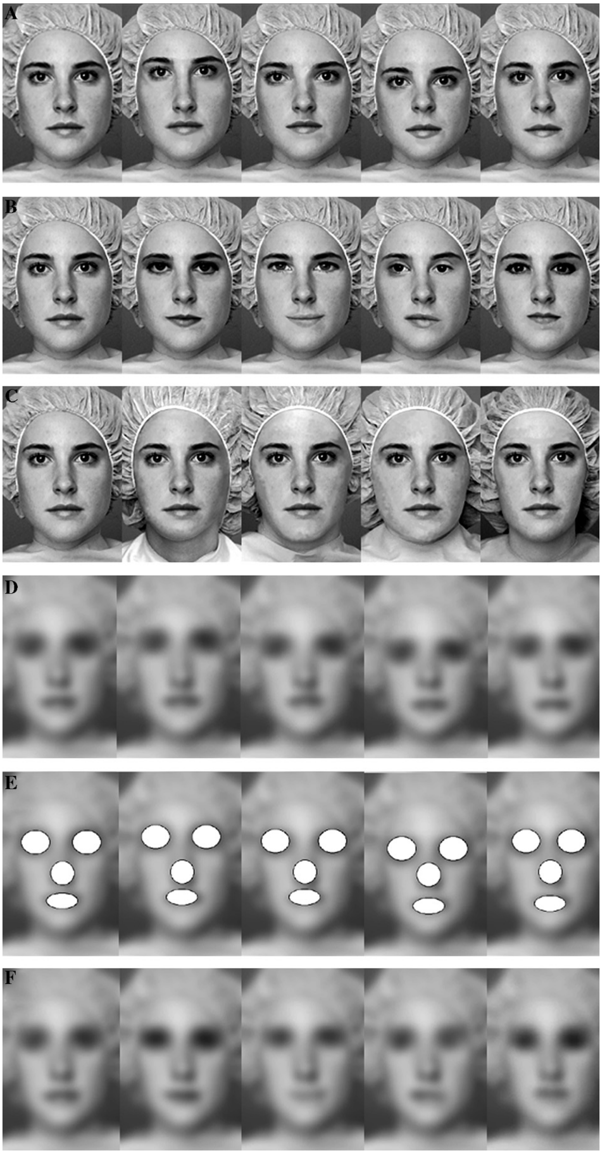

Fig. 1. "Jane" shown as the left-most face in each panel along with her sisters from the spacing set (A), the featural set (B), the contour set (C), the blurred spacing set (D), the occluded spacing set (E), and the blurred featural set (F).

over the nose, which was in the same position for all faces. Each version of the spacing set was presented both upright and inverted.

To measure whether this amount of blurring successfully reduces featural information, we presented two versions of the featural set: an unaltered version and a blurred version created using the same Gaussian filter described above (Fig. 1F). Because our ultimate goal was to use blurring to reduce featural information in the spacing set, which consisted of five faces with identical eyes, we included only the three sisters from the featural set with similarly light eyes (the first, third, and fourth columns in Fig. 1F). The stimuli were presented on a monochrome Radius 21-GS monitor controlled by a Macintosh LC-475 computer and Cedrus Superlab software. Participants signaled their responses using a joystick, and the experimenter initiated trials by pressing a key on the keyboard.

Procedure

Informed consent was obtained. Each participant sat in a darkened room with his or her eyes 100 cm from the monitor. The procedure began with a practice task requiring same/different judgments of two identical faces or two radically different versions of a face (e.g., a face with eyes rotated 45 degrees clockwise vs. counterclockwise). The first three pairs of faces were presented side by side, and the participant received feedback. There were then 12 trials in which faces were presented sequentially. The first face appeared for 360 ms, and the second face remained on the screen until the participant responded. To participate in the main experiment, participants were required to be correct on at least 10 of these 12 trials. All adults met this criterion on the first attempt.

Following practice, adults received four blocks of trials—blurred spacing set upright, blurred spacing set inverted, the unaltered upright featural set, and the blurred upright featural set—with the order counterbalanced across the 24 adults. Half (12) of the participants were then tested on the occluded spacing set upright, followed by the occluded spacing set inverted. During each trial, the first (model) face appeared for 200 ms, and following an interstimulus interval of 300 ms, the second (test) face appeared until the participant responded. Within each block, the correct response was "same" for half of the trials; each face served as the test face as often as the model face, and each face was presented half of the time on a "same" trial and half of the time on a "different" trial. All participants saw the same random order of trials within each block. Four practice trials were given prior to each block.

Results and discussion

Blurred featural set

As shown in Fig. 2, blurring the faces successfully reduced the salience of featural information. Accuracy was lower on the blurred featural set ($M = .78$, $SD = .12$) than on the unaltered featural set ($M = .92$, $SD = .06$), $t(23) = 6.9$, $p < .0001$.

Blurred spacing set

As shown in Fig. 2, adults were able to discriminate faces in the blurred spacing set and seemed to do so on the basis of second-order relations. They performed well when the stimuli were upright ($M = .76$ correct, $SD = .09$), and inversion severely impaired their performance ($M = .58$ correct, $SD = .10$), $t(23) = 0.18$, $p < .0001$.

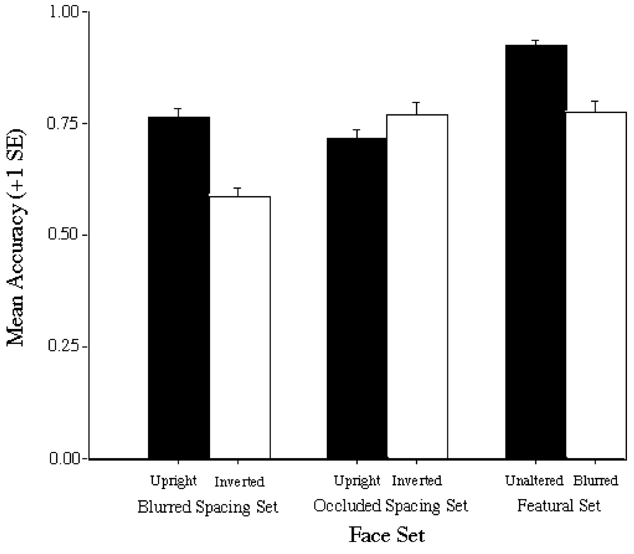

Fig. 2. Mean adults' accuracy (+1 *SE*) for each face set. Black bars indicate which conditions we predicted would be easier because the stimuli were either upright or unaltered.

Occluded spacing set

Adults' accuracy on the occluded spacing set did not vary across orientations (upright: $M = .72$, $SD = .07$; inverted: $M = .77$, $SD = .10$). Adults may have processed the area between well-demarcated occluders as features (e.g., broad low vs. skinny high bridge of the nose) rather than processing the spacing among occluders per se (see also Murray et al., 2000). Nonetheless, we included the occluded set in Experiment 1b because Baenninger (1994) reported that 8-year-olds are able to match facial identity when features are occluded. Based on these results, Baenninger concluded that "configural information provided in faces is very important for subjects as young as 8 years." If our 8-year-old participants had performed well on the occluded set, we would have had to test a second group of children on inverted versions of these same stimuli so as to verify that they were relying on second-order relations and not, like adults, discriminating the stimuli based on featural processing.

Experiment 1b

Method

Participants

A total of 20 Caucasian 8-year-olds participated. All participants passed the visual screening test described for adults and were right-handed, as measured by a modified version of the handedness test used for adults. An additional 12 children were tested but were excluded because they failed visual screening.

Stimuli

Three versions of the spacing set were used—the standard set, the blurred set, and the occluded set—along with control stimuli, called "Jane's cousins," consisting of Jane and three different females (faces that differed from the original Jane in features, spacing, and outer contour). The purpose of the control task was to ensure that children still were motivated and attentive at the end of the procedure and that they understood the experimental task. All children met our criterion of 70% correct on the control task (see also Mondloch et al., 2002).

Procedure

We obtained informed consent from each parent and assent from each child. Each child then completed the practice task (all children met the inclusion criterion) and then was tested on three blocks of 30 trials: the blurred spacing set, the occluded spacing set, and the original spacing set (in that order). Finally, the child was tested on the control stimuli (32 trials). Visual screening and the handedness test were inserted during testing to provide the child with breaks.

Results and discussion

The 8-year-olds' accuracy was not higher when stimuli were blurred ($M = .61$, $SD = .12$), or when features were occluded ($M = .61$, $SD = .12$), than it was for the standard spacing set in our previous study ($M = .66$, $SD = .13$) (Mondloch et al., 2002) or in the current study ($M = .69$, $SD = .12$). Furthermore, their performance on the standard spacing set was not higher in the current study than it was in our previous study, despite the children's having been exposed to 60 trials in which pairs of stimuli differed only in the spacing among features prior to this test block. Performance was above chance in all conditions (blurred set: $t(19) = 4.3$, $p < .001$; occluded set: $t(19) = 3.8$, $p < .001$; standard set: $t(19) = 6.9$, $p < .0001$).

The results of this study fail to support the hypothesis that children perform poorly when asked to discriminate pairs of faces that differ only in second-order relations because they are distracted by the facial contour and the shape of individual features. Blurring the stimuli reduced featural information while leaving spacing cues to second-order relations (Experiment 1a) but did not improve 8-year-olds' accuracy.

Experiment 2

In Experiment 2, we measured the influence of encoding speed and memory demands on 8-year-olds' relatively poor performance on the spacing set in our previous study (Mondloch et al., 2002). In the original study, the first face appeared for only 200 ms, a time that is sufficient for 8-year-olds to encode contour and featural information as well, or nearly as well, as do adults. Nevertheless, it might not be enough time for 8-year-olds to encode second-order cues. Moreover, children had to remember the first face during the 300 ms between its disappearance and the presentation of the second face. Although that short delay did not prevent nearly adultlike performance on the featural and contour sets, it is possible that children perceive spacing cues but then fail to store and/or retrieve them readily. Previous research has shown that adults' accuracy on face recognition tasks is limited by their ability to encode spacing cues and, possibly, by their ability to retrieve that information (Freire et al., 2000). In a delayed matching-to-sample task in which adults were asked to make same/different judgments about face pairs that differed only in the spacing of features, there was no effect of delay (1, 5, and 10 s), a result that Freire et al. (2000) interpreted as evidence that an encoding bottleneck limits performance. However, adults were more accurate when pairs were presented simultaneously for an unlimited time (82%) than when there was a delay between the two faces (74%). Adults may have performed better in the simultaneous matching task because they had more time to encode second-order relations and/or because there were no memory demands.

To assess whether there are similar influences on the performance of 8-year-olds, we tested adults and 8-year-olds with the original Jane task altered so that faces were presented in pairs and for an unlimited time. As in our previous study (Mondloch et al., 2002), there were three sets of faces—the spacing set, the featural set, and the contour set—presented both upright and inverted. This allowed us to assess whether reducing memory and encoding demands would eliminate 8-year-olds' poorer performance on the spacing set while maintaining their adultlike accuracy on the featural and contour sets.

Method

Participants

A total of 36 adults and 36 8-year-olds participated in this study. All participants were Caucasian, were right-handed, passed our criterion for visual screening (see Experiment 1a), and achieved accuracy of at least 70% on the control stimuli (Jane's cousins). An additional 15 individuals were excluded from the analyses because they were not right-handed (1 adult), failed visual screening (10 children and 2 adults), or were inattentive throughout testing (2 children).

Stimuli, apparatus, and procedure

The stimuli were the original featural set, the original spacing set, and the control stimuli described in Experiment 1a plus a fourth set—the contour set that had been created by pasting the internal portion of the original face within the outer contour of four different females (Fig. 1C). Faces were paired instead of being presented sequentially (inner edges were 2.5 cm apart (1.4 visual degrees)). Except as noted below, other details were the same as in Experiments 1a and 1b.

During each trial, two faces were presented side by side until participants signaled a response with a joystick. Participants were asked to respond as quickly as possible while being accurate. The procedure began with the same practice task used in Experiment 1 except that the stimuli were presented side by side. The first part of the main task involved 90 upright trials that were divided into three 30-trial blocks—spacing, featural, and external contour—preceded by six practice trials involving one same and one different trial from each stimulus set. Trials were blocked to encourage participants to use specific face processing strategies. Following the 90 upright trials, participants repeated the blocks in the same order with the faces inverted, preceded again by six practice trials. Three orders were used: spacing → featural → contour, featural → contour → spacing, and contour → spacing → featural (for additional details, see Mondloch et al., 2002). Following the final block of inverted trials, we presented a block of upright trials with Jane and her cousins. This block consisted of 32 trials with either the same face twice (16 trials) or two completely different faces (16 trials).

Data analysis

For each of the six blocks of trials for each participant in the main experiment, we calculated the proportion of responses that were correct and the median reaction time on correct trials. Unlike Experiment 1, we included an analysis of reaction time to determine whether differences in accuracy across face sets within an age group or a Face Set × Age interaction could be attributed to speed/accuracy trade-offs.

Accuracy

We conducted an analysis of variance (ANOVA) on proportion correct responses on upright trials that had one within-subjects factor (face set) and one between-subjects factor (age). Differences in the rate of development for different face processing skills would be revealed by a significant Age × Face Set interaction on upright trials. To determine whether the pattern of results could be due to children having response biases different from those of adults, we also conducted a d' analysis for upright trials.

To assess the effect of orientation on accuracy across face sets and age groups, we calculated a difference score (accuracy on upright trials minus accuracy on inverted trials) for each face set for each age group. If adults' larger inversion effect on the spacing set, relative to the featural and contour sets, is due to an encoding bottleneck, adults should show a larger inversion effect on the spacing set even when faces are presented simultaneously. If presenting faces simultaneously increases 8-year-olds' sensitivity to second-order relations, 8-year-olds [like 10-year-olds and adults in our previous study (Mondloch et al., 2002)] should show a larger inversion effect for the spacing set relative to the other two sets.

Reaction times

To determine whether differences in accuracy among face sets could be attributed to speed/accuracy trade-offs, we also conducted a similar ANOVA on median reaction time on correct trials. Fisher's protected least significant difference (PLSD) was used to make pairwise comparisons among face sets. To assess the effect of orientation on reaction times, we calculated a second ANOVA on difference scores (median reaction time on upright trials minus median reaction time on inverted trials) for each face set for each group.

Results

Accuracy on upright trials

As shown in Fig. 3, both age groups made very few errors on Jane's cousins, although adults ($M = .98$, $SD = .04$) were more accurate than 8-year-olds ($M = .94$, $SD = .06$), $t(70) = 4.1$, $p < .001$. The ANOVA on proportion correct for the three experimental face sets revealed main effects of age, $F(1,70) = 6.1$, $p < .02$, and face set, $F(2,140) = 76.8$, $p < .001$, and a significant Age × Face Set interaction, $F(2,140) = 3.74$, $p < .05$. Analyses of simple effects showed a significant effect of age for the spacing set, $t(70) = 3.2$, $p < .001$, but not for the featural and contour sets, $ps > .20$. As shown in Fig. 3, adults ($M = .83$, $SD = .10$) were more accurate than 8-year-olds ($M = .73$, $SD = .14$) on the spacing set.

d′ analysis

The ANOVA on d' revealed a main effect of age, $F(1,70) = 6.1$, $p < .02$, a main effect of face set, $F(2,140) = 98.8$, $p < .001$, and a significant Age × Face Set interaction, $F(2,140) = 3.5$, $p < .05$. As seen for accuracy, analyses of simple effects showed a significant effect of age for the spacing set, $t(70) = 3.4$, $p < .01$, but not for the featural and contour sets, $ps > .15$. Adults ($M = 2.3$, $SD = .8$) had a higher d' than did 8-year-olds ($M = 1.6$, $SD = 1.0$) only on the spacing set.

Reaction times on upright trials

The ANOVA of median reaction times on upright trials revealed a main effect of face set, $F(2,140) = 26.2$, $p < .001$. As shown in Fig. 4, reaction times were shorter on featural trials than on both spacing and contour trials, $ps < .001$, and were shorter on

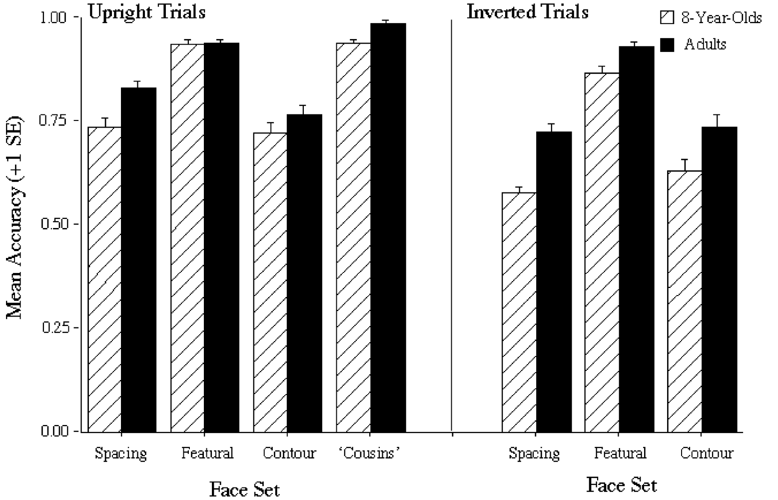

Fig. 3. Mean accuracy (+1 *SE*) for each face set and each age group when stimuli were presented upright (left panel) and inverted (right panel).

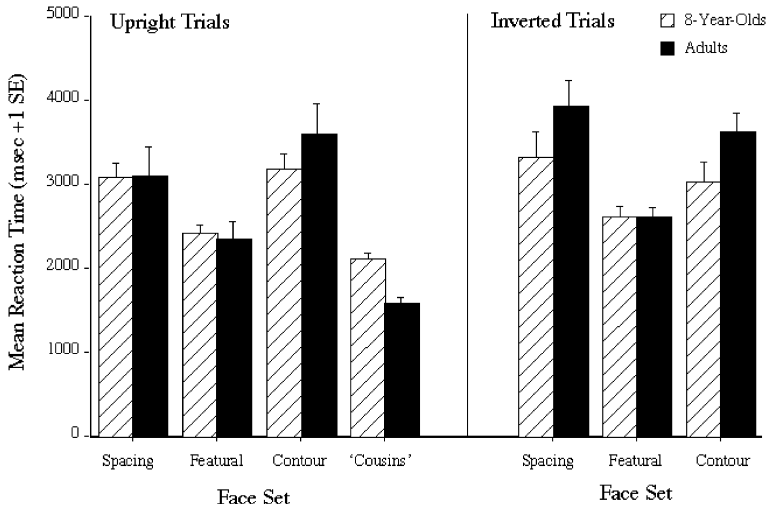

Fig. 4. Mean reaction times (+1 SE) on correct trials for each face set and each age group when stimuli were presented upright (left panel) and inverted (right panel).

spacing trials than on contour trials, $p < .05$. There was no main effect of age and no significant Age × Face Set interaction, ps > .20. These data show that 8-year-olds' poor performance relative to that of adults only on the spacing set cannot be attributed to a speed/accuracy trade-off.

Inversion effect

Inversion reduced the accuracy of both groups on the spacing set more than on the featural and contour sets (Fig. 5). The ANOVA on inversion scores revealed a main effect of age, $F(1,70) = 10.8$, $p < .01$, and a main effect of face set,

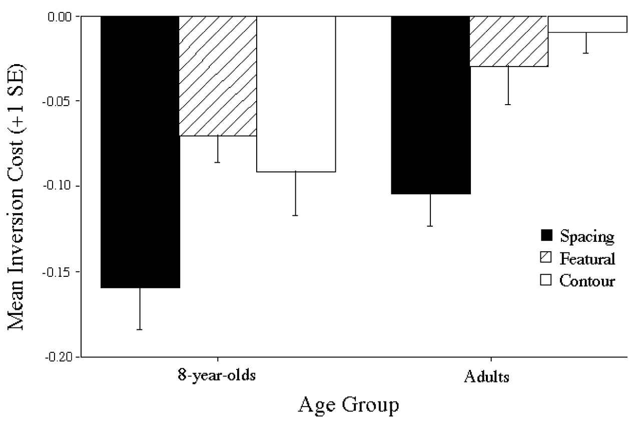

Fig. 5. Mean difference scores (accuracy on inverted trials minus accuracy on upright trials +1 *SE*) for each face set and each age group.

$F(1,140) = 11.8$, $p < .01$, but no Age × Face Set interaction, $p > .20$. The inversion effect was greater for the spacing set ($M = .13$, $SD = .13$) than for either the featural set ($M = .04$, $SD = .09$) or the contour set ($M = .06$, $SD = .15$), *p*s < .001, which did not differ from each other, $p > .20$. Inversion scores were larger for 8-year-olds than for adults.

Inversion increased adults' reaction times on the spacing set but had little or no effect on their reaction times in the other conditions or on 8-year-olds' reaction times in any condition (Fig. 6). The ANOVA on inversion effects for reaction times revealed a main effect of face set, $F(2,140) = 6.09$, $p < .01$, but no effect of age and no Age × Face Set interaction, *p*s > .10. Inverting the stimuli increased reaction times on the spacing set more than on the contour set, $p < .01$; no other pairwise comparisons were significant. Fig. 6 indicates that the effect of inversion on reaction times for the spacing set was carried by the adults. Because reaction times increased (adults) or remained unchanged (8-year-olds) when the spacing set was inverted, reduced accuracy when the spacing set was inverted cannot be attributed to speed/accuracy trade-offs.

Direct comparison with sequential data

The 8-year-olds made more errors than did adults on the spacing set when faces were presented in pairs for an unlimited period of time. Nonetheless, their mean accuracy ($M = .73$, $SD = .14$) was higher than that found in our previous study where stimuli were presented sequentially ($M = .65$, $SD = .13$), $t(70) = 2.40$, $p < .01$. In comparison, adults' accuracy on the spacing set did not vary across studies (*M*s = .82).

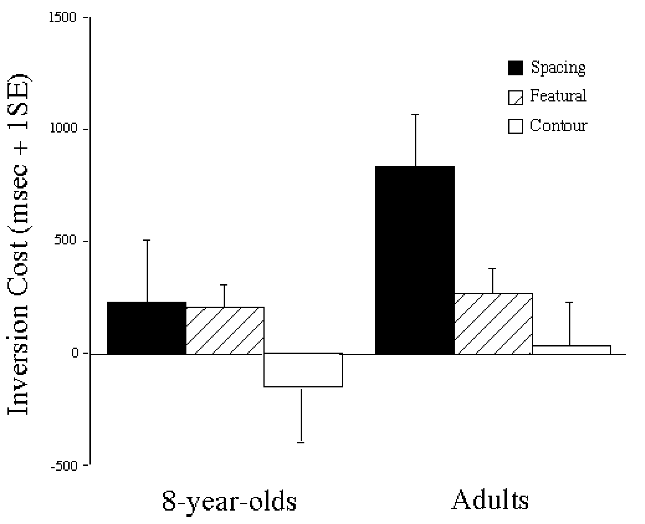

Fig. 6. Mean difference scores (reaction time on inverted trials minus reaction time on upright trials +1 *SE*) for each face set and each age group.

Discussion

In this study, 8-year-olds were more accurate when asked to discriminate faces that differed only in the spacing among features when stimuli were presented simultaneously for an unlimited time than they were in our previous study where the stimuli were presented sequentially (Mondloch et al., 2002). Furthermore, 8-year-olds (like adults) showed a larger inversion effect on the spacing set than on the featural and contour sets. This is the pattern expected when faces in the spacing set are discriminated with second-order relational processing and was the pattern shown by children 10 years of age or over, but not 8-year-olds, in our previous sequential study (Mondloch et al., 2002).

Two factors may have contributed to 8-year-olds' improved second-order relational processing in Experiment 2 compared with the previous sequential study. First, memory demands were eliminated in Experiment 2. Second, the children were allowed more time to encode second-order relations in the two faces. Although unlimited presentation time also may have allowed participants to engage in direct comparison between the stimuli, and thus rely on pictorial cues or feature-by-feature comparison, it is unlikely that this explains 8-year-olds' improved performance given that the children improved for the upright spacing set (73% correct vs. 65% correct in the previous sequential version) but not for the inverted spacing set (57% in both studies). One hallmark of relying on second-order relations when discriminating upright stimuli is a large inversion effect (Freire et al., 2000; Mondloch et al., 2002).

The comparison across studies suggests that 8-year-olds are unable to encode spatial relations as efficiently as are adults. Unlike adults, 8-year-olds were more accurate on the spacing set when stimuli were presented for an unlimited amount of time in Experiment 2 than when stimuli were presented very briefly (200 ms) in our previous study (Mondloch et al., 2002). Thus, one cause of their poor performance on tasks requiring the processing of second-order relations may be encoding inefficiency (or inability to remember the encoded relations across a short delay). Despite their improved performance, 8-year-olds made more errors than did adults on the spacing set, in contrast to their adultlike accuracy on the featural and contour sets. Their immaturity on the spacing task cannot be attributed to the spacing set being more difficult than the other sets. Adults' accuracy on the spacing set ($M = .83$, $SD = .10$), although lower than their accuracy on the featural set ($M = .94$, $SD = .06$), was actually higher than their accuracy on the contour set ($M = .76$, $SD = .14$). The 8-year-olds' immaturity also cannot be attributed to speed/accuracy trade-offs given that 8-year-olds did not respond faster on the spacing set than on the featural and contour sets. Rather, there seems to be an immaturity in processing second-order relations during childhood that cannot be eliminated by making featural cues less prominent (Experiment 1) and that can be only partially reduced by presenting faces simultaneously without a memory demand (Experiment 2).

The 8-year-olds were less accurate than adults on Jane's cousins, a task in which "different" trials involved presenting two completely different faces, despite being as accurate as adults on both the featural and contour sets and despite the fact that the faces were presented simultaneously for an unlimited amount of time. Adults' superior performance indicates the benefit of their greater sensitivity to second-order relations even when featural and contour cues are available, a benefit that is likely to increase when a face is seen from a novel point of view (Mondloch, Geldart et al., 2003).

Experiment 3

Like our previous study (Mondloch et al., 2002), Experiments 1b and 2 investigated 8-year-olds' sensitivity to second-order relations in recognizing facial identity. In Experiment 3, we asked whether 8-year-olds would demonstrate adultlike sensitivity to second-order relations if the manipulations were so extreme as to make the faces appear bizarre to adults. We made three types of alteration to each of four faces: featural distortions created by whitening the eyes and blackening portions of the teeth, "Thatcherization" created by inverting the eyes and mouth in an otherwise upright face, and spatial distortions created by altering the spacing among the eyes, nose, and mouth beyond natural limits. Each of these manipulations causes adults to rate faces as highly bizarre when faces are upright (Lewis, 2001; Murray et al., 2000; Searcy & Bartlett, 1996), and the effect of inversion on their ratings of bizarreness varies across distortions. Ratings of the bizarreness of spatially distorted and Thatcherized faces decrease nonlinearly with inversion (Lewis, 2001; Murray et al., 2000; Searcy & Bartlett, 1996), whereas ratings of featurally distorted faces do not change (Searcy & Bartlett, 1996) or actually increase slightly (Murray et al., 2000).

Even 6-month-olds discriminate between a normal face and a face that has been Thatcherized, at least if the faces are upright (Bertin & Bhatt, in press), and by 6 years of age children perform like adults when asked to rotate a Thatcherized face until it changes from looking bizarre to looking normal (Lewis, 2003). They also pick the Thatcherized version of a face as odd or unusual compared with the normal version with which it is paired, and if the presentation time is unlimited, their accuracy approaches a ceiling of 100% (Donnelly & Hadwin, 2003). Because of the ceiling effect, these data do not indicate whether children perceive Thatcherized faces to be as grotesque as do adults, nor do the data indicate whether children perceive Thatcherized faces to be as bizarre as featural distortions. In the current study, we asked 8-year-olds and adults to rate faces on a 7-point "bizarreness" scale. The 8-year-olds' sensitivity to second-order relations could be demonstrated in two ways: adultlike ratings of Thatcherized and spatially distorted faces and greater inversion effects for Thatcherized and spatially distorted faces relative to featurally distorted and unaltered faces.

Method

Participants

A total of 20 8-year-olds and 20 adults participated. The children completed this task immediately following the paired Jane task for Experiment 2. Because we had completed adult testing for Experiment 2 prior to designing the study using bizarre faces, a different group of adults participated in Experiment 3. All adults were Cau-

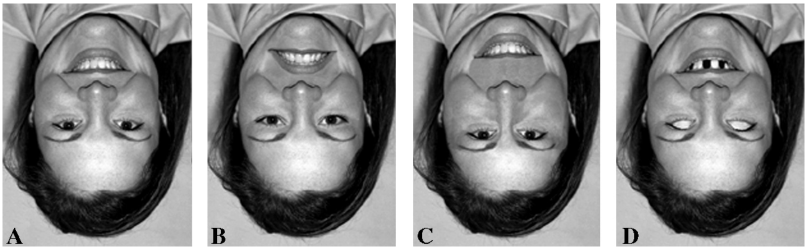

Fig. 7. Inverted versions of a single female face that was unaltered (A), Thatcherized (B), spatially altered (C), or featurally altered (D).

casian and right-handed. We did not administer a test for visual screening to the adults.[1]

Stimuli, apparatus, and procedure

Using Adobe Photoshop 4.0, we modified the original versions of four grayscale photographs of female faces that were *en face* and smiling and whose photographs included hair (Fig. 7).

To create featural distortions (Fig. 7D), we whitened the eyes and blackened portions of the teeth (see also Murray et al., 2000; Searcy & Bartlett, 1996). To create spatial distortions (Fig. 7C), we moved up the eyes so that the forehead comprised 36.5% of the length of the face, we moved the eyes closer together so that the bridge of the nose comprised 18.1% of the width of the face, and we moved the mouth down so that the chin comprised 13.9% of the length of the face (modeled after Murray et al., 2000; Searcy & Bartlett, 1996). To Thatcherize each face (Fig. 7B), we inverted the eyes and mouth within the context of the upright face (see also Lewis, 2001; Murray et al., 2000). The apparatus and informed consent procedures were the same as in Experiments 1 and 2.

Participants were asked to rate each of 32 faces on bizarreness, using a method based on Murray et al. (2000). To adapt the scale to 8-year-olds, the term "weird" was used instead of "bizarre." The experimenter explained the task to the 8-year-olds as follows: "These are pictures of people who went to a Halloween party. They tried to make their faces look as weird as possible. Some of them look really weird, and others look pretty normal. Now, you get to be the judge." The Halloween story was included because during pilot testing one child was reluctant to say that any face was weird; she thought that doing so would be "mean." Children were then asked to arrange seven cups in order of increasing size. Each cup was numbered, ranging from

[1] Having participated in Experiment 2 prior to rating bizarre faces should not have affected the performance of the 8-year-olds. In Experiment 2, children saw faces equally often where individuals differed in the shape of internal features or in the spacing of those features. These are also variations that are encountered in everyday life. Thus, making same/different judgments about featural and spacing differences in the laboratory, like those made in everyday life, should not have affected 8-year-olds' ratings of bizarre faces.

0 (smallest cup) to 6 (largest cup). The experimenter pointed to the cups while explaining that the scale ran from 0 (for *not at all weird*) to 6 (for *very, very weird*). The procedure for adults was the same except that they were not told the Halloween story.

Each version of the four manipulated faces was presented twice—once upright and once inverted within one of the four blocks of eight trials. Within each block and for each orientation, one of the faces was presented in its unaltered form, a different face in its featurally distorted form, a third face in its Thatcherized form, and the fourth face in its spatially distorted form. Within a block, each face was presented in a different form when upright (e.g., Thatcherized) than when inverted (e.g., spatially distorted). Half of the participants received one random order, and the remaining participants received the reverse order. Each face was shown on the computer screen for 3 s.

Results

The dependent variable in all analyses was the median rating by each participant across the four exemplars of each type: unaltered, featurally distorted, Thatcherized, and spatially distorted. Our primary question was the extent to which 8-year-olds' sensitivity to various kinds of distortions in upright faces resembles that of adults. We first conducted an ANOVA on upright ratings with one between-subjects factor (age) and one repeated-measures factor (manipulation). The ANOVA revealed a significant effect of manipulation, $F(3,114) = 336.03$, $p < .001$, and a significant Age × Manipulation interaction, $F(3,114) = 8.02$, $p < .001$, but no significant main effect of age, $p > .10$. Analyses of simple effects showed that 8-year-olds' ratings of unaltered faces and featurally altered faces did not differ from those of adults, $ps > .10$. In contrast, 8-year-olds rated both spatially distorted faces, $t(38) = 3.10$, $p < .01$, and Thatcherized faces, $t(38) = 2.18$, $p < .05$, as less bizarre than did adults. As shown in Fig. 8, adults' ratings did not vary across the three distorted versions, whereas children rated featurally distorted faces as more bizarre than spatially distorted and Thatcherized faces.

To determine whether children used changes in second-order relations to detect the spatial distortions and the Thatcherization, we calculated a difference score for each face type comparing upright and inverted ratings. The ANOVA revealed a significant effect of face type, $F(3,114) = 68.2$, $p < .001$, but no effect of age and no Age × Face Type interaction, $ps > .10$. As shown in Fig. 8, analyses of simple effects revealed no effect of inversion for featurally distorted faces, $p > .10$, but revealed a significant decrease in ratings of spatially manipulated faces, $t(39) = 5.82$, $p < .001$, and Thatcherized faces, $t(39) = 9.79$, $p < .001$, when stimuli were inverted. Interestingly, bizarreness ratings of unaltered faces increased, $t(39) = 2.45$, $p < .02$, when stimuli were inverted.

Discussion

As in previous studies (Lewis, 2001; Murray et al., 2000; Searcy & Bartlett, 1996), adults' ratings of Thatcherized and spatially distorted faces decreased when the stim-

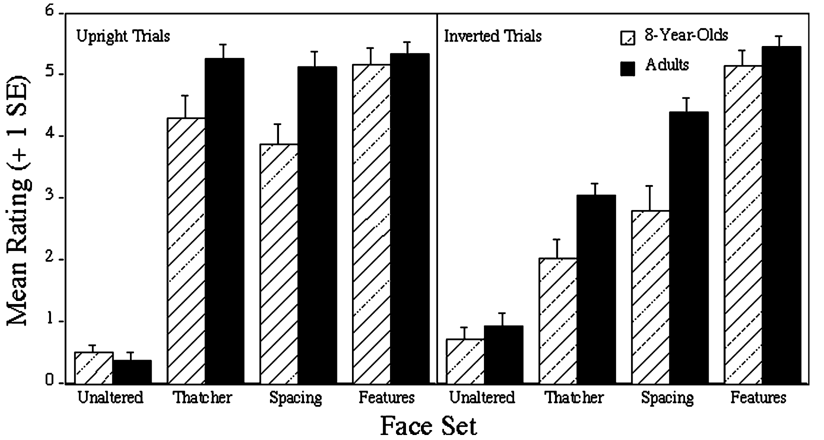

Fig. 8. Mean ratings (+1*SE*) for each version and each age group when stimuli were presented upright (left panel) and inverted (right panel).

uli were inverted, whereas their ratings of normal faces and featurally distorted faces increased slightly or did not change. These results confirm that adults' ratings of Thatcherized and spatially distorted faces were based on processing second-order relations—a type of configural processing.

The 8-year-olds showed some evidence of sensitivity to distortions of the relations among features. They rated spatially distorted and Thatcherized faces as more bizarre than unaltered faces. Like adults, they rated spatially distorted and Thatcherized faces as less bizarre when the faces were inverted, whereas their ratings of unaltered and featurally distorted faces either did not change or increased slightly. These results are consistent with those of two previous studies in which 6-year-olds judged a Thatcherized face to be less bizarre when it was rotated toward an inverted position (Lewis, 2003) or into an inverted position (Donnelly & Hadwin, 2003).

Nonetheless, 8-year-olds rated Thatcherized faces and spatially distorted faces as less bizarre than did adults. Their lower ratings cannot be attributed to their unwillingness to use both ends of the scale; like adults, they gave featurally distorted faces very high ratings and gave unaltered faces very low ratings. Rather, these results are consistent with the hypothesis that second-order relations are less salient to 8-year-olds than they are to adults. Given that even these very large spatial distortions are less salient to 8-year-olds than to adults, it is perhaps not surprising that 8-year-olds are less sensitive than adults to smaller differences in second-order relations that exist among faces in the real world—both when making same/different judgments for pairs of faces that differ only in the spacing of facial features (Mondloch et al., 2002) and when recognizing facial identity despite changes in point of view and lighting (Benton & Van Allen, 1973, as cited in Carey et al., 1980; Bruce et al., 2000; Mondloch, Geldart et al., 2003).

General discussion

This series of experiments indicates that under some conditions, 8-year-olds demonstrate moderately good sensitivity to second-order relations: when viewing time was unlimited and faces were presented simultaneously (Experiment 2) and when distortions of spacing exceeded natural limits (Experiment 3). These results indicate that the neural systems responsible for the processing of second-order relations are functioning reasonably well by 8 years of age (see also Brace et al., 2001; Carey & Diamond, 1994; Freire & Lee, 2001; Mondloch et al., 2002). Nevertheless, even under these conditions, 8-year-olds were not as sensitive as adults to the manipulations of spacing despite the fact that they showed adultlike performance for the featural changes tested in the same experiments, thereby ruling out explanations based on differences in motivation or other performance factors. Moreover, under real-world conditions—faces varying within natural limits and seen one at a time—8-year-olds apparently do not use second-order relations effectively to decode identity (Mondloch et al., 2002).

This series of experiments identified several factors that contribute to the limitations at 8 years of age. Experiment 3 indicated that second-order relations are less salient to 8-year-olds than to adults, even when those relations have been distorted so as to make the face appear grotesque. Experiment 2 indicated that 8-year-olds encode second-order relations less efficiently than do adults; unlike adults, they are more accurate when faces are presented simultaneously than when faces are presented sequentially, and it is only when faces are presented simultaneously that inversion disrupts the accuracy of 8-year-olds on the spacing set more than on the featural and contour sets. In addition, memory demands likely contribute to 8-year-olds' poorer performance when faces are presented sequentially; they may be unable to store and retrieve second-order relations, even after a brief delay. The elimination of memory demands likely contributed to their improved performance on the simultaneous task. Nevertheless, in both Experiments 2 and 3, 8-year-olds' performance in detecting second-order relations did not reach adult levels. They rated the faces with distorted second-order relations as relatively less bizarre than did adults, and their accuracy on the simultaneous version of the spacing set was significantly poorer than that of adults. This contrasts with adult levels of performance on the featural versions of both tasks.

This series of experiments also allows us to rule out some factors as explanations of 8-year-olds' poor performance on the spacing set. It is unlikely to be caused by their being distracted by individual internal and external features while performing the task. Blurring the faces reduced the salience of featural information (Experiment 1a), but neither blurring the features nor occluding them improved the performance of 8-year-olds (Experiment 1b). Their poor performance also cannot be attributed to speed/accuracy trade-offs given that 8-year-olds responded faster and more accurately on featural trials than on spacing trials in the paired Jane task (Experiment 2) and there was no time pressure in the final task where children were asked to rate faces on bizarreness (Experiment 3). Finally, their poor performance cannot be attributed to their inability to make same/different judgments about highly similar faces given that their performance was adultlike on the featural and contour sets (Experiment 2).

Children's relative insensitivity to second-order relations has implications for face recognition in the real world. It may explain why 8-year-olds made more errors than did adults on Jane's cousins even though Jane's cousins were three completely different women. Adults' higher accuracy may reflect their ability to make efficient use of differences in the spacing of features in addition to featural and contour processing. It likely also explains the results of a study in which children and adults were asked to match one of three target faces to a model based on identity despite changes in head orientation, identity despite changes in emotional expression, direction of eye gaze, emotional expression or vowel being mouthed. In that study, 10-year-olds were less accurate than adults on only one of the five tasks—matching facial identity despite changes in head orientation, a task that taps sensitivity to second-order relations (Mondloch, Geldart et al., 2003). These results are consistent with other evidence that children have more difficulty in recognizing faces in the real world whenever second-order relations are especially important, that is, when a face is seen from a new point of view, under new lighting conditions, or after a change in clothing/hairstyle (Benton & Van Allen, 1973, as cited in Carey et al., 1980; Bruce et al., 2000).

Future studies will determine the specificity of children's poor sensitivity to the spacing of features. Sensitivity to second-order relations has been attributed to expertise, gained through experience, at individuating a homogeneous class of objects (Diamond & Carey, 1986; Gauthier & Tarr, 1997; Valentine, 1991). According to the expertise argument, 8-year-olds' performance should be more adultlike when the children are tested with stimuli with which they have more experience (e.g., children's faces) or when tested with stimuli with which adults have less expertise (e.g., monkey's faces). Alternatively, and contrary to the expertise argument, children's poor performance might not be face specific; that is, they might be less sensitive to the spacing of features in nonface objects as well.

Acknowledgments

This study was funded by grants to Daphne Maurer from the Natural Sciences and Engineering Research Council of Canada and the Social Sciences Research Council of Canada.

References

Baenninger, M. (1994). The development of face recognition: Featural or configurational processing? Journal of Experimental Child Psychology, *57*, 377–396.

Bertin, E., & Bhatt, R. S. (in press). The Thatcher illusion and face processing in infancy. *Developmental Science*.

Bhatt, R., Bertin, E., & Hayden, A. (2004). *Face processing in infancy: Developmental changes in the use of different kinds of relation information*. Paper presented at the International Conference on Infant Studies, Chicago.

Brace, N. A., Hole, G. J., Kemp, R. I., Pike, G. E., Van Duuren, M., & Norgate, L. (2001). Developmental changes in the effect of inversion: Using a picture book to investigate face recognition. *Perception, 30*, 85–94.

Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Import, A., Langton, S., McAuley, S., & Wright, R. (2000). Testing face processing skills in children. *British Journal of Developmental Psychology, 18*, 319–333.

Campbell, R., & Tuck, M. (1995). Recognition of parts of famous-face photographs by children: An experimental note. *Perception, 24*, 451–456.

Campbell, R., Walker, J., & Baron-Cohen, S. (1995). The development of differential use of inner and outer face features in familiar face identification. *Journal of Experimental Child Psychology, 59*, 196–210.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195*, 312–313.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children?. *Visual Cognition, 1*, 253–274.

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component?. *Developmental Psychology, 16*, 257–269.

Collishaw, S. M., & Hole, G. J. (2000). Featural and configural processes in the recognition of faces of different familiarity. *Perception, 29*, 893–909.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*, 107–117.

Donnelly, N., & Hadwin, J. A. (2003). Children's perception of the Thatcher illusion: Evidence for development in configural face processing. *Visual Cognition, 10*, 1001–1017.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8*, 431–439.

Farkas, L. G. (1994). *Anthropometry of the head and face in medicine* (Second ed.). New York: Elsevier.

Freire, A., & Lee, K. (2001). Face recognition in 4- to 7-year-olds: Processing of configural, featural, and paraphernalia information. *Journal of Experimental Child Psychology, 80*, 347–371.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception, 29*, 159–170.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*, 1673–1682.

Goldberg, M. C., Maurer, D., & Lewis, T. L. (2001). Developmental changes in attention: The effects of endogenous cueing and of distractors. *Developmental Science, 4*, 209–219.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology, 53*, 513–536.

Lewis, M. (2001). The lady's not for turning: Rotation of the Thatcher illusion. *Perception, 30*, 769–774.

Lewis, M. (2003). Thatcher's children: Development and the Thatcher illusion. *Perception, 32*, 1415–1421.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Science, 6*, 255–260.

Mondloch, C. J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology, 86*, 67–84.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31*, 553–566.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2003). Early visual experience is necessary for the development of some—but not all—aspects of face processing. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood* (pp. 99–117). New York: Nova Science.

Murray, J., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychological Science, 11*, 492–496.

Peters, M. (1988). Description and validation of a flexible and broadly usable handedness questionnaire. *Laterality, 3*, 77–96.

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces?. *Cognition, 47*, 25–57.

Ridderinkhof, K., & van Molen, M. (1995). A psychological analysis of developmental difference in the ability to resist interference. *Child Development, 66*, 1040–1056.

Schwarzer, G. (2000). Development of face processing: The effect of face inversion. *Child Development, 71*, 391–401.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial–relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 904–915.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 43*, 161–204.