Developmental Psychology
2002, Vol. 38, No. 6, 1038–1051

Copyright 2002 by the American Psychological Association, Inc.
0012-1649/02/$5.00 DOI: 10.1037//0012-1649.38.6.1038

Children's Recognition of Caricatures

Paul P. W. Chang
Edith Cowan University

Susan C. Levine
University of Chicago

Philip J. Benson
University of Aberdeen

This study examined children's and adults' perception and recognition of facial stimuli that were either systematically exaggerated (caricatures) or de-exaggerated (anticaricatures) relative to a norm face. The results showed that all age groups perceived caricatures as the most distinctive versions of a face and anticaricatures as the least distinctive, although the effect was smallest for 6-year-olds. In general, caricatures were identified as quickly as the veridical faces and faster than the anticaricatures. Across all age groups, participants' familiarity with the stimulus faces interacted with degree of caricature to determine speed of processing as well as choice of best likeness. The results are discussed in relation to the idea that distinctiveness information in a face is represented in relation to a norm.

A well-known finding in the literature on face recognition is that faces judged as distinctive are more easily remembered than faces judged as typical. This recognition advantage is known as a *distinctiveness effect.* For adults, the distinctiveness effect has been found for both unfamiliar faces and familiar faces regardless of whether the faces are familiar through media exposure or whether the faces are personally known by the observer (Bartlett, Hurry, & Thorley, 1984; Going & Read, 1974; Valentine & Bruce, 1986a, 1986b). Moreover, distinctive faces are remembered better after long delays than are typical faces (Shepherd, Gibling, & Ellis, 1991). To date, only one study that we know of has compared children's recognition of distinctive versus typical faces. Johnston and Ellis (1995) tested children ranging in age from 5 to 13 years and found that 5-year-olds did not show a recognition advantage for distinctive faces over typical faces even though accuracy for both types of faces was well above chance. Instead, the characteristic advantage for processing distinctive faces began to emerge at around age 7 and resembled the adult pattern at around age 9.

An influential model to account for distinctiveness effects in adults proposes that adults' internal representations are encoded in a multidimensional "face space" (MDFS; Valentine, 1990, 1991). According to the model, dimensions of the space correspond to any features or spatial relationships between features that could help identify (or distinguish between) different individuals. Although the model does not specify which dimensions are used in encoding, several multidimensional scaling studies suggest that spatial relations between the eyes and eyebrows, the length and color of hair, and the chin shape may be regarded as likely candidates (e.g., Pedelty, Levine, & Shevell, 1985; Rhodes, 1988; Shepherd, Davies, & Ellis, 1981). Other dimensions may include the perceived age of the face (Johnston & Ellis, 1995), the gender of the face (Rhodes, Hickford, & Jeffery, 2000), or the race of the face (Byatt & Rhodes, 1998). The values on the feature dimensions are deemed to be normally distributed around a central tendency depending on the population of faces that a person has seen (Valentine, 1991). Faces that look similar to each other are represented by points or vectors that are close to each other in face space. Their actual *location* in space, however, is determined relative to the central tendency of the feature dimensions (see Figure 1). Using the framework of MDFS, Valentine (1991) proposed two models by which faces may be encoded. One is a norm-based coding model in which faces are represented in the face space as vectors derived from a norm or prototype face. The other is an exemplar-based model in which faces are encoded as discrete points in space. Both models support the notion that there is an area of central tendency in the face space that represents the area of highest exemplar density; as distance from the origin increases, the density decreases.[1] In general, the within-category exemplar density gradients assumed by the MDFS model are inherent in other models of categorization that attempt to explain

Paul P. W. Chang, School of Psychology, Edith Cowan University, Joondalup, Western Australia, Australia; Susan C. Levine, Department of Psychology, University of Chicago; Philip J. Benson, Department of Psychology, University of Aberdeen, Aberdeen, United Kingdom.

This article was based on Paul P. W. Chang's doctoral dissertation submitted to the University of Chicago. Portions of this research were presented at the Twelfth Annual Convention of the American Psychological Society, Miami, Florida, June 2000. We thank the students who kindly participated in this research.

Correspondence concerning this article should be addressed to Paul P. W. Chang, School of Psychology, Edith Cowan University, 100 Joondalup Drive, Joondalup, Western Australia 6027, Australia. E-mail: p.chang@cowan.edu.au

[1] The norm- and exemplar-based models are difficult to distinguish empirically (although see Byatt & Rhodes, 1998, for a recent study). Both models, however, predict a recognition advantage for distinctive facial stimuli given the assumption of a differential density gradient in face space.

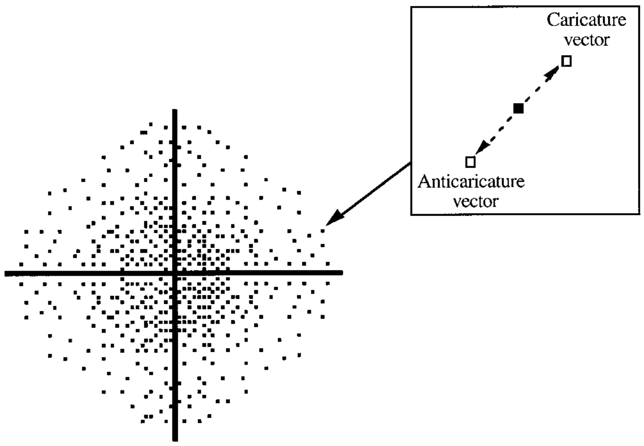

Figure 1. A hypothetical representation of face space. Each of the plotted points represents a previously seen face located in an *n*-dimensional face space (only two dimensions are shown here). The axes are not labeled but represent any feature dimension that could be used to individuate faces. The origin represents the central tendency of the feature dimensions for a population of faces (see Valentine, 1991, and Johnston & Ellis, 1995). The inset shows the relationship in face space between the veridical depiction and anticaricatured and caricatured versions of a face relative to the central tendency of the axes. The arrows in the inset show how caricatured and anticaricatured depictions may be more or less distinctive, respectively, in relation to a putative norm face or to the central tendency of faces.

the prototype advantage in shape-based classification tasks (e.g., Nosofsky, 1991).

One of the main assumptions of the MDFS model is that a person's lifetime experience with faces will contribute to the distribution of facial representations within the space. Typical faces, by definition, are more commonly encountered and look more similar to other faces than do distinctive faces. Thus, typical faces are represented around the origin, where a large number of close neighbors are clustered nearby. In contrast, distinctive faces are assumed to differ from the central tendency on at least one feature dimension and therefore are represented in the periphery of space, where there are fewer neighbors nearby. Valentine's (1991) model accounts for the distinctiveness effect by positing that the detection of target activity against a background of distractor activation is more difficult for typical faces than distinctive faces because typical faces are located in areas of space with a higher density of distractors. Hence, typical faces may be subject to a greater degree of interference from activated distractors than are distinctive faces.

One problem with examining a distinctiveness effect for faces is that most studies ask participants to recognize a range of different faces that may indeed vary in distinctiveness. As pointed out by Rhodes and her colleagues (e.g., Rhodes, Brennan, & Carey, 1987; Rhodes & Moody, 1990), however, faces that vary in distinctiveness may also covary on other dimensions (such as attractiveness, age, gender, perceived health, and so on). An alternative to presenting a range of distinctive and typical faces is to systematically vary distinctiveness within the same face in relation to a norm face (Brennan, 1985). These stimuli are known as *caricatures* (in which distinctive facial features are exaggerated) or *anticaricatures* (in which distinctive features are reduced). For example, if an individual has a narrow face, then in a caricature, the narrowness of the face is exaggerated in relation to other faces from the same population. In contrast, for the individual with a narrow face, an anticaricatured depiction would make the face less narrow, making it look more like the norm face. Multidimensional scaling analyses on adults' similarity ratings for pairs of caricatured and anticaricatured faces provide support for the idea that caricatures (compared with anticaricatures) are located further from the central tendency of a set of faces (Lee, Byatt, & Rhodes, 2000).

Using computer-generated images, studies have found that caricatures either are recognized as quickly as veridical depictions (Benson & Perrett, 1994; Calder, Young, Benson, & Perrett, 1996, Experiment 2; Rhodes & Tremewan, 1994) or are actually recognized faster than veridical depictions (e.g., Benson, 1997; Calder et al., 1996, Experiment 3; Rhodes et al., 1987). In these studies, anticaricatures are always recognized more slowly than both caricatures and veridical depictions. In addition, caricatures are often judged as better likenesses of well-known individuals than are veridical or anticaricatured depictions (Benson & Perrett, 1991a; Rhodes et al., 1987). In sum, distortions of the face in which distinctiveness information is exaggerated are recognized at least as well as undistorted versions of the face.

To date, there has been little systematic research on children's recognition of caricatures, except by Ellis (1991), who described two pilot experiments in an unpublished report. In the first experiment, Ellis found that 4–9-year-olds selected a caricature when asked to choose the best likeness of a male and a female celebrity. Indeed, the 4–6-year-olds selected a more extreme caricature than the one selected by the 7–9-year-olds. In the second experiment, children 6, 9, or 11 years of age selected a caricature as depicting the best likeness of an unspecified number of their classmates' faces. It was not reported in Ellis's summary, however, whether the mean caricature level (MCL) of the preferred images was significantly different from 0%, the value expected from the null hypothesis that caricaturing has no effect on best likeness judgments. Taken together, the pilot data showing that children as young as age 4 show caricature effects stand in contrast to results indicating that distinctiveness effects for faces emerge only at around age 7 (Johnston & Ellis, 1995).

The goal of the present research was to investigate whether or not distinctiveness was important for children's representations of faces. Two experiments were designed to examine caricature effects in children. In the first experiment, participants were asked to pick which of three types of depictions of unfamiliar faces—caricatures, veridical depictions, or anticaricatures—were the most or least distinctive versions of a face. If the adult space is differentiated (see Figure 1), as has been suggested by some researchers (e.g., Johnston & Ellis, 1995; Valentine, 1991), then adults should pick caricatures as the most distinctive versions of an individual's face and anticaricatures as the least distinctive versions. By contrast, if young children are not sensitive to distinctiveness information of faces, then their responses should be random or should at least show a pattern different from that of older children and adults.

The second experiment was designed to examine age-related changes in the participants' recognition of their classmates' faces compared with the faces of age-matched children who attended a different school but whose faces and names were learned from a photograph. The advantage of this design is that the same target faces are seen by two groups of participants, one that is familiar with the faces and one that is not. Thus, only familiarity is varied across the groups, and the effects of caricature recognition, related to the face stimuli per se, are mitigated. As described earlier, Ellis's (1991) pilot study found that young children showed a positive MCL (i.e., a caricature) for the preferred image, which is a result that is similar to that for adults. This ability, however, may not yet be fully developed. If children use distinctiveness information in recognizing faces, then we would expect them to show caricature effects in recognition performance as well, a finding that Johnston and Ellis (1995) did not observe in children younger than 7 years of age. In Experiment 2, two aspects of caricature recognition were examined. First, in an identification task, children and adults were asked to name caricatured, veridical, and anticaricatured faces; naming accuracy and speed were measured for the three types of depictions. A caricature advantage would be reflected by faster response times and more accurate recognition of caricatures than of veridical and anticaricatured depictions. Second, in a best likeness task, participants were asked to judge the various depictions according to their resemblance to the depicted individual. If there is a developmental change in the manner in which distinctiveness in faces is encoded, then caricature effects that are found in adults may not be found in young children. These experiments should provide basic information about whether young children are sensitive to distinctiveness information in faces and, if so, whether they use this information in performing a variety of tasks in the same manner as older children and adults. Obtaining this kind of information is an important first step in building more formal theories of developmental changes in face processing.

Experiment 1

It is possible that caricatured depictions of an individual's face are regarded as more distinctive versions of the face than are anticaricatures. To date, only research involving adults who viewed line drawings (Rhodes & Tremewan, 1996) or photographic images of faces (Rhodes, Sumich, & Byatt, 1999) has provided empirical evidence that supports this distinction. The purpose of Experiment 1 was to investigate whether children would be sensitive to variations in distinctiveness as the caricature level of faces was varied.

Method

Participants. A total of 56 primary school children (30 boys and 26 girls) participated in the experiment: 20 children were from Grade 1 (mean age = 6 years 10 months), 20 from Grade 3 (mean age = 8 years 11 months), and 16 from Grade 5 (mean age = 11 years 0 months). Twenty-two adult volunteers (mean age = 27 years 6 months) were also tested. All the participants were Caucasian, and the proportions of males and females in each age group were approximately equivalent.

Stimuli. Color photographs of 83 children and 26 adults were used to make the facial stimuli. All the photographic models were Caucasian (and all were the actual participants in Experiment 2). Photographs were taken of 27 children from Grade 1, 32 from Grade 3, and 24 from Grade 5. The 26 adults were graduate students attending the University of Chicago. Roughly half of the models at each age group were male. All the photographs were head shots (from the shoulders up) and were taken under constant background and lighting. The models were asked to look directly at the camera and to pose with a relaxed, neutral expression. Only adults who did not wear beards or mustaches served as models. The models who wore glasses (or other accessories such as earrings, hats, etc.) were asked to remove them, but none of the individuals depicted in the final stimulus set wore glasses.

The photographs were digitized (at a resolution of 100 dots per inch) so that they could be used to make the caricatures. The software used to create color photographic caricatures and anticaricatures has been described in detail elsewhere (Benson, 1994; Benson & Perrett 1991a, 1991b, 1992), so only a brief description is provided here. Caricature generation involved four steps. First, a fixed set of 208 landmark points was placed manually around clearly defined internal and external facial features such as the eyes, nose, mouth, and so on. The points were automatically joined with lines to produce a veridical tracing of the face. Second, a point-by-point, line-by-line match was made between a target face and its age- and gender-appropriate norm.[2] The norm faces were created automatically by averaging the coordinates of all the people who were photographed, blocked by age and gender. Eight norm faces were created: For each of the four age groups, there was one male norm and one female norm. At least 12 male and 12 female photographs at each of the four age groups were averaged to create the eight norm faces.

In the third step, all metric differences between the target face and its norm were increased to create caricatures or decreased to create anticaricatures. For example, a 100% caricature is operationally defined as an image in which the spatial differences between a stimulus face and a norm face are doubled. Finally, each veridical line drawing was divided into a lattice of 340 interconnected triangles, using the set of landmark points. The color pixel values contained in a triangulated veridical image were then remapped, pixel by pixel, into the corresponding triangles in the caricatured image, using the reference points of the caricatured image. In effect, the triangles of the caricatured image were "stretched" or "shrunk" in relation to the corresponding triangles on the veridical image (Benson & Perrett, 1991b).

The final stimulus set consisted of 48 faces. There were six male and six female faces from each of the four age groups (the mean ages of the faces for the first graders, third graders, fifth graders, and adults were 6 years 11 months, 8 years 9 months, 10 years 9 months, and 28 years 7 months, respectively). For each stimulus face, five depictions were generated: the –36% and –18% anticaricatures, the veridical (0%) depiction, and the 18% and 36% caricatures. These levels are within the range of those that have been used by other researchers who have found caricature effects in adults for familiar faces (e.g., Rhodes et al., 1987; Rhodes & Tremewan, 1994). A sample set of face depictions is shown in Figure 2.

[2] There were pragmatic and theoretical reasons underlying our decision to caricature individual faces against their age- and gender-appropriate norms. The decision was pragmatic in the sense that it seems more reasonable that a male Caucasian 6-year-old's face be caricatured against a norm derived from other male Caucasian 6-year-olds, and not a norm derived from, say, adult faces. The decision was theoretical in the sense that the primary aim of our study was not to directly examine the effect of different norms on recognition or likeness judgments (by contrast, see Perrett et al., 1998, and Rhodes et al., 2000, for evidence that a masculinized or feminized norm may affect attractiveness judgments). As pointed out by an anonymous reviewer, our procedure to generate caricatures carries with it the implicit assumption that perceived age, gender, and even race may be important dimensions of face space. This is a point we readily endorse, especially given evidence that young children can discriminate among the faces of young children more readily than among the faces of older children and adults (George, Hole, & Scaife, 2000) and evidence that adults may differentially encode faces according to age and gender (Johnston, Milne, Williams, & Hosie, 1997). As described in the Results section, however, the factors of age and gender of the facial stimuli were not significant and did not interact with other factors.

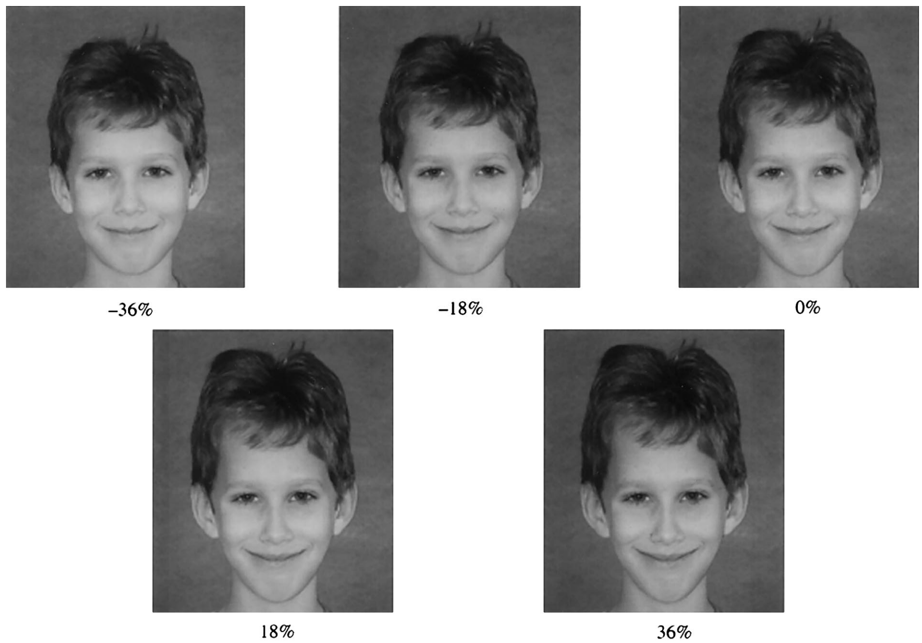

Figure 2. Sample stimulus set including anticaricatures (−36% and −18%), the veridical depiction (0%), and caricatures (18% and 36%) of an individual's face. (Note that the stimuli were presented in color.)

Apparatus and procedure. For this and the other experiments described later, the stimuli were presented in 12-bit color on a 27-cm wide monitor (with a screen resolution of 832 × 624 kHz and a refresh rate of 75 Hz) controlled by a Power Macintosh 7100 computer. The participant sat 60 cm away from the monitor, which resulted in each depiction subtending a horizontal visual angle of about 6.52° and a vertical visual angle of about 9.0°. Software for conducting psychology experiments (SuperLab 1.68) was used to store and display the stimuli and record responses.

Participants were tested individually in a quiet room at their school. On each trial, a set of five depictions of an individual's face was presented on the monitor: As demonstrated in Figure 2, three depictions were shown in the top row and two in the bottom row. At the beginning of the session, while the first set of depictions was on the monitor, the following instructions were read:

> I want you to think about all the faces you have ever seen. Now, look at these pictures on the screen. Which one of these faces do you think looks *most different from* (or *most like*) [emphasis added] the faces you have seen before? Please point to it.

Occasionally during subsequent trials, the instructions were repeated, along with a reminder that there were no right or wrong answers and that the experimenter was simply interested in what participants thought of the faces. After the participant pointed to one of the depictions, a new set of five depictions of another individual's face was then presented. The participants proceeded through the trials at their own pace.

A block of trials consisted of the faces of six individuals, blocked by age and gender of the face. The trials within a block were randomized. Each participant saw the caricature sets of all 48 individuals twice: once when they made a "most different" judgment and once when they made a "most like" judgment. Half the participants made all of their "most different" judgments first, and half made their "most like" judgments first. Each session lasted about 40 min and included a break after each block of six trials.

Results and Discussion

To obtain an estimate of the distinctiveness effects, we calculated the MCL of participants' judgments by averaging the caricature level of the depiction chosen as most different from or most like the faces they had seen before across participants for each age group. A preliminary analysis of variance (ANOVA) indicated that there was no difference between the MCL values as a function of age or gender of the facial stimuli and that these factors did not interact with any other factors (all Fs < 1); therefore, the MCL values were combined over these factors in subsequent analyses. As shown in Figure 3, a positive MCL was obtained for the "most different" condition for all age groups (6-year-olds, MCL = 4.37%, $SE = 0.98\%$; 8-year-olds, MCL = 22.56%, $SE = 1.83\%$; 10-year-olds, MCL = 27.96%, $SE = 1.08\%$; adults, MCL = 31.64%, $SE = 0.60\%$). By contrast, a negative MCL was obtained for the "most like" condition for all age groups (6-year-olds, MCL = –3.58%, $SE = 1.14\%$; 8-year-olds, MCL = –19.44%, $SE = 1.71\%$; 10-year-olds, MCL = –25.48%, $SE = 0.68\%$; adults, MCL = –27.14%, $SE = 0.68\%$). This pattern was confirmed by the results of a two-way ANOVA, with age (6 years, 8 years, 10 years, adult) as a between-subjects factor and instruction condition (most like or most different) as a repeated within-subject factor. The main effect of instruction condition was significant, $F(1, 74) = 1{,}738.75$, $p < .0001$ (Ms = –18.78% and 21.56% for the "most like" and "most different" conditions, respectively). This effect was moderated by a significant Age × Condition interaction, $F(3, 74) = 143.27$, $p < .0001$. Post hoc pairwise comparisons for the "most like" condition indicated that all adjacent age groups were significantly different from each other except the 10-year-olds and adults; for the "most different" condition, all adjacent age groups were significantly different from each other (Tukey's hon-

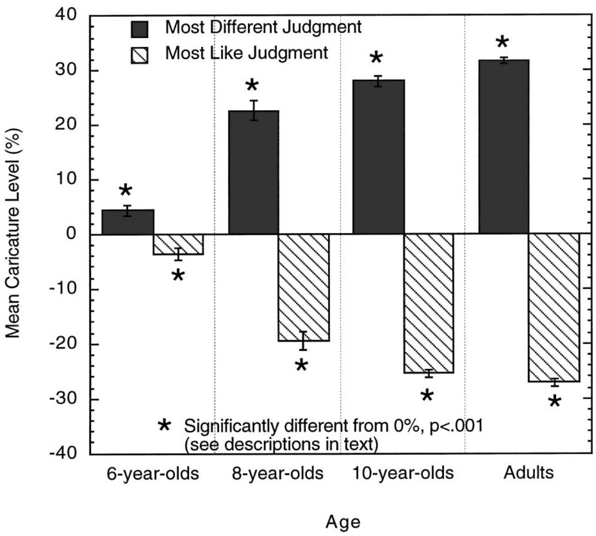

Figure 3. Mean caricature level (MCL) of distinctiveness judgments as a function of age in Experiment 1. Error bars depict the standard error of the MCL. In the graph, an MCL of 0% represents chance selection.

estly significant difference [HSD] tests, $ps < .01$). The main effect of age was not significant, $F(3, 74) = 1.38$, $p = .25$.

For the sake of clarity, separate ANOVAs on the MCLs were carried out for each age group to further examine the distinctiveness effects. For each age group, the null hypothesis was that the MCLs for the "most different" and "most like" conditions would not be different. In addition, a priori *t* tests were carried out to test the null hypothesis that varying the distinctiveness of the face would have no effect on the "most like" or "most different" judgments. Under this hypothesis, participants' MCLs would not be different from an expected MCL of 0%—the value expected from random responses. The *F* tests and planned contrasts are summarized in Table 1, which shows that the main effect of instruction condition was significant for every age group. In addition, the planned contrasts indicated that the MCLs for each age group were significantly different from 0%.

As Rhodes (1996) pointed out, a positive MCL suggests a preference for a caricatured depiction even though this value may also reflect either an equal preference for veridical and caricatured depictions or an outright preference for veridical depictions with caricatures preferred more than anticaricatures. The same reasoning applies to negative values of the MCL. For these reasons, it is important that the MCL values be considered in conjunction with the mean frequency with which each caricature level was selected. The frequency with which each depiction was selected under the "most like" and "most different" conditions was averaged across all stimulus faces and over all participants for the four age groups. These frequencies, expressed as a mean percentage of selection for each caricature level, are shown in Figure 4. In Figure 4, 20% represents chance because on any given trial, there is a .20 probability of randomly selecting a given depiction.

In comparing the selection performance, a justifiable concern is that the selection functions shown by the 6-year-olds relative to the older participants may indicate that the youngest participants did not understand the task instructions. If this was the case, then their selection performance should be random in at least one condition or random in both conditions. We suggest that there is evidence that at least some of the 6-year-olds understood the nature of the task. First, as described above, the MCLs for the 6-year-olds were positive under the "most different" condition and negative under the "most like" condition, just like those of the older participants in both conditions. Second, the main effect of instruction condition was significant for all age groups, and crucially, each of the MCLs was significantly different from zero. Third, an examination of Figure 4a—the "most different" condition—shows that for all participants, the 36% caricature was chosen most often, followed by the 18% caricature, and then by the rest of the depictions. Hence, the positive MCLs found under the "most different" condition were not the result of an equal preference for the veridical and the caricatured depictions, nor were they the result of an outright preference for veridical depictions with caricatures preferred more than anticaricatures. An examination of Figure 4b—the "most like" condition—shows that the 6-year-olds selected the two anticaricatures most often (the –18% depiction followed by the –36% depiction) followed by the rest of the depictions. The trend for the 6-year-olds, though not as pronounced, was similar to the pattern shown by the older participants (who clearly chose the –36% depiction most often). Hence, the negative MCLs found under the "most like" condition were not because of an equal preference for the veridical and the anticaricatures depictions, nor were they the result of an outright preference for veridical depictions with anticaricatures preferred more than caricatures. We tentatively suggest that the MCL analyses, coupled with the fact

Table 1
Main Effect of Instruction Condition on Mean Caricature Levels (MCLs) by Age Group

	ANOVA statistics			Planned contrasts[a]				
					Most like		Most different	
Age	*df*	*F*	*p*	*df*	*t*	*p*	*t*	*p*
6 years	1, 19	29.18	<.001	19	−3.14	<.01	4.46	<.001
8 years	1, 19	189.50	<.001	19	−11.36	<.001	12.31	<.001
10 years	1, 15	1700.45	<.001	15	−37.29	<.001	25.99	<.001
Adult	1, 21	2580.12	<.001	21	−51.36	<.001	53.07	<.001

[a] A priori *t* statistic (according to two-tailed tests) for testing the significance of the MCL difference from zero (see description in text).

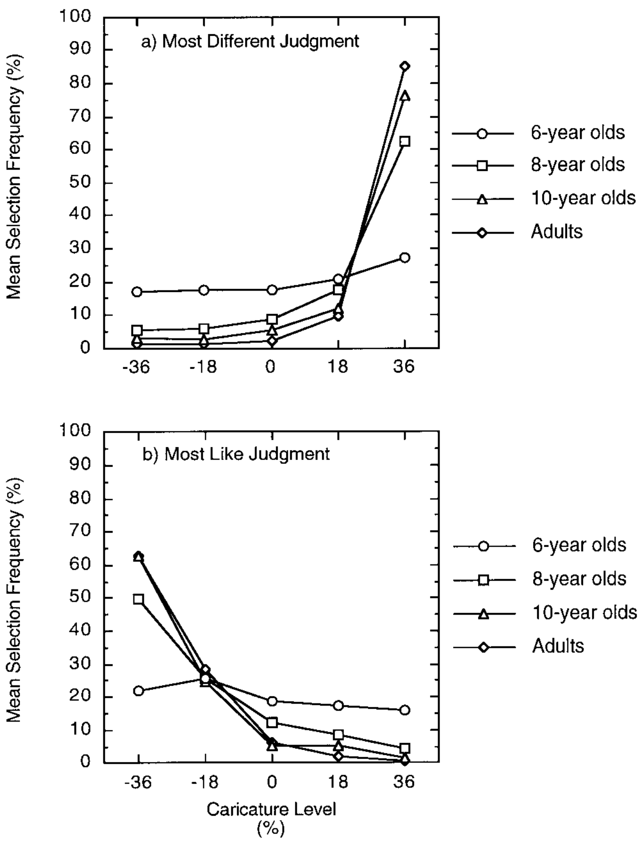

Figure 4. Mean selection frequency (expressed as a percentage) as a function of the caricature level for "most different" and "most like" judgments in Experiment 1. In the graphs, a 20% selection frequency represents chance selection.

that the 6-year-olds showed the same trends as the older participants in both conditions, indicate that at least some of the 6-year-olds understood the nature of the task and showed sensitivity to distinctiveness in faces. Clearly, however, the effect is not as strong as the effect in older participants.

The results are also consistent with Rhodes and Tremewan's (1996) findings that distinctiveness ratings for line drawings are positively correlated with degree of caricature. Along with findings from Rhodes et al. (1999), the present results confirm that this relationship holds for photographic images and serve as a manipulation check for the caricature sets in terms of perceived distinctiveness.

Experiment 2

As suggested by the results of Experiment 1, a set of depictions that varies in caricature level also seems to vary in perceived distinctiveness. The purpose of Experiment 2 was to investigate age-related changes in children's recognition of caricatures using depictions of familiar and initially unfamiliar faces. Two measures of recognition performance for faces were examined: which type of depiction (caricature, veridical, or anticaricature) would be identified faster and/or more accurately and which type of depiction would be judged as the best likeness of an individual's face. Previous developmental work suggests that young children do not show a recognition advantage for distinctive over typical faces, as is usually found for adults (Johnston & Ellis, 1995). Thus, one prediction that follows is that older children and adults, but not younger children, will show caricature effects in recognizing faces. If this is the case, then an interaction between age and caricature level should emerge.

When familiar faces were used as stimuli, caricature effects were found in several studies. The results, however, were not as clear when the faces used were initially unfamiliar. Some studies reported that for unfamiliar faces, veridical depictions were recognized faster than caricatured depictions (Rhodes & Moody, 1990) and were judged as better likenesses than were caricatures (Rhodes et al., 1987) even though a caricature advantage was found for familiar faces. These results were initially interpreted as evidence that representations for familiar faces are caricatured in memory (termed a *caricature-trace* model by Rhodes, 1996). By this account, caricatured depictions may be easier to recognize and may be judged as better likenesses because they match the stored (caricatured) representation available in memory more closely than do veridical or anticaricatured depictions. It seems reasonable to assume that familiarity established through one's daily interactions with people results in different face representations than familiarity acquired during an experiment (see Newcombe & Lie, 1995, for further discussion of this point). Thus, caricature effects through exposure to a photograph may be restricted to relatively familiar faces.

In contrast, other studies have reported caricature effects for initially unfamiliar faces. Rhodes and Tremewan (1994) found that enhanced line drawing caricatures were recognized more accurately than veridical depictions both by classmates of the people who were depicted and by another class who had only seen the faces as photographic images. Stevenage (1995), using artist-drawn images, found that the names for a set of unfamiliar faces were learned more quickly when the faces were presented as caricatures than when they were presented as veridical drawings. These results suggest that facial representations need not be represented as exaggerations in long-term memory for caricatured stimuli to be effective. In Experiment 2, the effects of familiarity were examined. Familiarity was manipulated by comparing recognition performance for photographs of participants' classmates with recognition performance for photographs of age-matched peers who were not seen prior to the experiment. If a caricature-trace account is correct, then caricature effects should be restricted to highly familiar faces.

Method

Participants. There were 123 participants, 109 of whom were photographed in the manner described in Experiment 1. (Note that only 48 of these faces composed the stimulus set.) There were 87 primary school children (46 boys and 41 girls) who were recruited from two Chicago area schools. The children were drawn from a total of six classrooms, one at each grade level at each school. Across the two schools, there were two first-grade classes (29 participants; mean age = 6 years 10 months), two third-grade classes (32 participants; mean age = 8 years 10 months), and two fifth-grade classes (26 participants; mean age = 10 years 9 months). In addition, 36 undergraduate and graduate students (mean age = 27 years 2 months) were tested: 24 were from the Department of Psychology, and 12 were from the Divinity School. As before, all the participants were

Caucasian, and the proportions of males and females in each age group were approximately equivalent.

Stimuli and apparatus. The stimuli comprising –36%, –18%, 0%, 18%, and 36% depictions of individuals' faces were generated in the manner described in Experiment 1. In each session, a participant learned the names of 18 faces, 6 of which were classmate faces and 12 of which had never been seen before. Half the stimulus faces were male and half were female. The faces were presented in six blocks, with the participant learning the names of three faces per block. For the children, six of the unfamiliar faces were same-age peers from another school, and six were adults (psychology graduate students). For the adults, the unfamiliar face stimuli were six graduate students from a different program and six unfamiliar children; equivalent numbers of adults viewed children's faces from each of the three age groups.

The computer, monitor, and software used to present the faces and record responses were identical to those used in Experiment 1 with the following exception: In the identification task, voice-activated response times (RTs) were recorded by a small microphone placed 15 cm in front of the participant.

Procedure. Each participant completed three tasks. First, participants learned the names of the three faces in a block in the training phase. (Only three faces were used in each block to try and ensure that a block of trials did not take too long to complete in order to possibly avoid frustrating the youngest participants.) Second, speeded identification performance for these three faces was tested. Third, immediately after the identification trials, participants made best likeness judgments for these three faces. This cycle was repeated until all six blocks of faces were tested. Each session lasted about 45 min; participants were tested individually in a quiet room at their school.

Part 1: Training phase. Digitized photographs of each face in a smiling pose (taken during the same photographic session) were used in the training phase. The aim of the training phase was for participants to learn the actual name of each of the faces. Although training on classmates' faces was probably unnecessary, the same training procedure was carried out for all of the faces to keep the procedures consistent. Familiar or unfamiliar faces were presented to each participant in blocks of three male or three female faces at a time, with the blocks presented according to 1 of 12 fixed random orders. The orders were counterbalanced across participants.

Throughout this study, the participant was seated at a table, 60 cm from the monitor. The following instructions were read:

> Now, I'm going to show you some faces. Some of the faces you have seen before, and some you haven't seen before. I'm going to tell you the names of the faces, and I want you to try and remember the name of each face. Then, I am going to show you the faces again and I want you to tell me the name of each face.

After the instructions were read, training proceeded as follows. For each of the three faces in a block, photographs with smiling poses were shown in the center of the monitor, one at a time, for 3 s each, with the experimenter saying the name for the first two training cycles. The faces were then presented again in a random order, and the participant was instructed to name each face. The experimenter provided the name when necessary. The training cycle was repeated until all three faces in a block were named correctly two times in a row.

Part 2: Identification task. The identification task was carried out immediately after the participant learned the names of the three faces in a block. At the beginning of the identification trials, the experimenter read the following instructions:

> Now, I want to show you some of the faces you just saw [in the training phase]. Look at the screen and let me know when you are ready to begin. Then, get ready, because a face will appear in the middle of the screen. Say the person's name out loud. Please try to say the right name as quickly as you can, without saying "umm" or "uhh" before the name. You will see each person's face a few times. The faces will be shown in a mixed-up order. Do you understand what you are supposed to do?

The correct procedure for making a vocal response was then demonstrated. The experimenter initiated the beginning of a block of trials with a keyboard press. Two seconds later, a single face depiction from the stimulus set was presented in the center of the monitor and remained on the screen until the participant said the name. The recording of the voice-activated RTs began as soon as the stimulus was presented and ended when a vocal response was recorded. Another face was presented 2 s after a response. There were 30 trials per block, consisting of two presentations of each of the five caricature levels for the three faces in a block. Trials within a block were randomized.

Part 3: Best likeness task. Immediately after the identification task, we asked participants to choose the best likeness for each face. Each trial consisted of a set of five face depictions presented on the screen in the arrangement shown in Figure 2. To guide participants' likeness judgments, we read the following instructions:

> Now, I've put some pictures of [John] on the screen. Which one of these pictures do you think looks *most* [emphasis added] like [John]? Please point to it.

Occasionally during the trials, the instructions were repeated, along with a reminder that there were no right or wrong answers and that the experimenter was simply interested in what they thought of the faces. After the participant responded by pointing to one of the depictions, a set of five depictions of another person's face was presented. Participants were allowed to proceed through the trials at their own pace. The depictions in the array were arranged according to 1 of 24 fixed random orders such that each depiction appeared at least four times in each position within the array. The order in which the faces were judged was random.

Results

The results of each task are presented in the order in which they were carried out: training, face identification, and best likeness judgments.

Training. For the familiar faces, none of the participants needed more than two training blocks to reach the criterion (i.e., correctly naming the three faces twice in a row). For the unfamiliar faces, an ANOVA was carried out on the mean number of training blocks needed to reach the criterion of naming all the faces, with age (6 years, 8 years, 10 years, adult) as a between-subjects factor. The ANOVA revealed a main effect of age, $F(3, 119) = 36.64$, $p < .001$; the 6-year-olds required more training blocks on average ($M = 2.6$) than did the 8- and 10-year-olds and the adults (Ms = 2.05, 2.06, and 2.02, respectively). Tukey's post hoc tests revealed significant differences between the 6-year-olds and the other age groups (ps $< .01$); none of the other age groups differed significantly from each other.

Identification accuracy. The overall error rate was less than 1.6% for all age groups. On trials in which an error was recorded, virtually all of them (91.2%) occurred when the vocal response did not trigger the microphone. Because there were so few misidentifications, no further analyses of identification accuracy are reported here.

Identification speed. A preliminary ANOVA was conducted on the mean correct RTs to examine the effects of age and gender of the facial stimuli (within-subject factors) and the participant's gender (between-subjects factor). None of these factors reached significance (Fs < 1), nor did these factors interact. Hence, the correct RTs were combined over these factors in subsequent analyses.

The mean correct RTs of each participant were computed for each caricature level as a function of familiarity.[3] The mean RTs averaged over participants are presented in Figure 5. The RTs were analyzed in a three-way ANOVA, with age (6 years, 8 years, 10 years, adult) as a between-subjects factor and caricature level (–36%, –18%, 0%, 18%, 36%) and familiarity (familiar or unfamiliar) as within-subject factors.

The main effect of age was significant, $F(3, 119) = 58.23$, $p <$ .0001. Pairwise comparisons showed that the 6-year-olds were slower than all other age groups (M = 918 ms) and that the adults (M = 704 ms) were significantly faster than the 8-year-olds (M = 765 ms) but not the 10-year-olds (M = 734 ms), Tukey's HSD tests, ps $< .05$. None of the other differences between the age groups were significant. The main effect of familiarity was also significant, $F(1, 119) = 387.78$, $p < .0001$, indicating that familiar faces were named faster than unfamiliar faces (716 ms vs. 837 ms).

Finally, the main effect of caricature level was significant, $F(4, 476) = 35.58$, $p < .0001$. Pairwise comparisons showed that the 36% caricature was named significantly faster than the 0%, –18%, and –36% depictions; that the 18% caricature was named faster than the –18% and –36% anticaricatures; and that the 0% depiction was recognized faster than the –36% anticaricature (Tukey's HSD tests, ps $< .05$). None of the other depictions differed significantly from each other.

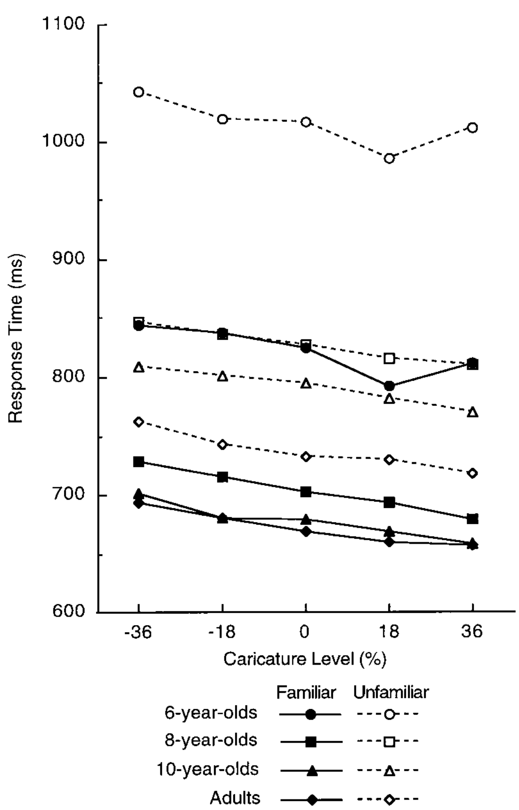

Figure 5. Mean correct response times as a function of familiarity and caricature level for the 6-, 8-, and 10-year-olds and the adults in Experiment 2.

The main question asked in this experiment was whether or not children, like adults, would show caricature effects in recognition, so the interactions that address this question are described first. The analysis indicated that neither the Age × Caricature Level interaction, $F(12, 476) = 1.24$, $p = .3$, nor the Age × Caricature Level × Familiarity interaction, $F < 1$, was significant. The absence of interactions involving both age and caricature level provides no evidence that caricature level affected some age groups differently than others.

The only interaction that was significant was the Age × Familiarity interaction, $F(3, 119) = 18.22$, $p < .0001$. These data are presented in Table 2. Post hoc pairwise comparisons for the familiar faces indicated that the 6-year-olds were significantly different from the other age groups; for the unfamiliar faces, all of the age groups were significantly different from each other except for the 8-year-olds and the 10-year-olds (Tukey's HSD tests, all ps $< .01$). None of the other age groups differed significantly from each other. As shown in Table 2, the difference between naming familiar and unfamiliar faces decreased with increasing age; this difference was largest for the 6-year-olds (192 ms), followed by the 8-year-olds (124 ms), the 10-year-olds (114 ms), and the adults (65 ms). As shown in Table 2, however, this decrease was due mainly to a marked improvement in naming speed for unfamiliar faces with age.

Best likeness judgments. The frequency with which each caricature level was selected as the best likeness was averaged across all stimulus faces and over all participants for the four age groups as a function of familiarity. These frequencies, expressed as mean percentages of selection for each caricature level, are presented in Figure 6.

We calculated the MCLs of the depictions chosen as the best likeness by each age group. A preliminary ANOVA indicated that there was no difference between the MCL values as a function of age or gender of the facial stimuli and that these factors did not interact with any other factors (all Fs < 1); therefore, the MCL values were combined over these factors in subsequent analyses. The MCLs for the best likeness task are shown in Figure 7. For the children, a positive MCL was obtained for the familiar faces (6-year-olds, MCL = 3.41%, SE = 2.06%; 8-year-olds, MCL = 15.66%, SE = 0.99%; 10-year-olds, MCL = 8.89%, SE = 1.26%). For familiar faces, the adults showed a negative MCL value (adults, MCL = –2.83%, SE = 1.36%). By contrast, a negative MCL was found for unfamiliar faces for all age groups (6-year-olds, MCL = –3.27%, SE = 1.71%; 8-year-olds, MCL = –8.47%, SE = 1.94%; 10-year-olds, MCL = –9.62%, SE = 1.47%; adults, MCL = –3.31%, SE = 1.58%). The MCLs were analyzed in a three-way ANOVA, with age (6 years, 8 years, 10 years, adult) as a between-subjects factor and caricature level (–36%, –18%, 0%, 18%, 36%) and familiarity (familiar or unfamiliar) as repeated within-subject factors. The main effect of age was significant, $F(3, 119) = 5.33$, $p < .01$ (the MCLs of the 8-year-olds were significantly larger than those of the adults according to Tukey's pairwise comparisons, $p < .05$; none of the other groups differed significantly from each other). The main

[3] The same pattern of significant results reported in this section was also obtained when the median scores and a log transform of the median scores were used.

Table 2
Means and Standard Errors of Correct Response Times (in Milliseconds) for Familiar and Unfamiliar Faces

Age	Familiar faces	Unfamiliar faces	Difference
6 years			
M	822	1,014	192
SE	8.40	8.59	
8 years			
M	703	827	124
SE	5.50	6.82	
10 years			
M	677	791	114
SE	6.27	8.61	
Adults			
M	672	737	65
SE	5.92	5.85	

effect of familiarity was also significant, $F(1, 119) = 156.18$, $p < .0001$, suggesting that a caricature was preferred for familiar faces ($M = 5.93\%$), whereas an anticaricature was preferred for unfamiliar faces ($M = -6.0\%$).[4] Finally, the Age × Familiarity interaction was significant, $F(3, 119) = 31.61$, $p < .0001$.

For the sake of clarity, separate ANOVAs on the MCLs of each age group were carried out to further examine the Age × Familiarity interaction. For each age group, the null hypothesis was that the MCLs for familiar faces would not be different from the MCLs for unfamiliar faces. In addition, a priori *t* tests were carried out to test the null hypothesis that varying the distinctiveness of the face would have no effect on best likeness judgments of the face. Under this hypothesis, participants' MCLs would not be different from an expected MCL of 0%; that is, there would be no difference from the veridical.

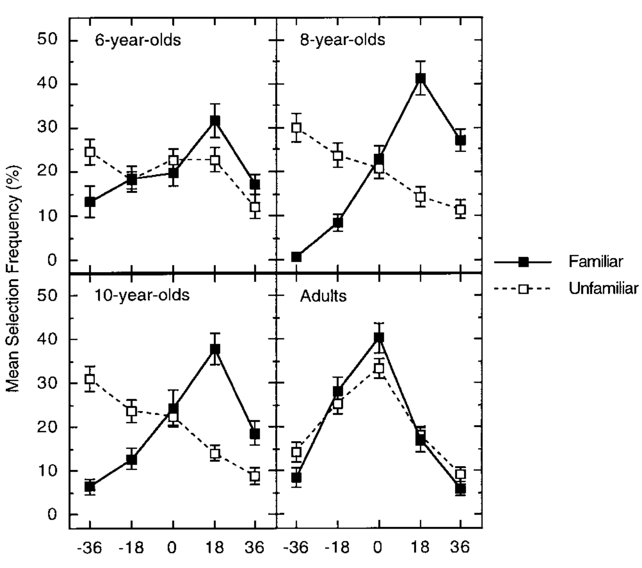

Figure 6. Mean selection frequency (expressed as a percentage) as a function of age and familiarity for best likeness judgments in Experiment 2. Error bars depict the standard error of the mean. In the graphs, a 20% selection frequency represents chance selection.

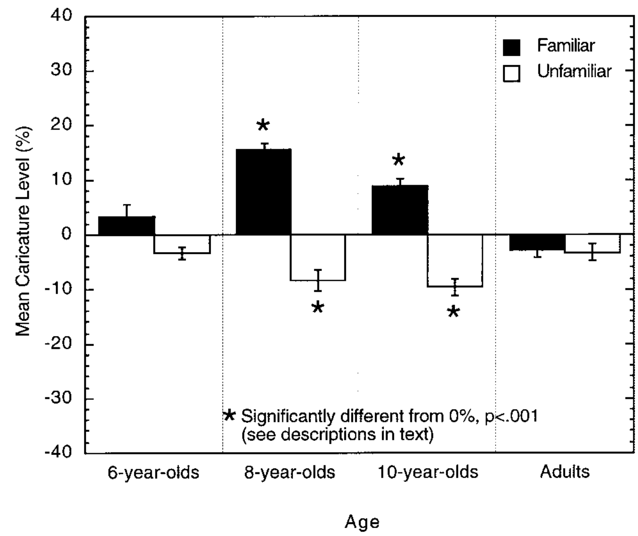

Figure 7. Mean caricature level (MCL) of best likeness judgment as a function of age and familiarity in Experiment 2. Error bars depict the standard error of the MCL. In the graph, an MCL of 0% represents chance selection.

For the children, the main effect of familiarity was significant: 6-year-olds, $F(1, 28) = 9.53$, $p < .01$; 8-year-olds, $F(1, 31) = 119.35$, $p < .0001$; 10-year-olds, $F(1, 25) = 74.59$, $p < .0001$. As suggested by the selection frequency data in Figure 6, the positive MCL values for the familiar faces seem to be due to an overall preference for the caricatures; conversely, the negative MCL values for the unfamiliar faces seem to be due to an overall preference for the anticaricatures. The planned contrasts indicated that the MCLs were significantly different from zero for the older children: 8-year-olds, $t(31) = 15.85$, $p < .001$ for familiar faces and $t(31) = -4.37$, $p < .001$ for unfamiliar faces; 10-year-olds, $t(25) = 7.04$, $p < .001$ for familiar faces and $t(25) = -6.54$, $p < .001$ for unfamiliar faces (all are according to two-tailed tests). For the 6-year-olds, the MCLs were in the same direction as those for the older children, but they were not significantly different from zero: $t(28) = 1.7$, $p > .1$ for familiar faces, and $t(28) = -1.92$, $p > .1$ for unfamiliar faces.

For the adults, the main effect of familiarity was not significant, $F(1, 35) = 4.01$, $p = .1$. Moreover, the planned contrasts indicated that the adult MCLs were marginally significantly different from 0%: $t(35) = -2.08$, $.05 < p < .1$ for familiar faces, and $t(35) = -2.1$, $.05 < p < .1$ for unfamiliar faces (all are according to two-tailed tests). For the adults, as suggested by the mean selection frequency data in Figure 6, the slightly negative MCLs for familiar and unfamiliar faces were due to an outright preference for the veridical depiction, combined with a preference for the –18% anticaricature over the 18% caricature.

[4] These overall MCLs included data from the adult participants. To anticipate Experiment 3, the strategy used by the adults to select best likeness may have underestimated these values. When the adult data were excluded, the MCLs for the children alone were 9.55% and –7.08% for the familiar and unfamiliar faces, respectively.

Discussion

The results of the identification task indicated a caricature advantage (one of the pairwise comparisons showed that the 36% caricature was recognized significantly faster than the veridical depiction) as well as an anticaricature disadvantage (one of the pairwise comparisons showed that the –36% anticaricature was recognized significantly more slowly than the veridical depiction). Thus, the manipulation of caricature level affected naming speed: Increasing the distinctiveness of a face increased naming speed, whereas reducing the distinctiveness of a face (i.e., making it more similar to the central tendency of faces) made face recognition slower.[5]

The best likeness task also provided evidence of the superiority of caricatures. For the children, the MCL of the image chosen as the best likeness for familiar faces was a positive value. Furthermore, the 8- and 10-year-olds (but not the 6-year-olds) showed a caricature advantage for best likeness judgments in which the positive MCL was significantly different from 0%. These results show the same trend as the pilot results involving children's best likeness judgments that were reported in Ellis's (1991) summary. For the children in the present study, the MCL of the image chosen as the best likeness for unfamiliar faces was a negative value. Only the 8- and 10-year-olds (but not the 6-year-olds), however, showed an anticaricature preference for unfamiliar faces in which the negative MCL was significantly different from 0%. (These data and possible explanations for the effect are discussed later.) Finally, it should be noted that for the 6-year-olds, the best likeness data (see Figure 6) show a different pattern of selection compared with the data from the distinctiveness task (Experiment 1; see Figure 4) even though we used similar procedures, but entirely different instructions, for the two tasks. This finding suggests that during the two tasks, the 6-year-olds were neither simply responding randomly nor showing a bias to respond in a certain way regardless of the instructions.

In contrast to the children, adults showed a preference for the veridical depictions as the best likeness for both familiar and unfamiliar faces. Our next experiment, motivated partly by our adult participants' feedback, explored whether the presentation of the depictions in an array played a role in this discrepancy. Several adult participants told the experimenter that they selected the best likeness through a process of elimination. They first scanned the set of five depictions and eliminated the two extremes (the 36% and –36% depictions). They then eliminated the next two extremes (the 18% and –18% depictions), which allowed them to effectively target the veridical depiction. As previously noted, the findings from Experiment 1 indicated that the adults and the older children could easily pick the extreme depictions when selecting the "most like" or "most different" depictions even when they had never seen the faces before. Perhaps, then, we should not be surprised that an elimination strategy may have been used to select the best likeness. It should be pointed out that there is no evidence that this elimination strategy was used by the children given that they showed an opposite pattern of bias in selecting caricatures and anticaricatures for familiar and unfamiliar faces, respectively. Thus, in the next experiment, only adults were tested.

Experiment 3

The purpose of this experiment was to examine adults' best likeness judgments under a different procedure; we specifically tried to eliminate the conditions under which participants could use a process of elimination to select the best likeness of an individual's face. In this experiment, we asked adults who participated in Experiment 2 to rate likeness on a 1–7 Likert-type scale when the face depictions were presented one at a time on the computer monitor.

Method

Participants. Twenty-five adults from the Department of Psychology (mean age = 28 years 1 month) were tested; 21 had participated in Experiment 2 about 11 weeks earlier. As before, all the participants were Caucasian, and the proportions of men and women were approximately equivalent.

Apparatus and procedure. The computer, monitor, software, and facial stimuli were identical to those described previously.

At the beginning of a session, participants learned the actual names of three faces in a block, according to the training procedure described in Experiment 2. Immediately after learning the three names in a block, participants made likeness ratings for a single face depiction as it was shown on the monitor.

To guide participants' likeness ratings, we read the following instructions:

> Now, I'm going to put up some pictures of [John] on the screen, one at a time. If you think that the picture is a good likeness of [John], then you should make a mark near these numbers here [experimenter pointed to the top]. If you think that the picture is a poor likeness of [John], then you should make a mark near these numbers here [experimenter pointed to the bottom]. Please try to use the whole scale, including the numbers in the middle, to tell me how good each picture is of [John].

Occasionally during the trials, the instructions were repeated, along with a reminder that there were no right or wrong answers and that the experimenter was simply interested in what participants thought of the faces. A response to each face depiction was made on a 7-point scale in which 1 represented a *poor* likeness and 7 represented a *good* likeness. After a response, another depiction was presented on the screen. The five depictions of an individual were presented in a random order. Each set of depictions of an individual's face was presented twice before another person's face was presented. This was done to try and mitigate the possibility that participants could keep track of which face depiction was shown on any given trial. Participants were allowed to proceed through the trials at their own pace.

Results and Discussion

No participant needed more than two training blocks to reach the criterion (i.e., correctly naming the three faces twice in a row). The mean likeness ratings averaged over participants as a function of caricature level and familiarity are presented in Figure 8.

A three-way repeated measures ANOVA was carried out on the mean ratings, with caricature level (–36%, –18%, 0%, 18%, 36%),

[5] As Rhodes and Tremewan (1994) pointed out, "The stipulation that caricatures and veridicals should both be better than anticaricatures ensures that the equivalence of caricatures and veridicals is non-trivial (e.g., not due to their being perceptually indistinguishable)" (p. 278).

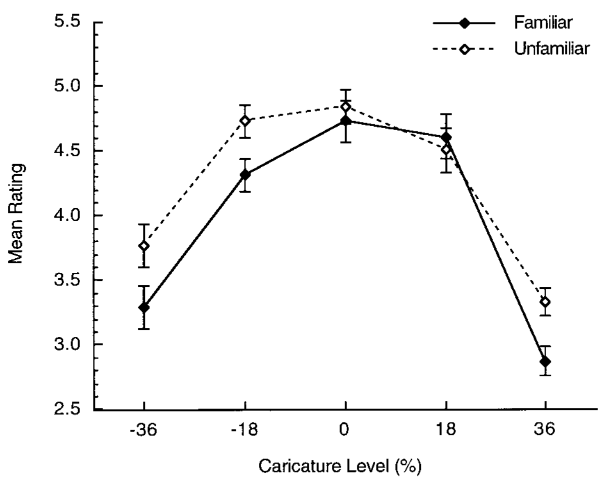

Figure 8. Mean best likeness ratings as a function of caricature level in Experiment 3. Error bars depict the standard error of the mean.

familiarity (familiar or unfamiliar), and presentation order (first or second) as within-subject factors. The main effect of familiarity was significant, $F(1, 24) = 9.41$, $p < .01$, indicating that familiar faces received lower ratings than unfamiliar faces (3.96 vs. 4.23). The main effect of presentation order was significant, $F(1, 24) = 26.17$, $p < .0001$, indicating that the ratings were higher on the second presentation than the first (4.26 vs. 3.93). Finally, the main effect of caricature level was also significant, $F(4, 96) = 20.62$, $p < .0001$. As can be seen in Figure 8, this was entirely due to the low ratings for the –36% and 36% depictions.

The only interaction between the factors that was significant was the Caricature Level $\times$ Familiarity interaction, $F(4, 96) = 3.17$, $p < .05$. For the familiar faces, Figure 8 shows that the 0% and 18% depictions received the highest ratings. Post hoc pairwise comparisons indicated that the 36% caricature received the lowest ratings and the –36% anticaricature received the next lowest ratings, with each differing significantly from all other caricature levels; the –18% depiction was rated lower than the 0% and 18% depictions (*t*s ranged from –9.9 to 12.79, *p*s $< .05$; all are according to two-tailed tests). The difference between the 0% and 18% depictions was not significant. In addition, a planned contrast that was carried out to test the hypothesis that the caricatures would be rated higher than the anticaricatures did not reach significance ($F < 1$).

For the unfamiliar faces, Figure 8 shows that the 0% and –18% depictions received the highest ratings. Post hoc pairwise comparisons indicated that the 36% caricature received the lowest ratings and the –36% anticaricature received the next lowest ratings, with each differing significantly from all other caricature levels; the rating for the –18% anticaricature was not significantly different from the ratings for the 18% or 0% depictions, but the 0% depiction was rated higher than the 18% caricature (*t*s ranged from –7.37 to 10.38, *p*s $< .001$; all are according to two-tailed tests). Recall that an unexpected finding from the best likeness task of Experiment 2 was that adult participants (and the other age groups) selected an anticaricature (rather than a caricature) as the best likeness of an unfamiliar face. This result was confirmed in Experiment 3 by a planned contrast which showed that the anticaricatures were rated higher than the caricatures, $F(1, 96) = 10.42$, $p < .01$.

The present results do not provide support for the caricature-trace hypothesis (the idea that face representations are caricatured in long-term memory) because the veridical depictions for both familiar and unfamiliar faces were given the highest ratings when the best likeness of a face was assessed. Moreover, these results are consistent with the interpretation that adult participants in Experiment 2 used a process of elimination that allowed them to target the veridical depiction. When the conditions for applying this strategy were removed, an interaction between caricature level and familiarity was found, similar to the pattern shown by the children in Experiment 2.

General Discussion

In Experiment 1, participants selected anticaricatures when asked to choose the face most like the ones they had seen before and caricatures when asked to choose the face most different from the ones they had seen before. These distinctiveness effects were observed in participants of four age groups, but the size of the effects varied with age: The youngest participants showed the smallest distinctiveness effects, and the oldest participants showed the largest effects.

The results of the identification task (Experiment 2) indicate that all of the age groups recognized caricatures at least as quickly as veridical depictions, and both of these were recognized faster than anticaricatures. This result was found regardless of whether the faces were personally familiar or newly learned. The results of the best likeness task (which was also included in Experiment 2) indicate that the 6-, 8-, and 10-year-olds, but not the adults, preferred a caricature when the face was familiar and an anticaricature when the face was unfamiliar. Adults tended to pick the veridical depiction as the best likeness regardless of familiarity. The preference for the veridical depiction was confirmed when adults were asked to rate the depictions for best likeness one at a time (Experiment 3). Under these conditions, however, there was evidence of a preference for caricatures over anticaricatures for familiar faces (the 18% caricature was rated higher than the –18% anticaricature), whereas for unfamiliar faces, there was evidence of a preference for anticaricatures over caricatures (the overall ratings for the anticaricatures were higher than the ratings for the caricatures).

The present results may now be considered together with the aim of arriving at the most parsimonious face-space model that accounts for both the younger and older participants' data. We suggest that a model of face space in which the density of space decreases as distance from the origin increases (Valentine, 1991) can account for the present results. The model assumes that in a recognition task, the observer is trying to access the target representation among distractors in memory. According to this model, the proximity of neighboring representations may determine the efficiency of recognition. Thus, the presentation of anticaricatures may result in high levels of distractor activation (due to anticaricatures' similarity to a large group of faces), which consequently increases the likelihood of interference during speeded recognition. Conversely, accessing the target representation may be relatively more efficient when the overall level of distractor activation is low (there may be fewer neighbors in close proximity), which may be the case when the stimulus (i.e., the veridical and caricatured depictions) is more distinctive than the anticaricature.

We favor a model of children's face space that resembles the adult space in organization but has fewer total exemplars in space, reflecting the fact that children have seen a fewer number of faces than have adults. This interpretation is supported by the two main results of the present experiments. First, 6-year-old children showed the same pattern of distinctiveness judgments as the older age groups (in Experiment 1, the MCLs for most like and most different judgments were significantly different from 0%). Second, all of the age groups recognized caricatures at least as quickly as veridical depictions, and caricature level did not interact with the factors of age or familiarity.

Caricature effects are generally taken to indicate that distinctiveness information in faces is encoded relative to a norm or central tendency of faces. If there was a developmental change in the processing of distinctiveness information when recognizing familiar but not unfamiliar faces, then an interaction among age, caricature level, and familiarity (Experiment 2) should have been evident. This interaction was not significant (nor was the interaction between age and caricature level). If we assume that the presence of caricature effects may be an indicator of the use of distinctive information in faces, then children as young as 6 years of age seem to be able to code whatever is distinctive about a face, though not to the same extent as older children and adults. Similar conclusions have been reached by other researchers, who have suggested that age-related differences in face recognition are due to older participants being able to simultaneously encode more features than young children (e.g., Pedelty et al., 1985; Winograd, 1981) or to the fact that older participants can encode a greater amount of facial information in a shorter amount of time than can younger children (e.g., Ellis & Flin, 1990).

The idea that an internal norm is constructed and plays a role in recognition is certainly not novel and is supported by the results of numerous category formation studies using both abstract stimuli, such as well-learned dot patterns (Posner & Keele, 1968), and schematic faces in research on adults (Reed, 1972; Solso & McCarthy, 1981) and on 10-month-old infants (Strauss, 1979). There is evidence that children as young as age 6 may be able to abstract a prototype from the configuration of well-learned dot patterns and to classify novel dot patterns in relation to the prototypical pattern (Diamond & Carey, 1990). With regard to facial stimuli, Walton and Bower (1993) found, using a preferential operant-sucking procedure, that newborns 8 to 72 hours old preferred to look at a composite (prototype) of previously seen faces rather than at a composite of unseen faces. Taken together, these results provide evidence that a general ability to encode patterns of a shared configuration is already present in early childhood.

By considering the role of a norm or central tendency of faces in the face space, we may be able to provide some account for the unexpected finding of an anticaricature preference for best likeness judgments when faces were unfamiliar (data from 8- and 10-year-olds in Experiment 2; moreover, the 6-year-olds also showed a trend in this direction). Compared with how one learns faces in everyday life, our training procedure may be adequate for the purposes of speeded identification, but it clearly does not provide the same level of experience for all of the possible variations (i.e., in expression, pose, hairstyle, etc.) that a face might have when seen in everyday life. We suggest that an anticaricature preference may represent a regression to the norm, and in this context, such biases may be interpreted by appealing to general models of categorization. A model by Huttenlocher, Hedges, and Duncan (1991) proposes that category information affects reports of particular experiences in adults and in young children (Huttenlocher, Newcombe, & Sandberg, 1994). The model assumes that memory is hierarchically organized and that information is stored at two levels of detail: One is at a specific level that captures fine-grain detail (e.g., distinctive properties of a single pose), and the other is at a category level that contains global information about the category (e.g., a norm or information about the central tendency of faces). In addition, the model makes the reasonable assumption that people's memory for fine-grain values may be inexact. In the present training procedure, learning a name from a single pose is likely to engender a large degree of uncertainty about the face even though the face can be accurately identified. At the heart of Huttenlocher et al.'s (1991) model is the assumption that people will use information about the category—in particular, knowledge about the central or prototypical value of the category—when they are asked to reconstruct a particular instance from memory. Thus, even if memory itself is unbiased, people's reports of the fine-grain value may be systematically biased toward the central tendency of the category. In this context, one speculation is that if the distinctive properties of faces have to be encoded from a single pose, as was the case in the present study, then participants may be more likely to report as the best likeness a depiction that is closer to the norm face (an anticaricature) than the veridical depiction. Adults, in contrast to the younger participants, may have less uncertainty at the specific level of representation for a newly encoded face and thus may rely less on category information.

Research that has compared the attractiveness of averaged versus nonaveraged faces is also consistent with the idea of a regression to the norm. When a composite face is created by photographically averaging a set of individual faces, the composite is generally regarded as being more attractive than the component faces that make up the average. This result holds for adults (Langlois & Roggman, 1990; Rubenstein, Langlois, Kalakanis, & Larson, 1996) and is reflected by longer looking times for infants as young as 6 months (Rubenstein, Kalakanis, & Langlois, 1999). There also appears to be a negative correlation between caricature level and perceived attractiveness regardless of whether the faces are line drawn (Rhodes & Tremewan, 1996) or photographic images (Rhodes et al., 1999). Thus, it may be that if the distinctive properties of a face are encoded from a single pose, then participants' judgments of best likeness may be disproportionately influenced by the inherent (but artificial) attractiveness of anticaricatures. In contrast, if the face is familiar and the representation contains more stable or more complete information about the distinctive properties of faces, or if the representation is exaggerated in memory (e.g., the caricature-trace model proposed by Rhodes, 1996), then participants may be better able to ignore the influence of attractiveness and focus on the distinctive properties of the image.

We interpret the present results as evidence that faces are encoded in relation to a norm or central tendency of faces. It seems that 6-year-olds may be able to use this type of information, although they may not be as proficient at using it as older participants. Within the MDFS framework, we speculate that as children grow, there may be an increasing ability to extract a norm from a range of experienced exemplars, in particular, a norm based on more information from a range of different faces. Perhaps this point may be exemplified in the recognition of unfamiliar faces of another race (although it should also apply to other classes of homogeneous stimuli with which we are less familiar). Even adults perform more poorly at recognizing unfamiliar other-race faces than same-race faces (e.g., Valentine & Endo, 1992). This may be because they have not yet fully acquired the specific information about patterns of facial variations in other-race faces or perhaps because a highly detailed norm has not been fully fleshed out. Young children may perform more poorly on same-race recognition tasks for similar reasons.

The interpretation of child face space offered here is that face recognition develops quantitatively in the sense that the number of faces in face space increases as children experience more faces. This view differs slightly from an influential theory of how face recognition develops in young children that is known as the "encoding shift" theory (Carey, 1982). According to this theory, children go through a qualitative transition between the ages of 6 and 10 years, from encoding featural information (i.e., local elements of a face that can be specified independently, such as a scar or mustache) to encoding configural information (i.e., the so-called second-order spatial relations, such as ratios of distances between the features and the global shape of the face). In the present study, we found that children as young as 6 years of age found caricatures to be highly effective depictions of faces, although the reason for this finding is difficult to disentangle vis-à-vis the featural–configural question. It may be that children show a caricature effect because caricatures emphasize the configural information that they *should* be attending to and that adults seem to attend to in noncaricatured faces. Another possibility is that caricatures emphasize the distinctive features that young children use for face recognition (e.g., Carey & Diamond, 1977). In any case, caricatured depictions seem to be effective as early as age 6, which suggests that adultlike expertise is not necessary to make use of the distinctiveness information emphasized in caricatures (see also Rhodes & Tremewan, 1994).

We are then left with the question of what might underlie the better overall level of performance at recognizing faces shown by older participants. As young children experience more faces, they become more proficient at attending to and extracting information that is potentially distinctive about an individual's face in relation to a norm or central tendency of faces. In this way, it may be worthwhile to use the caricature-generation process as a heuristic: Improvements in face recognition may be schematized by having more points outline particular parts of the face, by having more parts of the face outlined, by having the outline trace the contours of the face in a much more accurate way, and/or by having a stable norm. All of these factors would result in the distinctiveness information in a face being represented with a greater degree of accuracy. The central idea here is that young children already have the facility to encode distinctiveness information in relation to a norm or central tendency. Perhaps, however, it is only with an increasing ability to take into account more facial information that there are improvements in their model of face space, particularly in relation to the distribution of representations in face space. We suggest that the process by which faces are encoded in relation to normative information improves throughout childhood. Sensitivity to distinctiveness in the form of caricatures is one way of indexing this improvement.

References

Bartlett, J. C., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition, 12,* 219–228.

Benson, P. J. (1994). Morph transformation of the facial image. *Image and Vision Computing, 12,* 691–696.

Benson, P. J. (1997). *On the stability of recognising famous faces from photographic-quality caricatures and a conceptualisation of similarity-based processes.* Unpublished manuscript.

Benson, P. J., & Perrett, D. I. (1991a). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology, 3,* 105–135.

Benson, P. J., & Perrett, D. I. (1991b). Synthesising continuous-tone caricatures. *Image and Vision Computing, 9,* 123–129.

Benson, P. J., & Perrett, D. (1992, February 22). Face to face with the perfect image. *New Scientist,* 32–35.

Benson, P. J., & Perrett, D. I. (1994). Visual processing of facial distinctiveness. *Perception, 23,* 75–93.

Brennan, S. E. (1985). Caricature generator: The dynamic exaggeration of faces by computer. *Leonardo, 18,* 170–178.

Byatt, G., & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision Research, 38,* 2455–2468.

Calder, A. J., Young, A. W., Benson, P. J., & Perrett, D. I. (1996). Self priming from distinctive and caricatured faces. *British Journal of Psychology, 87,* 141–162.

Carey, S. (1982). Face perception: Anomalies of development. In S. Strauss & R. Stavy (Eds.), *U-shaped behavioral growth* (pp. 169–190). New York: Academic Press.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195,* 312–314.

Diamond, R., & Carey, S. (1990). On the acquisition of pattern encoding skills. *Cognitive Development, 5,* 345–368.

Ellis, H. D. (1991). *The development of face processing skills* (Final Report to the Economic and Social Research Council on Grant XC15250003). Cardiff, Wales: University of Wales College of Cardiff.

Ellis, H. D., & Flin, R. H. (1990). Encoding and storage effects in 7 year olds' and 10 year olds' memory for faces. *British Journal of Developmental Psychology, 8,* 77–92.

George, P. A., Hole, G. J., & Scaife, M. (2000). Factors influencing young children's ability to discriminate unfamiliar faces by age. *International Journal of Behavioral Development, 24,* 480–491.

Going, M., & Read, J. D. (1974). Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. *Perceptual and Motor Skills, 39,* 109–110.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98,* 352–376.

Huttenlocher, J., Newcombe, N., & Sandberg, E. H. (1994). The coding of spatial location in young children. *Cognitive Psychology, 27,* 115–147.

Johnston, R. A., & Ellis, H. D. (1995). Age effects in the processing of typical and distinctive faces. *Quarterly Journal of Experimental Psychology, 48A,* 447–465.

Johnston, R. A., Milne, A. B., Williams, C., & Hosie, J. (1997). Do distinctive faces come from outer space? An investigation of the status of a multi-dimensional face-space. *Visual Cognition, 4,* 59–67.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1,* 115–121.

Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science, 11,* 379–385.

Newcombe, N., & Lie, E. (1995). Overt and covert recognition of faces in children and adults. *Psychological Science, 6,* 241–245.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 3–27.

Pedelty, L., Levine, S. C., & Shevell, S. (1985). Developmental changes in face processing: Results from multidimensional scaling. *Journal of Experimental Child Psychology, 39,* 421–436.

Perrett, D., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D., et al. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature, 394,* 884–887.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3,* 382–407.

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 17,* 43–63.

Rhodes, G. (1996). *Superportraits: Caricatures and recognition.* Hove, England: Erlbaum.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representation of faces. *Cognitive Psychology, 19,* 473–497.

Rhodes, G., Hickford, C., & Jeffery, L. (2000). Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive? *British Journal of Psychology, 91,* 125–140.

Rhodes, G., & Moody, J. (1990). Memory representations of unfamiliar faces: Coding of distinctive information. *New Zealand Journal of Psychology, 19,* 70–78.

Rhodes, G., Sumich, A., & Byatt, G. (1999). Are average configurations attractive only because of their symmetry? *Psychological Science, 10,* 52–58.

Rhodes, G., & Tremewan, T. (1994). Understanding face recognition: Caricature effects, inversion, and the homogeneity problem. *Visual Cognition, 1,* 275–311.

Rhodes, G., & Tremewan, T. (1996). Averageness, exaggeration, and facial attractiveness. *Psychological Science, 7,* 105–110.

Rubenstein, A. J., Kalakanis, L., & Langlois, J. H. (1999). Infant preferences for attractive faces: A cognitive explanation. *Developmental Psychology, 35,* 848–855.

Rubenstein, A., Langlois, J., Kalakanis, L., & Larson, A. (1996, July). *Attractive faces really are only average.* Paper presented at the 8th annual meeting of the American Psychological Society, San Francisco, CA.

Shepherd, J., Davies, G., & Ellis, H. (1981). Studies of cue saliency. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 105–132). London: Academic Press.

Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay on face recognition. *European Journal of Cognitive Psychology, 3,* 137–145.

Solso, R. L., & McCarthy, J. E. (1981). Prototype formation of faces: A case of pseudo-memory. *British Journal of Psychology, 72,* 499–503.

Stevenage, S. V. (1995). Can caricatures really produce distinctiveness effects? *British Journal of Psychology, 86,* 127–146.

Strauss, M. S. (1979). Abstraction of prototypical information by adults and 10-month-old infants. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 618–632.

Valentine, T. (1990). Representation and process in face recognition. In R. Watt (Ed.), *Vision and visual dysfunction: Vol. 14. Pattern recognition by man and machine* (pp. 107–124). London: Macmillan.

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

Valentine, T., & Bruce, V. (1986a). The effects of distinctiveness in recognising and classifying faces. *Perception, 15,* 525–535.

Valentine, T., & Bruce, V. (1986b). Recognizing familiar faces: The role of distinctiveness and familiarity. *Canadian Journal of Psychology, 40,* 300–305.

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology, 44A,* 671–703.

Walton, G. E., & Bower, T. G. R. (1993). Newborns form "prototypes" in less than 1 minute. *Psychological Science, 4,* 203–205.

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 181–190.

Received April 10, 2001
Revision received June 24, 2002
Accepted June 25, 2002 ■